9-2019

Effects of Pre-tests and Feedback on Performance Outcomes in Massive Open Online Courses: What Works and What Doesn’t?

Maria Janelli

The Graduate Center, City University of New York

How does access to this work benefit you? Let us know!

Follow this and additional works at: https://academicworks.cuny.edu/gc_etds

Part of the Adult and Continuing Education Commons, Educational Psychology Commons, and the Online and Distance Education Commons

Recommended Citation

https://academicworks.cuny.edu/gc_etds/3286

This Dissertation is brought to you by CUNY Academic Works. It has been accepted for inclusion in All Dissertations, Theses, and Capstone Projects by an authorized administrator of CUNY Academic Works. For more information, please contact deposit@gc.cuny.edu.
EFFECTS OF PRE-TESTS AND FEEDBACK ON PERFORMANCE OUTCOMES IN MASSIVE OPEN ONLINE COURSES: WHAT WORKS AND WHAT DOESN’T?

by

MARIA JANELLI

A dissertation submitted to the Graduate Faculty in Educational Psychology in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

2019
Effects of Pre-tests and Feedback on Performance Outcomes in Massive Open Online Courses:
What Works and What Doesn’t?

by

Maria Janelli

This manuscript has been read and accepted for the Graduate Faculty in Educational Psychology in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

Date

Anastasiya Lipnevich
Chair of the Examining Committee

Date

Bruce Homer
Executive Officer

Supervisory Committee:

Dr. Joan Lucariello, Professor of Educational Psychology, The Graduate Center, CUNY

Dr. David Randle, Senior Manager of Professional Development,
American Museum of Natural History

Dr. Jeffrey Smith, Professor, College of Education, University of Otago, New Zealand

THE CITY UNIVERSITY OF NEW YORK
Abstract of the Dissertation

Effects of Pre-tests and Feedback on Performance Outcomes in Massive Open Online Courses: What Works and What Doesn’t?

by

Maria Janelli

Advisor: Dr. Anastasiya Lipnevich

This experimental study examined the effects of pre-tests and feedback on learning outcomes in a five-week massive open online course (MOOC). The participants (N = 399) were adults from around the world who self-enrolled in the American Museum of Natural History’s (AMNH) climate change MOOC (called Our Earth’s Future) offered on the Coursera platform. Participants were randomly assigned to one of four conditions. Learners in the first treatment group took pre-tests without receiving feedback. Learners in the second treatment group took pre-tests and received basic (correct/incorrect) feedback. Learners in the third treatment group took pre-tests and received elaborate feedback. The fourth group was the control. Post-tests were administered to measure learning outcomes. Additionally, we examined links among self-efficacy, persistence, and outcome measures. Of the 606 participants assigned to the four conditions, 399 met the criteria for inclusion in the final analysis. Results of this study indicate that: (1) among all users in a MOOC, pre-tests and feedback do not affect learning outcomes; (2) the presence of pre-tests significantly and negatively affects persistence and completion, deterring some participants from progressing through the course; (3) among those who do persist
and complete the course, those who take pre-tests achieve higher learning outcomes than those who do not; and (4) among those who take pre-tests, there is a positive, cumulative effect of persistence (module completion) on learning outcomes. These findings represent a new contribution to the literature on assessment and feedback, expanding the field to include adult participants from around the world who enrolled in a self-paced, not-for-credit online science course. The results pave the way for future research in this area with this population and have a direct practical application for online course developers, offering them information to help improve student learning outcomes and engagement.
Acknowledgements

This research was made possible by generous financial and intellectual support from the American Museum of Natural History, the City University of New York, and Coursera.

Thank you to Dr. Anastasiya Lipnevich - professor, advisor, colleague, and friend. Without her, this project would have remained just an idea.

Thank you to Dr. Bruce Homer and Dr. Joan Lucariello, for reading drafts of this work and helping to shape it into something meaningful; and to Dr. David Randle and Dr. Jeffrey Smith, for serving on my committee.

Thank you to the brilliant Dr. Neil Sarnak, who taught me SQL and gave me scotch.

Finally, thank you to my dearest, Genna, who scoured the internet to find PDFs of obscure articles; helped me study for my qualifying exams while we were on dates and on vacation; deleted extraneous commas; and said she’d marry me despite the chaos of graduate school. Being a doctor is nothing compared to being her partner.

For Gram.
# Table of Contents

Abstract of the Dissertation .................................................................................. iv  
Acknowledgements ............................................................................................... vi  
List of Tables ......................................................................................................... ix  
List of Figures ........................................................................................................ x  
Introduction .......................................................................................................... 1  
Literature Review .................................................................................................. 3  
  Assessment ........................................................................................................... 3  
  Feedback ............................................................................................................. 14  
  Self-Efficacy ....................................................................................................... 27  
  Massive Open Online Courses ......................................................................... 30  
  Summary of the Present Study and Research Questions ................................. 31  
Method ................................................................................................................. 34  
  Institutional Background .................................................................................... 34  
  Participants ........................................................................................................ 36  
  Experiment ......................................................................................................... 38  
  Instrumentation ................................................................................................. 39  
  Analytic Plan ..................................................................................................... 41  
Results ................................................................................................................. 45  
  Data Preparation and Coding ........................................................................... 45  
  Attrition .............................................................................................................. 48  
  Descriptive Statistics ......................................................................................... 51  
  Analysis .............................................................................................................. 58  
Discussion ............................................................................................................ 72  
  Feedback Findings in the Present Study ............................................................ 73  
  Effects of Pre-tests on Post-test Scores .............................................................. 80  
  Effects of Pre-tests on Persistence and Course Completion ............................ 83  
  Effects of Self-efficacy as a Covariate ............................................................... 85  
  Contextualizing Findings in the MOOC Literature .......................................... 85
List of Tables

Table 1  Total Number of Test Submissions Per Module Before Data Preparation for 606 Participants
Table 2  Total Number of Test Submissions Per Module After Data Preparation for 399 Participants
Table 3  Attrition and Differential Attrition Per Measure
Table 4  Sex
Table 5  Ethnicity
Table 6  Country of Residence
Table 7  Education Level
Table 8  Sample Size
Table 9  Pre-test Means
Table 10 Post-test Means
Table 11 Course Persistence: Mean Number of Modules Completed
Table 12 Course Completion
Table 13 Mean Post-test Scores by Number of Modules Completed
List of Figures

Figure 1  AMNH Coursera test submissions as of August 2017................................. 36

Figure 2  Each of the four course samples and the course material that is included for each.............................................................. 39

Figure 3  Mean post-test composite scores of course completers for the treatment (pre-test) group versus the control group ........................................ 62

Figure 4  Mean post-test composite score based on number of modules completed for participants in the treatment (pre-test) groups........................................ 66

Figure 5  The number of modules completed by participants in all four conditions… 67

Figure 6  The number of post-tests completed per module by participants in all four conditions.............................................................. 68

Figure 7  The percentage of participants in the treatment and control groups who completed the course modules.................................................... 69

Figure 8  Course completion by condition........................................................................... 70

Figure 9  The universal, linear structure of a traditional course compared to the unique and varied structure of a MOOC............................................... 88
CHAPTER I

Introduction

Testing is often associated exclusively with summative assessment, in which it is used after a unit of instruction to measure whether or not participants have achieved desired learning outcomes. This is testing of learning. When used in formative assessment, testing supports the ongoing adaptation of teaching to improve student learning (Gikandi, Morrow, & Davis, 2011). This is testing for learning.

Research indicates that testing for learning can be highly effective (Beckman, 2008; Bjork, Storm, & deWinstanley, 2010; Kornell, Hays, & Bjork, 2009; Richland, Kornell, & Kao, 2009). For example, educational psychology studies have shown that pre-tests before instruction can help students learn and encode important concepts that are then presented in detail in future lessons (Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013). Research has also shown that the effectiveness of tests-as-instruction can be dependent on the feedback that students receive after taking a test (Richland et al., 2009), and that feedback works best when context (student ability, consequences, receptivity, etc.) is considered (Lipnevich, Berg, & Smith, 2017).

Studies about the effects of pre-testing and feedback have focused exclusively on K-12 or undergraduate populations in traditional face-to-face classrooms. To our knowledge, no studies have included an international community of adult online learners as participants. With more than 58,000,000 learners enrolled in Massive Open Online Courses (MOOCs) in 2016, and global MOOC enrollment continually growing (Global MOOC enrolment jumped again last year, 2017), this population should not be ignored.

---

1 The term adult refers to people 18-years of age and older who pursue free, informal online learning opportunities.
The experimental study herein described addressed this gap in the literature by examining the effects of pre-testing with and without feedback on individuals’ performance in a science MOOC. We investigated whether or not the testing and feedback effects observed in traditional classrooms with traditional students are also observed in a non-traditional population of international online learners. Further, we identified links among learning outcomes, persistence in the course, course completion, and student self-efficacy. Findings from this study expand the literature in this domain to include a new population and also provide educators and instructional designers with information to improve teaching, learning, and engagement in online courses.
CHAPTER II

Literature Review

This research study examined the effects of pre-testing and feedback on performance and persistence in an online science course for adults. The study also sought to describe the relations among learning outcomes, self-efficacy, and course completion. The following section is a review of the literature about assessment and feedback, including the advantages, disadvantages, implementation challenges, and best practices of pre-testing and feedback. This review also includes literature about self-efficacy and its role in educational research broadly and in feedback specifically. Lastly, there is an overview of massive open online courses.

Assessment

Assessment is an integral part of education, discussed and debated by teachers, administrators, parents, policy-makers, and researchers alike. Product-oriented assessment measures students’ achievement on a test or project. Process-oriented assessment measures the effort students put forth on homework, quizzes, and classroom engagement. Progress-oriented assessment measures learning growth. Good assessment uses distinct indicators for each of these three categories (Guskey & Jung, 2006).

Typically, the literature about assessment juxtaposes formative and summative assessment. Whereas formative assessment supports learning through feedback, summative assessment measures learning after a unit of instruction (Bull & Stephens, 1999; Gikandi et al., 2011). In other words, formative assessment can be viewed as assessment for learning and summative assessment can be viewed as assessment of learning (Gikandi et al., 2011; Scriven, 1967; Scriven, 1991). The current experiment fused the two approaches, and examined whether a summative, product-oriented assessment could be used formatively.
Like many aspects of education, formative assessment lacks a universal definition (Black & Wiliam, 1998). Formative assessment has been defined as “a diagnostic tool” (Koedinger, McLaughlin, & Heffernan, 2010, p. 490) that provides information to students about their learning progress and to teachers about which aspects of instruction should be modified (Smith & Lipnevich, 2018). Importantly, formative assessment is not used to evaluate students at the end of a unit or course (Buchanan, 2000). Rather, formative assessment is used to establish “what, how much, and how well students are learning in relation to the learning goals and expected outcomes in order to inform tailored formative feedback and support further learning” (Gikandi et al., 2011, p. 2337). It is the teaching and learning activities that occur after the formative assessment that yield change (Guskey, 2018).

In a meta-analysis of 250 publications, Black and Wiliam (1998) found that formative assessment, done well, can lead to meaningful improvements in learning (with a mean effect size range of .4 to .7). The results of a second meta-analysis of formative assessment, conducted by Kingston and Nash (2011), had a median effect size of .25 and a mean effect size of .2. These meta-analyses confirm the findings from multiple studies that formative assessment supports learning (Brookhart, 2018).

**Testing.** There are many types of assessment activities, one of which is testing (including pre-testing). Testing can be used formatively to inform students about existing gaps in their learning and to show teachers potential areas that require more instructional time (Beckman, 2008; Guskey, 2018). Formative assessments/pre-tests can also be used for instructional placement, such as identifying students who might benefit from a gifted and talented program (Guskey, 2018). Additionally, pre-tests can increase students’ awareness of instructional expectations, help students organize content (Hartley & Davies, 1976), affect memory (Richland
et al., 2009) and focus students’ attention on specific content (Hartley & Davies, 1976; Richland et al., 2009).

At the same time, testing serves another purpose. “[T]ests...become learning events in their own right” (Richland et al., 2009, p. 243). Indeed, researchers who study cognition and assessment have concluded that testing can be used for both assessment and learning new material (Koedinger et al., 2010; Richland et al., 2009). That is, in the process of taking a test the nature of students’ knowledge may be altered (Marsh, Roediger, Bjork, & Bjork, 2007). Put simply, tests facilitate learning (Bjork et al., 2010).

Testing-as-instruction is not a new idea (Guskey, 2018). Nearly a century ago, Pressey (1926) designed a simple device that could test intelligence by automatically delivering and grading objective, multiple-choice tests that previously could only be administered on paper. His device included what he described as a teaching component: new questions were not revealed until the current question was answered correctly. This feature of the testing machine immediately “taught” students which questions they got right and wrong (Pressey, 1926) — which information was true, and which was not. Admittedly, this is not a very sophisticated interpretation of teaching. But even simple tests, like those administered by Pressey in the 1920s, force students to engage with material and be active participants in their own learning (Black & Wiliam, 1998). And if one goal of instruction is to foster active learning (Kornell et al., 2009), then testing is a useful instructional strategy that can help achieve that goal.

The relation between learning and assessment is complex (Whitelock, 2011), and several researchers have conducted studies designed to help us better understand the varied learning benefits of testing. We know from this research, for example, that testing has positive effects on retention (Dunlosky et al., 2013). In multiple studies (Carpenter, 2009; Glover, 1989; Karpicke
Roediger, 2008; Richland et al., 2009; Roediger & Karpicke, 2006), students who were given a practice test performed better on a post-test than those who were not exposed to the practice test. Similar pre- and post-test studies (Bjork et al., 2010) have also shown that testing improves learning outcomes even when the tests are administered prior to learning new material. Here, the act of taking a test before a unit of instruction helps students to develop the ability to process and retain information that will be learned in the future.

Studies about testing broadly, and pre-testing in particular, abound. For example, Liggitt-Fox (1997) found that students who answered pre-test questions incorrectly were more curious about learning that particular content. In other experiments, students who had taken a practice test performed better on post-tests than those who had not taken a practice test (Dunlosky et al., 2013). In a study of undergraduate students in a science course, one class was given a pre-test with questions derived from unit learning objectives before a unit of study. A second class was given a list of learning objectives instead of a pre-test before a unit of study. At the end of the unit, both groups were given a post-test in order to identify which was more effective: reading a list of learning objectives prior to a new unit of instruction, or taking a pre-test, based upon learning objectives, prior to a new unit of instruction. The participants in the treatment group scored significantly higher than those in the control group on both the unit post-tests and the final exam, and they reported that the pre-tests motivated them to monitor and measure their own learning (Beckman, 2008).

In an earlier study about the learning that occurs from the act of taking pre-tests, Pressley and Tanenbaum (1990) found that undergraduates who answered the questions on a pre-test had significant learning gains compared to the control group, in which the pre-test questions were
read but not answered. This study supports the belief that it is the active task of **answering** pre-test questions, and not the passive task of **reading** pre-test questions, that leads to learning gains.

Scholars have several hypotheses about how and why testing is beneficial. Bjork, Storm, and deWinstanley (2010) write that one possible reason learning occurs during the test-taking process is because successfully retrieving the facts on the test modifies and reinforces the way the material is represented in the memory; thus, the very act of retrieval during a test makes the information more easily retrievable during future moments of recall. This is called the transfer-appropriate multifactor account (Bjork et al., 2010) or the retrieval hypothesis (Glover, 1989). The more times students successfully recall information, the more likely they are to successfully recall that information in the future.

Similar to the transfer-appropriate multifactor account is the amount of processing hypothesis, which is the possibility that successful recall from memory is determined by how much processing is spent on discrete information. The act of test-taking forces students to spend processing time on specific information. Therefore, any information presented on one test (which represents an opportunity for processing time) should be more easily recalled on subsequent tests (Glover, 1989).

Another hypothesis is called the procedural account of testing benefits. The procedural account of testing benefits is based on the idea that testing compels students to generate information (rather than simply read it). This act of information generation uses processes of encoding that are re-used when a test is taken again at a later time. Thus, using and re-using these encoding processes creates an advantage during subsequent test-taking (Bjork et al., 2010).

These hypotheses represent the potential direct effects of test-taking on student learning outcomes. There are potential indirect effects, as well. For example, frequent testing can
encourage distributed practice (Bjork et al., 2010), which has proven to be a highly effective study strategy (Dunlosky et al., 2013). Additionally, test-taking provides students with information (feedback) about their current level of understanding, which can help them identify areas on which they should focus future attention (Smith & Lipnevich, 2018). Finally, and perhaps most simply, tests highlight important information to which students should pay particular attention (Bjork et al., 2010).

Of course, we must consider the effects of testing if and when students answer questions incorrectly. Although it is possible that answering questions incorrectly might reinforce the wrong retrieval routes, studies show that it is also possible that answering questions incorrectly can yield positive learning outcomes. In one experiment, for example, 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). In another set of experiments, students who answered pre-test questions incorrectly and received feedback did better on the post-test than those who simply read the pre-test questions and answers (Kornell et al., 2009).

There are several possible hypotheses for these outcomes, some of which are similar to the hypotheses for the benefits of testing described previously. At its simplest level, the act of test-taking, regardless of whether or not the answers are correct, might direct students’ attention to important concepts that they will encounter in the future. Requiring knowledge to answer a question, even if that knowledge has not yet been acquired, primes the mind to establish that particular retrieval route (Richland et al., 2009). The act of test-taking itself might also result in deep processing that is similar to the deep processing that occurs during encoding (Kornell et al.,
In other words, answering questions encourages deep processing, whereas passive learning (such as reading) does not (Beckman, 2008; Richland et al., 2009).

Another hypothesis is related to retrieval. The act of recall might strengthen retrieval routes from the question to the right answer, even if the initial retrieval is not correct (Kornell et al., 2009; Richland et al., 2009). Conversely, exploring incorrect retrieval routes might weaken those routes/connections (Kornell et al., 2009).

In sum, research indicates that testing yields positive learning outcomes, even if the test questions are not answered correctly, provided that feedback and/or subsequent instruction guides students toward the correct information (Richland et al., 2009). The current study incorporated these findings by providing feedback and relevant curricular material to students who took a pre-test prior to a unit of instruction. A post-test measured learning gains. In this way, we attempted to understand whether or not the pre-test findings observed with traditional students in face-to-face classrooms are also present among non-traditional students in self-guided online courses.

**Test types.** Two types of tests are relevant for the current research study: multiple choice tests and digital tests. In the following section, we will examine each in turn. Afterward, these tests will be situated within the broader context of the study.

**Multiple-choice tests.** One of the most common types of test is the multiple-choice test. Multiple-choice tests provide benefits to both students and educators. From a student/learning perspective, these tests can serve as study guides, and they can activate prior knowledge (Marsh et al., 2007), which can improve recall on tests administered in the future (Butler & Roediger, 2008; Roediger & Marsh, 2005). They also have the potential to teach students material through the process of correctly identifying and eliminating lures (wrong answers) to arrive at the right
answers (Marsh et al., 2007). From an educator/teaching perspective, multiple-choice tests are easy to grade and they can be administered to large numbers of students (Roediger & Marsh, 2005).

Like any assessment instrument, multiple-choice tests are not perfect. Some scholars posit that these tests take a surface approach to learning due to relying on recognition and not recall (Bull & Stephens, 1999). Others note that multiple-choice tests are difficult to write, and that students may spend less time studying for a multiple-choice test (Roediger & Marsh, 2005). The overwhelming criticism, however, is about the potentially detrimental nature of lures. Exposure to incorrect information, in the form of lures/distractors, can cause students to acquire false knowledge (Butler & Roediger, 2008; Marsh et al., 2007; Roediger & Marsh, 2005). Indeed, after choosing a lure on one test, students are likely to select that same lure on future tests (Butler & Roediger, 2008; Roediger & Marsh, 2005). Lures reinforce false knowledge.

In an experiment with 24 undergraduate students, Roediger and Marsh (2005) found that taking a multiple-choice test caused some students to answer questions incorrectly on future tests\(^2\). The outcome was minimized when fewer lures were used. This is an important test development finding/strategy. Many educators often use 3-4 lures on a multiple-choice test to limit the possibility that a student will simply guess the correct answer, but it is important to remember that every lure in a test question exposes students to additional incorrect information (Butler & Roediger, 2008; Roediger & Marsh, 2005). This should be carefully considered during the creation of multiple-choice tests.

\(^2\) Though this seems to contradict the positive findings previously reported, it is important to note that this study did not include feedback or follow-up instruction. Participants in the second part of this study who took a multiple-choice pre-test \textit{and received feedback} did better on the post-test.
As noted in the aforementioned studies about assessment, answering questions incorrectly is not always harmful. One way to counter the potential negative effect of lures is to provide students with immediate feedback to adjust any misunderstanding that may have been caused by lures (Butler & Roediger, 2008; Marsh et al., 2007). In one study, Butler and Roediger (2008) examined whether or not feedback could be used to enhance the positive effects and diminish the negative effects of multiple-choice tests. The results indicate that feedback can indeed achieve those objectives. In their findings, the number of correct responses on the post-test increased when pre-test questions included feedback. Additionally, students performed better on post-tests comprised of questions that had been included on the pre-test. Not surprisingly, students who spent more time studying prior to the post-test answered more questions correctly, an effect that was observed regardless of the number of lures in a question.

Based on these findings, the researchers recommend that teachers and instructional designers incorporate feedback into multiple-choice assessments to enhance the positive effects and diminish the negative effects of the multiple-choice test format (Butler and Roediger, 2008). That is what we did for the current research study in order to determine if the findings observed by Butler and Roediger (2008) would also be true for adult online learners. Additionally, the current study allowed us to understand which specific type of feedback, if any, is optimal for use in online multiple-choice tests.

**Digital tests.** Teachers at many schools use computers to facilitate assessment (Zakrzewski & Bull, 1998), and the use of multiple-choice tests has increased (Denton, Madden, Roberts, & Rowe, 2008), in part, because new software made – and continues to make – administering multiple-choice questions fast and easy (Butler & Roediger, 2008).
There are several benefits to using computers for tests. From an educator’s perspective, computerized tests are easy to use and easy to scale (Debuse & Lawley, 2016). They can give reliable scores using fewer questions than paper-based tests, so they take less time for students to complete (Parshall, 1995). Digital assessments can provide real-time results (Office of Educational Technology, 2015), along with the ability to test new questions and collect additional information such as student response time and time between questions (Parshall, 1995). Students benefit from immediate feedback (Buchanan, 2000; Denton et al., 2008; Gikandi et al., 2011; Office of Educational Technology, 2015), which can be automated on a digital test. Additionally, Parshall (1995) noted that students prefer computerized tests to paper-based tests, a finding that was supported in surveys taken by students in Louisiana and Florida who recently took statewide standardized tests via computer. In those post-test surveys, students from both states indicated that they preferred the digital version of the test over the paper version (Louisiana Department of Education, 2014; Postal, 2015). Finally, using technology can make assessments accessible to diverse populations, including those with differing abilities, disabilities, and multi-lingual or non-English language learners (Gikandi et al., 2011; Office of Educational Technology, 2015). The features that support this extended access are language dictionaries, text-to-speech software, and the ability to increase font size and change the color contrast (Office of Educational Technology, 2015). Given the recent emphasis on digital accessibility (Randazzo, 2016), these are timely and important considerations.

Several studies have been conducted in the area of digital assessment. Buchanan (2000), for example, conducted two studies to examine (1) whether or not undergraduate students in a face-to-face course who took required online multiple-choice practice-tests during the semester scored higher than the control group on a summative assessment at the end of the course; and (2)
whether or not undergraduate students in a face-to-face course who took optional online multiple-choice practice-tests scored higher than the control group on a summative assessment at the end of the course. Results of the first study indicated that incorporating required online practice-tests into the instruction of a face-to-face course improved learning outcomes as measured on the course final exam. Results of the second study indicated that incorporating optional online practice-tests into a face-to-face course also improved learning outcomes as measured on the course final exam (Buchanan, 2000). Overall, findings from both studies support the use of online pre-tests, although a very low sample size in both studies minimizes the generalizability of the study.

Gikandi, Morrow, and Davis (2011) conducted a review of online education studies that focused on formative assessment, or assessment for learning. Their goal was to identify the ways in which online formative assessment supports students’ acquisition of domain-specific content knowledge and professional skills. Of the 91 studies that were reviewed, 18 were included in the analysis (nine online studies; eight blended studies; one that was both online and blended). The findings were mixed, and the authors concluded that online formative assessment is a strategy that needs more research. However, they acknowledged that evidence supports the use of online formative assessment for three purposes. First, online formative assessment has the ability to provide students with immediate feedback, removing the delay often associated with paper-based assessment. Second, online formative assessment has the potential to create opportunities for active learning and engagement with content by embedding the assessment into a rich learning environment. Third, online formative assessment can create equitable learning opportunities by leveraging the flexibility of technology to make material accessible to a diverse population of learners. This focus on digital accessibility is a growing trend, as indicated by the emergence of
the Web Content Accessibility Guidelines, a framework for developing online material for audiences with differing abilities. (See https://www.w3.org/TR/WCAG20/ for more information about creating accessible digital content.)

The current research study contributes to this body of research about digital testing. The experiment used pre-tests and immediate feedback within a digital learning environment to determine what kind of pre-instructional assessment and feedback, if any, is the most effective. Where the existing research focuses on traditional students, the current study focuses on non-traditional adult learners.

**Feedback**

Feedback in education is information that is provided to students about their performance relative to a particular desired outcome (Lipnevich & Smith, 2008). The information provided should help students improve their future performance (Smith & Lipnevich, 2018). Educational psychologists have long been fascinated with the effects of feedback and how to administer it in such a way that students will be compelled to act upon it (Bangert-Drowns, Kulik, Kulik, & Morgan, 1991). From a cognitive and information processing perspective, the goal of feedback is to improve students’ long-term memory of the information addressed by the feedback (Wiliam, 2018). Another goal of feedback is to help students correct misinformation and motivate them to continue learning (Mason & Bruning, 2001). There are many ways to provide feedback, and educational psychologists have conducted a variety of studies to try to understand how feedback works; why feedback works; and which types of feedback are better than others. The body of feedback research is vast, varied, and lacks a specific conclusive finding. Though researchers are not sure how feedback works, they know that, *generally speaking*, when feedback is appropriately delivered, it can improve teaching and learning (Shute, 2008; Lipnevich & Smith,
As with the review of assessment, we start the review of feedback by articulating definitions.

Defining feedback is not easy. Ramaprasad (1983) proposed a definition of feedback that could be adopted by theorists across academic domains. Feedback is “information about the gap between the actual level and the reference level of a system parameter, which is used to alter the gap in some way” (p. 4).

Denton, Madden, Roberts, and Rowe (2008) adapted Ramaprasad's (1983) definition for educational scholars. Feedback, they state, is “information about the gap between the learner's performance level and the reference level, which is used by the student to narrow that gap” (p. 487). Similarly, Shute (2008) defines formative feedback as information that indicates the gap between current and desired levels of understanding. Kluger and DeNisi (1996) and Hattie and Timperley (2007) define feedback interventions broadly as actions or agents that provide performance information to users. Lipnevich and Smith (2008) and Shute (2008) expand on that description by adding that feedback is not just information about one’s performance, but information about one’s performance that can be used to modify thoughts or behaviors as they relate to a desired state. It is important to note that information presented to a student is only feedback if it can be used by the student to help close the gap between the actual and desired states (Ramaprasad, 1983). Students must have an opportunity to process the feedback information that they receive in order for the feedback to have a positive effect on learning outcomes (Lipnevich & Smith, 2008; Lipnevich & Smith, 2009a).

---

Note, however, that feedback does not always yield positive learning outcomes, and can in some cases yield negative outcomes.
At a basic level, feedback gives students information and an opportunity to fix their mistakes and correct their misunderstandings (Butler & Roediger, 2008). But feedback is about so much more than that; there are a variety of benefits to incorporating feedback into instruction. Feedback can promote deep learning (Higgins, Hartley, & Skelton, 2002) by directing a student’s attention to task-related thoughts and behaviors (Kluger & DeNisi, 1996) and encouraging students to engage more actively both with the material (Lipnevich & Smith, 2009a) and with their own learning processes (Nicol & Macfarlane, 2006). Explanatory feedback can reduce students’ cognitive load (Shute, 2008) whereas feedback on multiple-choice tests can help mitigate some of the risks that are posed by the lures’ potential to teach students wrong information (Richland et al., 2009). The following sections will examine the features of feedback, feedback taxonomies, challenges to feedback implementation, and findings from feedback studies.

**Features of feedback.** Good feedback comprises four components: (1) data about the actual level of a measurable attribute (where the student is); (2) data about the reference level of the measured attribute (where the student needs to be); (3) information about the gap between the measured level and the reference level; and (4) a way to use that information to help close the gap (Black & Wiliam, 1998; Denton et al., 2008). Good feedback is provided to the student in an appropriately timely manner\(^4\) (Black & Wiliam, 1998; Bull & Stephens, 1999; Denton et al., 2008; Hattie & Timperley, 2007; Higgins et al., 2002; Office of Educational Technology, 2015; Shute, 2008). Good feedback is relevant to students (Black & Wiliam, 1998; Bull & Stephens, 1999) and related to a specific task (Black & Wiliam, 1998; Stobart, 2018), and it provides students with enough information to support their progress toward improvement (Bull &

---

\(^4\) For some tasks, immediate feedback is optimal; for others, delayed feedback works.
It is not enough to simply tell students whether or not they have answered a question incorrectly. Mistakes should be explained so that students can improve (Higgins et al., 2002). Even when students answer a question correctly, feedback matters (Butler & Roediger, 2008; Hattie & Timperley, 2007) because it allows students to assess their knowledge while simultaneously encoding the correct response into their memory (Butler & Roediger, 2008).

After evaluating the vast literature in the area of feedback, Nicol and Macfarlane (2006) articulated seven principles of good feedback. Good feedback, they contend:

1. helps students understand performance goals;
2. facilitates the development of students’ ability to self-assess and reflect on their work;
3. delivers high quality information to students about their learning;
4. encourages teacher and peer dialogue around learning;
5. encourages positive motivational beliefs and self-esteem;
6. provides opportunities to close the gap between the actual level and the reference level;
7. and provides information to teachers that can be used to help improve future instruction.

These features of feedback are oft-cited in the literature. However, recent scholarship in the area of instructional feedback expands upon and sometimes challenges these core tenets. In their recently published book about feedback, Smith and Lipnevich (2018), offer a broader definition of feedback: “Any information about a performance that a learner can use to improve that performance or grow in the general domain of the performance” (p. 592). This definition acknowledges that feedback can come from sources other than the classroom teacher and that,
within a given domain, learning can be both general and specific. It allows for feedback that is more complex than simply right/wrong, and allows for both mastery learning and improvement/progress along a continuum.

In addition to this new definition of feedback, Smith and Lipnevich (2018) suggest that we move away from general depictions of feedback and toward models of feedback that account for subject matter and age. Feedback in the domain of math is different than feedback in the domain of writing. Feedback for a five-year-old is different than feedback for an undergraduate. Feedback for a novice is different than feedback for an expert. These differences should be considered in feedback research and implementation.

Like Smith and Lipnevich (2018), Stobart (2018) also challenges us to move away from a model of feedback that focuses on “closing the gap.” Instead, he suggests a proficiency-based model of feedback that is based on the literature in the novice-to-expert domain. A proficiency-based model of feedback would incorporate three central components. First, it would expand the definition of feedback to include environmental inputs and input from other people who are not necessarily teachers; this acknowledges that feedback does not come solely from a teacher and that environmental factors are important. Second, it would recognize that emotions affect the way feedback is received by students and that negative emotions that arise from feedback may require mediation in the form of self-level feedback. Finally, proficiency-based feedback would focus on feedback that compels students to take immediate action. This is based on the idea of deliberate practice, in which failure to perform at a certain level or meet certain criteria is identified and immediately practiced in order to be improved.

With these three concepts at the core of proficiency-based feedback, the overall experience of feedback evolves with mastery. Novices, or advanced beginners, experience
feedback as a finite set of choices to be made. As advanced beginners become competent learners, they assume more agency in their decision-making processes. At the next stage, students become proficient, and the feedback they receive starts to include self-feedback in addition to feedback from others. Finally, experts experience feedback from environmental sources and make adjustments accordingly. The proficiency-based model of feedback is less restrictive than models typically used in education, and it closely mirrors the way we receive and act upon feedback in the real world (Stobart, 2018).

Another new model of feedback is the Student Interaction Model, which focuses on student receptivity to feedback. This model begins with the premise that feedback is not received in isolation. Contexts such as subject area, consequences of the feedback, and student expectations are important. With that central idea in mind, the model has four components: feedback, student, response, and action. Feedback includes many characteristics such as timeliness, tone, elaboration, and accuracy, among others. The feedback generated by a teacher is given to a student. Students are unique, and their reactions to feedback will reflect their uniqueness. Factors that affect students’ reactions are their ability, their history with feedback, their positive or negative dispositions toward feedback, and the type of feedback they prefer receiving. After receiving feedback, students respond both cognitively (Can the student understand and process the feedback?) and affectively (How does the feedback make the student feel?). Finally, students act on the feedback. Action includes using the feedback to improve and activate coping strategies to counter any negative feelings that may arise. Of course, students may also choose inaction, which renders the feedback meaningless (Lipnevich et al., 2017). This Student Interaction Model is designed to put the student first. By considering the unique abilities, experiences, and potential responses of individual students, teachers can tailor feedback to
minimize the likelihood that feedback will be ignored and maximize the positive effects of the students’ relations to and with feedback.

Brookhart (2018) echoes the importance of a student-centric approach to feedback, noting that feedback cannot be a teacher-based, unidirectional enterprise. For feedback to be effective, students must use it, and students’ use of feedback depends upon their views and receptivity to it. Furthermore, Brookhart (2018) posits that feedback is most useful in formative learning situations, when students have the time and ability to incorporate it into future learning. Feedback after summative assessments is, to put it simply, too late to be useful.

These recent models of feedback emphasize the importance of context. People in different life stages and in different learning environments receive and respond to feedback in different ways. Context matters. The current study contributes to this articulation of feedback by situating pre-tests and feedback within a new context: online adult science education.

Feedback taxonomies. There are many types of feedback that educators can employ. Corrective feedback is used to help redirect students who veer off course, whereas reinforcement feedback is used to convey to students that they are on the right track (Wiliam, 2018). Verification feedback indicates the correctness of a response (Mason & Bruning, 2001; Shute, 2008). Elaborative feedback, also called evaluative feedback, describes why an answer is or is not correct (Lipnevich & Smith, 2008; Mason & Bruning, 2001; Shute, 2008). Although information about the correctness of responses is valuable, students have indicated that receiving detailed comments is among the most useful forms of feedback (Lipnevich & Smith, 2009b). Finally, even no feedback is a type of feedback (Bull & Stephens, 1999).

The body of research about different types of feedback is vast, and some researchers have aggregated findings into feedback taxonomies. Kulhavy and Stock (1989) organized feedback by
content type. *Load* is the total amount of information, from a simple correct/incorrect annotation to elaborate feedback. *Form* is the degree to which the feedback is similar to the information presented during instruction. And *type of information* refers to whether the information provided in the feedback appeared in the assessment, appeared during instruction, or appeared for the first time in the feedback.

Bangert-Drowns et al. (1991) created a taxonomy of feedback organized by questions: Is the feedback intentional? How is the feedback delivered? What is the goal of the feedback? What is the content of the feedback? These questions guide the *creation of feedback*, whereas their five-stage model describes the *experience of feedback*: (1) The students’ initial state (interest, prior knowledge, etc.) is triggered by questions that cue (2) search and retrieval mechanisms. After retrieval, the students (3) respond to the question, (4) evaluate their response in light of the feedback they receive, and (5) modify their knowledge.

Hattie and Timperley’s (2007) taxonomy is also guided by questions: *Where am I going? How am I going? And where to next?* Good feedback answers these questions and can be categorized in one of four levels. The first level is feedback about a task, sometimes called corrective or verification feedback. The second level is feedback about the processes that are used to complete a task. The third level is feedback about self-regulation and the fourth level is personal feedback about the self.

Task-level feedback, the level most relevant for the current research study, can be about the correctness of an answer, or about helping the student find new or different information to supplement their knowledge. Task-level feedback is more likely to be useful if it is delivered immediately, rather than delayed. Additionally, task-level feedback should be simple; the more complex the task-level feedback becomes, the less useful it is to students. Of the four levels of
feedback, feedback about a task is the most common type used in education (Hattie & Timperley, 2007) and also the type that school principals believe to be the most effective (Lipnevich, McCallan, & Smith, 2013). However, feedback at this level does not typically generalize to other tasks, and too much feedback at this level can actually deter performance and learning by focusing students’ attention on the immediate goal rather than synthesizing feedback and instruction into a broader learning effort. Task-level feedback is powerful on its own, but it is most effective when it helps a student progress from the task to the process, and from the process to self-regulation. A meta-analysis of feedback studies indicates that task-level feedback has an effect size that ranges from .74 - 1.13 (Hattie & Timperley, 2007). Clearly, having the right information is a foundational part of learning and instruction!

Written computer-mediated feedback. Written comments on student work are a central component of good feedback. Quality written feedback is not easily scalable. It is not often given to students in a timely manner, and frequently lacks detailed suggestions about how to improve (Nicol, 2010). One way to expedite the delivery of feedback is to use computers (Mason & Bruning, 2001; Smith & Lipnevich, 2018). Indeed, in research conducted by Gikandi et al. (2011), the authors found that students preferred that their instructors provide digital feedback instead of written feedback because e-feedback (typed) was easier to read and made it easier to determine what they did wrong and what they did well.

In addition to timeliness and clarity, computer-based feedback has several other advantages over written feedback. For example, after the initial programming is complete, computers are able to provide feedback to an unlimited number of students, for an unlimited period of time. Automated computer feedback is unbiased and provides accurate, consistent feedback to students, regardless of individual student characteristics (Mason & Bruning, 2001).
Learning Management Systems (LMSs) allow for the easy creation, storage, and automation of feedback, and they track student usage data to help educators and instructional designers understand not only the effectiveness of the feedback, but also student behavior as it relates to engagement with feedback presented through the LMS (Munshi & Deneen, 2018).

Though the effects of computer-based feedback are not entirely clear (Lipnevich & Smith, 2008), several scholars have conducted experiments to better understand it. The results of these studies have been promising, with computerized feedback often yielding better learning outcomes than non-computerized feedback (Kluger & DeNisi, 1996). For example, in one experiment, conducted by Lipnevich and Smith (2009b), the authors found that students who received digital feedback perceived the computer to be unbiased; this perception helped them to feel freer to focus on their work without worrying about being judged by their instructor.

Koedinger, McLaughlin, and Heffernan (2010) also examined the use of feedback in digital assessments. Specifically, they studied the effectiveness of an online math tutoring program called ASSISTments. ASSISTments is a digital math assessment that provides instructional support (feedback) to the students who use it, and detailed student activity reports (feedback) to teachers. Participants included 1,240 seventh grade students from three different treatment schools. Students in a fourth school served as the control. The pre-test variable was the students’ 6th grade final grade in math. The post-test variable was the students’ 7th grade final grade in math. Results of the study indicated that those who used the ASSISTments software had better learning outcomes at the end of the 7th grade than those in the control group. Additional statistical analyses determined that general education populations benefitted from the software, but the gains were much greater for special education students.
In another study, researchers sought to understand student and teacher perspectives of a computer-based assessment and feedback delivery system in undergraduate and post-graduate classes. Six teachers who taught eight classes were included in the study. During the course of the trial, these teachers delivered feedback to their students via computers. Follow-up interviews and surveys were used to understand student and faculty opinions about the system and the experience of e-grading. Results indicated that legibility was the biggest benefit of using the system. Timeliness, specificity, detail, and amount of feedback were also commonly listed benefits. All survey categories were positive except one: students wanted the feedback to be more constructive; note that this is a shortcoming of the individual teachers who wrote the feedback, not a shortcoming of the digital feedback delivery system itself. Overall, the teachers included in the study were able to use the digital feedback system to provide their students with personalized feedback. The time spent setting up the system initially was offset by the time savings the teachers realized once they became familiar with the system (Debuse & Lawley, 2016).

Finally, and notably, meta-analyses of computer-assisted feedback show positive effects. Hattie & Timperley’s (2007) meta-analysis of feedback studies indicates that computer-assisted feedback has an effect size of .52, making it one of the most effective types of feedback educators can employ. A meta-analysis of computer-based feedback conducted in 2015 showed mixed results, with the largest effect size (.49) from elaborate feedback and feedback in the math domain, and the smallest effect size (.05) from simple corrective feedback (Van der Kleij, Feskens, & Eggen, 2015). Finally, three meta-analyses of technology-assisted feedback on student writing also showed positive effects (Bangert-Drowns, 1993; Cochran-Smith, 1991; Goldberg, Russell, & Cook, 2003).
Feedback challenges. Incorporating feedback into instruction and assessment is not easy. In addition to the many feedback considerations discussed thus far, Hattie and Timperley (2007) note that it is important that the language of feedback be appropriate for the student’s educational level. Higgins, Hartley, and Skelton (2002) concur, asserting that teachers and instructional designers who create feedback must carefully consider the language that they use to ensure that it is meaningful to the students who will receive the feedback.

For web-based interventions, such as the one used in this research study, identifying the appropriate language for feedback is difficult to do. The varied nature of a public online audience makes it impossible to determine the reading level of users. Additionally, most consumers of web-based content scan web pages looking for information relevant to their interests. This is a useful strategy for high-literacy readers, but not for low-literacy readers (Nielsen, 2005). Though it is not easy to identify the literacy levels of individual users, we know the literacy levels of the broader American population. As of 2003, 29% of adults were reading at a basic level. Even more alarming is the fact that 14% of adults were reading at a below basic level (National Assessment of Adult Literacy, 2003).

How do we meet the literacy needs of adult learners in an online educational space? Experts from two organizations offer similar solutions. Nielsen (2005) recommends that all web-based text that appears on main pages should be tailored for a 6th grade reading level and that all web-based text that appears on other pages should be tailored for an 8th grade reading level. The organization that oversees the Web Content Accessibility Guidelines recommends that “[w]hen text requires reading ability more advanced than the lower secondary education level after removal of proper names and titles, supplemental content, or a version that does not require reading ability more advanced than the lower secondary education level, is available” (How to
Meet WCAG 2.0, 2016, n. pg.). This, as one can imagine, is an expensive solution, as it requires that all material (content, assessments, and feedback) be developed multiple times for audiences of varied literacy levels.

Language and accessibility are not the only challenges to implementing feedback. Large classes can make it difficult for teachers to provide feedback in a timely manner (Bull & Stephens, 1999; Denton et al., 2008). Even in small classes, many teachers struggle to write feedback that can help their students close the gap between their current and desired learning states (Bull & Stephens, 1999). Digital assessment and feedback are not immune to these challenges (Denton et al., 2008), and even have additional obstacles to overcome. These include needing tech-savvy administrators, an on-site technical support staff, technological infrastructure, software, and the financial resources to pay for them (Bull & Stephens, 1999).

Adopting and sustaining new technologies is costly, and therefore not widespread. Furthermore, many technology-based feedback interventions are not grounded in educational theories. Good intentions abound, but more than that is needed (Munshi & Deneen, 2018).

One last and not insignificant challenge to successful feedback implementation is that the effects of feedback are not well understood. Many studies have been conducted to try to better understand the effects of feedback. Findings have been conflicting. Feedback can have short-term benefits with no long-term benefits. It can also have long-term benefits with no short-term benefits. Feedback findings are often contradictory (Wiliam, 2018) and few, if any, studies examine the effects of feedback with international adult populations enrolled in informal online classes that never meet face-to-face. There is much that we know about feedback; there is much more we have yet to understand. Smith and Lipnevich (2018) offer several areas of focus for future research. First, they recommend that researchers try to understand feedback with more
precision; by studying the nature of specific feedback, we can move toward a better understanding of its effects. Second, they acknowledge the need for more efficient feedback; here, technology can be especially helpful. Lastly, they suggest examining non-cognitive factors that relate to feedback, such as motivation and self-efficacy.

The current research study addressed these needs by attempting to identify the specific type of digital feedback (if any) that works for adult learners; testing the efficiency of delivering elaborate feedback in an online Learning Management System; and seeing if there are relations among self-efficacy, perceptions of feedback, learning outcomes, and course persistence.

**Self-Efficacy**

Pre-testing with feedback is certainly not the only way to support learning outcomes. Indeed, there are many non-cognitive skills, behaviors, and strategies that also promote academic performance (Bertling, Borgonovi, & Almonte, 2016; Gutman & Schoon, 2013; Murano, Martin, Burress, & Roberts, 2018). Among these skills is self-efficacy (Gutman & Schoon, 2013), defined as a person’s beliefs about her or his ability to perform a series of actions in service of a particular task (Bandura, 2006; Gutman & Schoon, 2013; Pajares, 1996; Schunk, 1991; Zimmerman, 2000) or beliefs about her or his ability to cope with potential situations (Bandura, 1982). Bandura (2006) posits that self-efficacy is not a global trait that can be universally measured; rather, it varies by domain.

There are three properties of self-efficacy beliefs: level, generality, and strength (Bandura, 2006). The level of one’s self-efficacy depends on the difficulty of a task. For example, she can solve addition problems, but not algebraic equations. The generality of self-efficacy relates to transferability. For example, if she can add, can she also multiply? Finally, the
strength of self-efficacy is about certainty (Zimmerman, 2000). For example, my classmates and I are certain that we can complete our doctorates.

Self-efficacy is important to understand because studies indicate that people’s self-efficacy affects behavior, cognition, and affect (Gutman & Schoon, 2013; Pajares, 1996). Cognitively, individuals’ beliefs about their abilities influence how they think (strategically, positively, negatively, etc.) (Bandura, 2006). These beliefs also affect behavior: the goals people pursue/abandon, the challenges they seek to overcome, the effort they are willing to expend, the degree to which they will persist in the face of adversity, and the choices they will make throughout their lives (Bandura, 1982; Bandura, 2006; Pajares, 1996). From an emotional perspective, people with low self-efficacy believe that tasks are difficult, which can cause stress and/or an inability to find solutions to problems. People with high self-efficacy, on the other hand, tend to feel calmer in the face of challenging tasks or circumstances. Due to the variety of ways in which it influences thoughts, actions, and feelings, self-efficacy is a strong predictor of achievement (Pajares, 1996; Schunk, 2012).

Self-efficacy is important for more than its predictive value. It is also important because it is closely related to motivation (Bandura, 1993; Schunk, 1991; Zimmerman, 2000). Students with high self-efficacy are eager about the task at-hand and can expect successful outcomes for themselves. This expectation fosters positive thinking and motivation, both of which support achievement. Students with low self-efficacy, on the other hand, are task-avoidant. This imagined failure fosters doubt and negativity, which undermine motivation and, consequently, achievement (Bandura, 1993; Schunk, 1991; Schunk, 2012).

If self-efficacy is important for student learning, and we take a student-centered approach to feedback, then it would serve us well to incorporate self-efficacy into any feedback framework
or intervention. To do this, we turn once again to the Student Interaction Model by Lipnevich, Berg, and Smith (2017). Recall that this model of feedback attempts to minimize the likelihood that feedback will be ignored and maximize its utility by taking into account students’ unique abilities, experiences, and potential responses to feedback. Specifically, the student part of the model includes three factors: (1) the student’s ability; (2) the student’s prior success with the task or domain; and (3) the student’s receptivity to feedback. Self-efficacy (a student’s beliefs about her ability) is directly affected by these factors. By considering these factors as we design assessments and deliver feedback, we attempt to strike a balance between correcting errors, providing guidance, and supporting students’ beliefs in themselves.

In a study about student perspectives on feedback, Lipnevich and Smith (2009b) found that students’ self-efficacy can be maintained in the face of poor grades when detailed feedback that includes praise is provided. Praise in feedback softens the blow, though it does not directly affect learning. The authors also found that providing students with grades can negate the positive effects of detailed feedback because grades have the potential to reduce students’ self-efficacy and create negative feelings toward the assessment itself. Finally, students reported that low grades decreased their sense of self-efficacy, whereas high grades decreased their motivation.

Research like this supports the consideration of self-efficacy in the development of assessments and feedback. The current study contributes to this body of work by identifying relations between self-efficacy and learning outcomes; self-efficacy and course completion; and self-efficacy and persistence. Most of the research on self-efficacy in education is about children or young adults in college. The current study expands upon that work by including adult participants from around the world.
Massive Open Online Courses

Massive Open Online Courses (MOOCs) are a popular self-guided online learning resource. Though they are often compared to traditional higher education courses, the comparison is not always appropriate. For example, unlike traditional online courses, MOOCs are usually free, cannot be taken for credit, have no cap on enrollment, and do not have much engagement with faculty (Pappano, 2012). With regard to the present study, two aspects of MOOC development and implementation are important to keep in mind: (1) MOOC research is still in its infancy; and (2) MOOCs deserve special attrition considerations. Let us address each of these in turn.

First, MOOCs became popular in 2012 (Pappano, 2012) and the vast amount of data collected by MOOC providers (Coursera, edX, Udacity, and the like) made it quickly apparent that these platforms could be used for educational research. Today, many educators and scholars use MOOCs for learning science research, convening at regularly scheduled annual conferences to share their methods and findings. However, because this type of online teaching and learning is just seven years old, published MOOC research is not nearly as voluminous as traditional teaching and learning research. For this reason, there are no related MOOC studies upon which to draw for the current experiment. This is the first formal exploration of automated multiple-choice pre-tests with varying feedback types in MOOCs.

Second, 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). By comparison, the freshmen retention rate in higher education is 81% with a six-year graduation rate of 60% (Undergraduate Retention and Graduation Rates, 2018). The current study demonstrates the
ways in which this high attrition rate requires that the data be analyzed from multiple perspectives, rather than the single perspective of traditional higher education research.

**Summary of the Present Study and Research Questions**

This literature review identifies the effects of testing and feedback on learning outcomes. In summary, both testing and feedback can positively affect students’ learning, but they can also have no effects or even negative effects. Context matters, and it is important to try to understand and implement the best testing and feedback applications for a particular group of learners. Non-cognitive factors, such as self-efficacy, also affect learning outcomes. Students with different levels of self-efficacy will differentially respond to assessment results and feedback.

The purpose of the current study is to contribute to this body of knowledge by identifying the effects of pre-tests, feedback, and self-efficacy on performance, persistence, and completion in an online science course for adults. The study took place in an online climate change course (Our Earth’s Future) created by the American Museum of Natural History. The course was comprised of five learning modules, each of which started with a pre-test and ended with a post-test. The study had four conditions: Participants randomly assigned to the first condition took pre-tests with no feedback. Participants randomly assigned to the second condition took pre-tests with basic corrective (right/wrong) feedback. Participants randomly assigned to the third condition took pre-tests with elaborate (detailed) feedback. The fourth condition was the control group in which participants only took the post-test. Participants in all four conditions were asked to take a pre-course survey that included questions about self-efficacy and receptivity to feedback.

The following research questions were addressed. Questions that are not italicized were part of the initial approved dissertation proposal. Questions that are italicized were incorporated
into the study after it became clear that additional analyses would provide greater insight into the potential effects of the intervention.

1. Effects on Post-test Scores

a. Were there differences in post-test scores among conditions?
   
   i. Were there differences in post-test scores among conditions within the five separate modules?
   
   ii. Were there differences in post-test composite scores between the treatment group and the control group?
   
   iii. Were there differences in post-test scores between the treatment group and the control group within the five separate modules?
   
   iv. Were there differences in post-test composite scores among conditions for the sample that submitted a post-test in all five modules?
   
   v. Were there differences in post-test composite scores between the treatment group and the control group for the sample that completed all five modules?

b. Controlling for condition, did self-efficacy explain variability in post-test scores?

c. Was there an effect of condition on post-test scores for participants who took all five post-tests?
   
   i. Was there an effect of treatment on post-test scores for participants who took all five post-tests?

---

5 Throughout the study, the treatment group is defined as the participants who were in the pre-test groups, regardless of the type of feedback that they received.
ii. *Was there an effect of the number of modules completed on post-test composite scores?*

2. Effects on Course Persistence

   a. Were there group differences, by condition, in the level of course persistence, as indicated by the number of modules that were completed?

   i. *Were there group differences, by treatment, in the level of course persistence, as indicated by the number of modules that were completed?*

   b. Controlling for condition, did self-efficacy predict the level of course persistence?

3. Effects on Course Completion

   a. Were there group differences, by condition, in the likelihood of course completion as indicated by attempting all five module post-tests?

   b. Controlling for condition, did self-efficacy predict the likelihood of course completion?

The findings described in the following chapters are an important contribution to the literature on assessment, feedback, and self-efficacy, expanding the field to include adult participants from around the world enrolled in a not-for-credit massive open online science course. The results are applicable to online educators, educational technologists, and instructional designers, offering them information to help improve student learning outcomes and engagement for adults in informal online courses.
CHAPTER III

Method

The goal of this experimental investigation was to identify the effects of pre-tests, feedback, and self-efficacy on learning outcomes and persistence in Our Earth’s Future, a climate science MOOC offered by the American Museum of Natural History on the Coursera platform. The study used a multivariate design with random assignment. The following sections include: (a) background information about the institutions through which the study was conducted; (b) information about the people who have enrolled in Our Earth’s Future; (c) a description of the experimental design; (d) the instrumentation used for the study; and (e) the analytic plan.

Institutional Background

Massive Open Online Courses (MOOCs) are a popular online resource for self-guided learning. The course model is simple: Universities and educational institutions partner with a MOOC provider. Faculty from those educational institutions create courses on the MOOC platform. People from around the world create an account with a MOOC provider and enroll in the courses that interest them. Course content typically includes lecture videos, freely available readings, discussion forums, and assessments. Enrollment, engagement, and completion are voluntary. MOOC popularity soared in 2012 (Pappano, 2012; Special report: Lifelong learning, 2017), and MOOCs remain a compelling option for anyone in pursuit of low-stakes, high-quality, informal educational content.

With more than 190 institutional partners, more than 3200 courses, more than 40 million learners from around the world (Maggioncalda, 2019), and more than 130 million course enrollments (Hickey & Urban, 2019), Coursera is one of the leading providers of MOOCs.
Coursera grew out of a belief that the best courses, at the best schools, taught by the best teachers, should not be exclusively available to the handful of participants who attend elite universities (Koller, 2012). The founders of Coursera recognized that learning lasts a lifetime, and that informal learning is a critical part of one’s lifelong education.

In 2013, the American Museum of Natural History partnered with Coursera to offer science MOOCs. Since then, more than 150,000 people have enrolled in AMNH’s six Coursera MOOCs (Coursera Our Earth’s Future Analytics, 2018). Each course spans four or five modules/weeks, and each weekly module is comprised of essays, videos, links to related resources, and multiple-choice tests. Half of the AMNH Coursera courses also include a peer-reviewed essay assignment.

*Our Earth’s Future*, a five-module course about climate change, was co-designed by a team of scientists, educators, instructional designers, videographers, and graphic designers. In this course, essays, images, videos, and tests are used to teach people from around the world about the evidence for climate change and how to communicate this evidence to others. The recorded lectures feature Dr. Debra Tillinger, an AMNH online educator who holds a Ph.D. in ocean and climate physics from Columbia University. *Our Earth’s Future* was selected for this study because it has higher student engagement than the other courses as measured by total number of test submissions per module. See Figure 1 for the number of quiz submissions per module across the AMNH MOOCs from September 2015 – August 2017 (the time when this study was designed).
Participants

The participants of this experiment were adults from all over the world who self-enrolled in *Our Earth’s Future*. In keeping with the terms of the IRB approval for this study, the participants received two informed consent disclosures in the course: one in the syllabus, for the AMNH IRB, and one sent via email, for the CUNY IRB.

Based on demographic data collected since 2015, it was expected that the majority of participants (43%) would be from North America; that 8.4% of participants would be from India; that 5% would be from the United Kingdom; that 4.9% would be from Canada; and that 3.1% would be from China⁶ (Coursera *Our Earth’s Future* Analytics, 2018).

---

⁶ Based on data from 7,151 learners. Estimates accurate to ± 1.2 percentage points.
It was also expected, based on several years of demographic data, that 42% of course participants would be female and 57% would be male. The expected age range was from 13-years-old to 65+, with the majority of participants (46.3%) falling within the 25 to 44-year-old range\(^7\) (Coursera Our Earth’s Future Analytics, 2018). Note that per the IRB agreement, data from minors (anyone under the age of 18) was excluded from this study.

Additionally, previously collected demographic data indicated that 30% of the course participants were likely to be enrolled in full- or part-time formal education programs\(^8\). 76.8% of previous course participants had at least a Bachelor’s degree and 97.5% had at least a high school diploma\(^9\). 62.8% of previous course participants had either full- or part-time jobs\(^10\). Of the 8,782 people who enrolled in the course since the launch in September 2015 to December of 2017, 13.8% (1,219) had completed it (Coursera Our Earth’s Future Analytics, 2018).

This study included test and survey data from the adults who enrolled in the course from January 8, 2018 - November 12, 2018 and met the criteria for inclusion in the analyses. The course offering that began on November 12, 2018 concluded on December 24, 2018. Thus, data was collected for nearly one year. It was expected that the demographics of the participants in the experiment would approximate the 2015 – 2017 demographics of the broader Our Earth’s Future roster. During the data collection period, 606 participants enrolled in the course and submitted tests. Of those, 399 participants met the criteria for inclusion in the analyses.

\(^7\) Based on data from 1,807 learners. Estimates accurate to ± 2.3 percentage points.
\(^8\) Based on data from 891 learners. Estimates accurate to ± 3.3 percentage points.
\(^9\) Based on data from 1,364 learners. Estimates accurate to ± 2.7 percentage points.
\(^10\) Based on data from 884 learners. Estimates accurate to ± 3.3 percentage points.
Experiment

This experiment used a multivariate design and random assignment. After enrolling in the course, participants were randomly assigned to one of the following conditions: (1) pre-tests with no feedback; (2) pre-tests with basic feedback; (3) pre-tests with elaborate feedback; or (4) the control group (no pre-test). (See Appendix A for the five pre-tests with question-level feedback.) Participants in the three treatment groups were able to take a pre-test at the start of each of the five course modules to assess their existing knowledge of the evidence for climate change. Participants randomly assigned to the first treatment group received a pre-test score without an indication of which questions they answered incorrectly. Participants randomly assigned to the second treatment group received a pre-test score that indicated which questions they answered correctly and incorrectly. Participants randomly assigned to the third treatment group received a pre-test score and elaborate feedback for each question they answered incorrectly. All three treatment groups also took module-level post-tests. Participants randomly assigned to the fourth group – the control – took just the post-tests. Participants in all four groups were invited to take the pre-course self-efficacy survey prior to starting module one. See Figure 2 for the course content included in each of the four experimental conditions.

The Our Earth’s Future course was offered twelve times during the data collection period with all four conditions in each offering. For all twelve offerings, participants were randomly assigned to one of the four conditions. Data collection began on January 8, 2018 and concluded on December 24, 2018. Analysis was done in the spring of 2019.
Figure 2. Each of the four experimental conditions and the course material included for each

It is important to note that Coursera typically charges $49 for participants to access the tests in an AMNH course. In support of this research study, the administrators at Coursera generously waived this fee for the duration of the data collection period. Thus, cost was not a barrier to participation. As is true of all AMNH MOOC tests, participation was optional, and anyone who chose not to complete the tests still had access to the course content. Additionally, those who wanted to complete the tests without participating in the study were able to receive the pre- and post-tests via e-mail.

**Instrumentation**

**Pre- and post-tests.** *Our Earth’s Future* launched in September of 2015. At the time it was created, each of the five weekly modules concluded with a post-test that participants could use to assess their understanding of the content. For this study, a pre-test was added to the start of each module for the treatment groups. The pre-tests were not identical to the post-tests, but they
were organized around the same key concepts that serve as the framework for the course. The following is an example of a pre-test question, answer options, and elaborate feedback. Below that is a post-test question about the same topic.

Module One Pre-test Question

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.
   a. Option 1 feedback: Correct!
2. 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.
3. 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.
4. 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

Question 7: 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 degrees 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.

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 analysis. (5) To avoid test-taking fatigue, pre-tests had just five questions whereas post-tests had ten questions.
Post-test composite scores were used to analyze the effectiveness of the pre-test conditions. Post-test composite scores were the mean of each participant’s first post-test submission per module. See Appendix B for the five post-tests.

**Self-efficacy survey.** After enrolling in the *Our Earth’s Future* MOOC, participants received a welcome email that included a link to the pre-course survey. A link to this survey was also available in the syllabus, located on the course homepage. The self-efficacy measure was four indicators on the survey. The indicators were co-created with Dr. Anastasiya Lipnevich and the survey was administered via Survey Monkey. Each indicator was measured on a scale of 1 to 5 (disagree strongly to agree strongly). A sample item is “I can explain the evidence for climate change”. (See Appendix C for a copy of the pre-course survey.) Reliability analysis indicated high internal consistency; Cronbach’s alpha for the four self-efficacy indicators was .875. A composite self-efficacy variable was created using regression factor scores for the self-efficacy scale.

**Analytic Plan**

This study was designed to better understand three broad categories of research questions: (1) effects on post-test scores; (2) effects on course persistence; and (3) effects on course completion. These issues were examined through the following fifteen questions. Questions that are not italicized were part of the initial approved proposal. Questions that are italicized were incorporated into the study after examination of the data made it clear that additional analyses would provide greater insight into the potential effects of the intervention.
1. Effects on Post-test Scores

a. Were there differences in post-test scores among conditions? This question was answered using an ANOVA with post-hoc tests to compare the conditions.

i. *Were there differences in post-test scores among conditions within the five separate modules?* This question was answered using an ANOVA with post-hoc tests to compare the conditions.

ii. *Were there differences in post-test composite scores between the treatment group and the control group?* This question was answered with an ANOVA.

iii. *Were there differences in post-test scores between the treatment group and the control group within the five separate modules?* This question was answered with an ANOVA.

iv. *Were there differences in post-test composite scores among conditions for the sample that submitted a post-test in all five modules?* This question was answered using an ANOVA with post-hoc tests to compare the conditions.

v. *Were there differences in post-test composite scores between the treatment group and the control group for the sample that completed all five modules?* This question was answered with an ANOVA.

b. Controlling for condition, did self-efficacy explain variability in post-test scores? This question was answered with an ANCOVA by re-running the model for Q1a and adding self-efficacy as a covariate.
c. Was there an effect of condition on post-test scores for participants who took all five post-tests? This question was examined using a repeated measures ANOVA.

i. *Was there an effect of treatment on post-test scores for participants who took all five post-tests?* This question was examined using a repeated measures ANOVA.

ii. *Was there an effect of the number of modules completed on post-test composite scores?* This question was answered using an ANOVA with post-hoc tests to compare the modules.

2. Effects on Course Persistence

a. Were there group differences, by condition, in the level of course persistence, as indicated by the number of modules that were completed? This question was examined using ordinal logistic regression analysis.

i. *Were there group differences, by treatment, in the level of course persistence, as indicated by the number of modules that were completed?* This question was answered using ordinal logistic regression analysis.

b. Controlling for condition, did self-efficacy predict the level of course persistence? This question was answered by re-running the logistic regression for Q2a and adding self-efficacy as a covariate.
3. Effects on Course Completion

a. Were there group differences, by condition, in the likelihood of course completion as indicated by attempting all five module post-tests? Logistic regression analysis was used to answer this research question.

b. Controlling for condition, did self-efficacy predict the likelihood of course completion? Logistic regression with self-efficacy as a covariate was used to answer this research question.
CHAPTER IV

Results

This chapter is presented in six sections. First, we review the data preparation and coding processes that were used to export the data from Coursera and clean it for use in this study. Second, we examine attrition data, paying particular attention to the ways in which attrition in MOOCs is fundamentally different from attrition in formal educational settings. Third, descriptive statistics of the sample, pre-tests, post-tests, and surveys are presented. Fourth, effects on post-test scores are analyzed. Fifth, effects on course persistence are analyzed. Finally, effects on course completion are analyzed.

Data Preparation and Coding

The *Our Earth’s Future* course was offered twelve times for this study and data was collected from January 8, 2018 to December 24, 2018. At the conclusion of the data collection period, anonymized student test submissions and demographic data were exported from Coursera and imported into Postico. (Postico is software for aggregating data from multiple tables into a single spreadsheet using SQL code.) After importing the individual data tables from Coursera into Postico, two SQL scripts were used to aggregate the relevant data into one file. The first script aggregated test scores. The second script aggregated responses to the demographic survey administered by Coursera. Both SQL scripts were co-created with Dr. Neil Sarnak who holds a Ph.D. in computer science from New York University and is the managing director of risk technology at Credit Suisse. The scripts were verified by data engineers at Coursera. See Appendix D for the two scripts that were developed for this study.

The test data exported from Coursera included the following fields: user ID, treatment group, module number, test name, test question, test question grade, and test submission
timestamp. The demographic data exported from Coursera included the following fields: user ID, sex, country of origin, country of residence, age, ethnicity, highest level of education, employment status, student enrollment status, and English language proficiency\(^{11}\). Once this data was exported from the .csv tables and compiled into a single file it was prepared for analysis.

The file that was generated in Postico contained question-level data which was aggregated by test to create an individual test score for each user’s test submissions. This initial dataset included 4,396 pre- and post-test submissions for 606 participants. See Table 1 for the total number of test submissions per module prior to preparing the data for analysis.

Table 1

<table>
<thead>
<tr>
<th>Test Condition</th>
<th>Module 1</th>
<th>Module 2</th>
<th>Module 3</th>
<th>Module 4</th>
<th>Module 5</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-pre-test no feedback</td>
<td>168</td>
<td>97</td>
<td>71</td>
<td>58</td>
<td>56</td>
<td>450</td>
</tr>
<tr>
<td>Pre-pre-test basic feedback</td>
<td>186</td>
<td>113</td>
<td>89</td>
<td>71</td>
<td>71</td>
<td>530</td>
</tr>
<tr>
<td>Pre-pre-test elaborate feedback</td>
<td>165</td>
<td>95</td>
<td>79</td>
<td>63</td>
<td>62</td>
<td>464</td>
</tr>
<tr>
<td>Total</td>
<td>519</td>
<td>305</td>
<td>239</td>
<td>192</td>
<td>189</td>
<td>1444</td>
</tr>
<tr>
<td>Post-pre-test no feedback</td>
<td>213</td>
<td>152</td>
<td>118</td>
<td>98</td>
<td>94</td>
<td>675</td>
</tr>
<tr>
<td>Post-pre-test basic feedback</td>
<td>231</td>
<td>167</td>
<td>147</td>
<td>126</td>
<td>115</td>
<td>786</td>
</tr>
<tr>
<td>Post-pre-test elaborate feedback</td>
<td>233</td>
<td>165</td>
<td>115</td>
<td>101</td>
<td>107</td>
<td>721</td>
</tr>
<tr>
<td>Post-Control</td>
<td>218</td>
<td>211</td>
<td>125</td>
<td>104</td>
<td>112</td>
<td>770</td>
</tr>
<tr>
<td>Total</td>
<td>895</td>
<td>695</td>
<td>505</td>
<td>429</td>
<td>428</td>
<td>2952</td>
</tr>
</tbody>
</table>

Included in these results were multiple test submissions (when a user took the same test more than once), un-matched test submissions (when a user took a pre-test, but not the corresponding post-test, or vice versa), and non-sequential test submissions (when a user took the post-test before the pre-test to try to game the system). For these reasons, 606 is not the final sample size. The dataset had to be cleaned for analysis.

\(^{11}\) All of this demographic data was not needed for the present study but was exported due to its potential inclusion in future studies.
The first step in preparing the data was to remove test submissions from the six participants who reported that they were younger than 18. The next step was to delete multiple test submissions from participants, keeping only their first submission for each pre- and post-test. Though second, third, and fourth test attempts might be of interest for a future study, this particular research required only a participant’s initial submission for both the pre- and post-tests.

The next step was to delete submissions from participants whose pattern of course activity would be a confounding variable for the study. To that end, test submissions from participants who took a post-test before a pre-test were deleted. Also deleted were the test submissions from users who took a post-test less than 20 minutes after taking a pre-test, as this was evidence that they did not spend time on the actual instructional material. Finally, tests that did not have a corresponding match were deleted from the dataset. For example, if a participant took the module one pre-test but not the module one post-test, then the pre-test score was excluded from the study. Similarly, if a participant took the module four post-test but not the module four pre-test, then the post-test score was excluded from the study. Only matched pairs were kept.

After these deletions were complete, three new variables were created. The first variable was a composite post-test score for each participant (the mean score of their post-test submissions). The second variable indicated the number of modules completed by participants. And the third variable indicated whether or not participants completed the course, as measured by the presence of post-test submissions for all five modules. Finally, the Coursera demographic data and self-efficacy responses from the pre-course survey were appended to the dataset, joined
on the shared user_id variable. Upon completing these steps, the dataset was imported into SPSS for analysis.

Whereas the initial dataset included 4,396 test submissions from 606 participants, the final dataset that was used for analysis included 2,046 test submissions from 399 participants. See Table 2 for the total number of test submissions per module after the data had been prepared for analysis.

Table 2

<table>
<thead>
<tr>
<th>Test Condition</th>
<th>Module 1</th>
<th>Module 2</th>
<th>Module 3</th>
<th>Module 4</th>
<th>Module 5</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre test no feedback</td>
<td>84</td>
<td>55</td>
<td>42</td>
<td>40</td>
<td>34</td>
<td>255</td>
</tr>
<tr>
<td>Pre test basic feedback</td>
<td>84</td>
<td>71</td>
<td>57</td>
<td>51</td>
<td>46</td>
<td>309</td>
</tr>
<tr>
<td>Pre test elaborate feedback</td>
<td>86</td>
<td>66</td>
<td>48</td>
<td>45</td>
<td>37</td>
<td>282</td>
</tr>
<tr>
<td>Total</td>
<td>254</td>
<td>192</td>
<td>147</td>
<td>136</td>
<td>117</td>
<td>846</td>
</tr>
<tr>
<td>Post pre-test no feedback</td>
<td>84</td>
<td>55</td>
<td>42</td>
<td>40</td>
<td>34</td>
<td>255</td>
</tr>
<tr>
<td>Post pre-test basic feedback</td>
<td>84</td>
<td>71</td>
<td>57</td>
<td>51</td>
<td>46</td>
<td>309</td>
</tr>
<tr>
<td>Post pre-test elaborate feedback</td>
<td>86</td>
<td>66</td>
<td>48</td>
<td>45</td>
<td>37</td>
<td>282</td>
</tr>
<tr>
<td>Post Control</td>
<td>101</td>
<td>80</td>
<td>64</td>
<td>57</td>
<td>52</td>
<td>354</td>
</tr>
<tr>
<td>Total</td>
<td>355</td>
<td>272</td>
<td>211</td>
<td>193</td>
<td>169</td>
<td>1200</td>
</tr>
</tbody>
</table>

Attrition

Attrition is the difference between the original sample and the sample for which data has been obtained. This decline in data can affect study outcomes by introducing bias which creates a threat to internal validity. Though many studies ignore attrition, it is important to present baseline data tables so that readers can be aware of this decline and the potential bias introduced by it (Dumville, Torgerson, & Hewitt, 2006).

Attrition occurs in all MOOCs. Our Earth’s Future is no exception. Five measures were examined for this study: post-test composite scores, change in scores over time, persistence, course completion, and self-efficacy. Of the five, three (post-test composite scores, persistence,
and course completion) had a 0% attrition rate. The change in scores over time and self-efficacy measures both had high attrition rates. See Table 3 for a summary of the attrition rates per measure for each of the four conditions.

Table 3

Attrition and Differential Attrition Per Measure

<table>
<thead>
<tr>
<th>Measures</th>
<th>Original Sample</th>
<th>Test Sample</th>
<th>Attrition</th>
<th>Differential Attrition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Post-test composite scores</td>
<td>399</td>
<td>399</td>
<td>0.00%</td>
<td>--</td>
</tr>
<tr>
<td>Pre-test no feedback</td>
<td>98</td>
<td>98</td>
<td>0.00%</td>
<td>--</td>
</tr>
<tr>
<td>Pre-test basic feedback</td>
<td>102</td>
<td>102</td>
<td>0.00%</td>
<td>--</td>
</tr>
<tr>
<td>Pre-test elaborate feedback</td>
<td>96</td>
<td>96</td>
<td>0.00%</td>
<td>--</td>
</tr>
<tr>
<td>Control</td>
<td>103</td>
<td>103</td>
<td>0.00%</td>
<td>--</td>
</tr>
<tr>
<td>Change in scores over time</td>
<td>399</td>
<td>138</td>
<td>65.41%</td>
<td>--</td>
</tr>
<tr>
<td>Pre-test no feedback</td>
<td>98</td>
<td>22</td>
<td>77.55%</td>
<td>28.04%</td>
</tr>
<tr>
<td>Pre-test basic feedback</td>
<td>102</td>
<td>34</td>
<td>66.66%</td>
<td>17.15%</td>
</tr>
<tr>
<td>Pre-test elaborate feedback</td>
<td>96</td>
<td>30</td>
<td>68.75%</td>
<td>19.24%</td>
</tr>
<tr>
<td>Control</td>
<td>103</td>
<td>52</td>
<td>49.51%</td>
<td>--</td>
</tr>
<tr>
<td>Persistence</td>
<td>399</td>
<td>399</td>
<td>0.00%</td>
<td>--</td>
</tr>
<tr>
<td>Pre-test no feedback</td>
<td>98</td>
<td>98</td>
<td>0.00%</td>
<td>--</td>
</tr>
<tr>
<td>Pre-test basic feedback</td>
<td>102</td>
<td>102</td>
<td>0.00%</td>
<td>--</td>
</tr>
<tr>
<td>Pre-test elaborate feedback</td>
<td>96</td>
<td>96</td>
<td>0.00%</td>
<td>--</td>
</tr>
<tr>
<td>Control</td>
<td>103</td>
<td>103</td>
<td>0.00%</td>
<td>--</td>
</tr>
<tr>
<td>Course completion</td>
<td>399</td>
<td>399</td>
<td>0.00%</td>
<td>--</td>
</tr>
<tr>
<td>Pre-test no feedback</td>
<td>98</td>
<td>98</td>
<td>0.00%</td>
<td>--</td>
</tr>
<tr>
<td>Pre-test basic feedback</td>
<td>102</td>
<td>102</td>
<td>0.00%</td>
<td>--</td>
</tr>
<tr>
<td>Pre-test elaborate feedback</td>
<td>96</td>
<td>96</td>
<td>0.00%</td>
<td>--</td>
</tr>
<tr>
<td>Control</td>
<td>103</td>
<td>103</td>
<td>0.00%</td>
<td>--</td>
</tr>
<tr>
<td>Self-efficacy</td>
<td>399</td>
<td>48</td>
<td>87.97%</td>
<td>--</td>
</tr>
<tr>
<td>Pre-test no feedback</td>
<td>98</td>
<td>11</td>
<td>88.78%</td>
<td>5.28%</td>
</tr>
<tr>
<td>Pre-test basic feedback</td>
<td>102</td>
<td>13</td>
<td>87.25%</td>
<td>3.75%</td>
</tr>
<tr>
<td>Pre-test elaborate feedback</td>
<td>96</td>
<td>7</td>
<td>92.71%</td>
<td>9.21%</td>
</tr>
<tr>
<td>Control</td>
<td>103</td>
<td>17</td>
<td>83.50%</td>
<td>--</td>
</tr>
</tbody>
</table>

The attrition rate for the self-efficacy measure is higher than the attrition rate for the change in scores over time measure. This is because the source data is different. The repeated measures analysis that was conducted to analyze the change in scores over time used post-test data from the course; the analysis of self-efficacy used data from the pre-course survey. The data
source notwithstanding, there is no denying that both attrition rates are abysmal. However, though they are concerning, they are not unusual. Attrition within MOOCs is notoriously high – approximately 92 – 97% (Hew & Cheung, 2014; Williams, Stafford, Corliss, & Reilly, 2018). Indeed, within the AMNH MOOC ecosystem, the mean attrition rate for all six courses is 91.78% (Coursera Our Earth’s Future Analytics, 2019). The 12.03% response rate for the pre-course survey is also typical of MOOCs (Evans, Baker, & Dee, 2015).

The high attrition rates for these two measures were handled differently. Because the repeated measures analyses automatically excluded users who stopped participating over time, no missing values were replaced for those analyses; instead, the repeated measures analyses were conducted with the data from the 138 participants who completed the course.

The attrition rate for self-efficacy was addressed in several ways. The dataset violated the Missing Completely At Random (MCAR) assumption that is a requirement for multiple imputation. The results of Little’s MCAR test were inconclusive because the proportion of present versus missing data was too high to generate a p value. A follow-up test of the MCAR assumption was conducted by creating a dummy-coded variable for the self-efficacy aggregate score in which 1 = “data was present” and 0 = “data was missing”. An independent samples t-test of this new variable and the post-test composite score was then conducted. Unfortunately, the difference between the groups was significant, indicating that the dataset violated the Missing At Random assumption, \( t(397) = -4.268, p < .001 \) (in addition to violating the Missing Completely At Random assumption).

With these assumptions violated, using multiple imputation was no longer an option. A second option, mean imputation, was considered but quickly dismissed. Replacing the 351 missing values with the mean self-efficacy score would have biased the relation between that
variable and the outcome measures and underestimated the standard error of the self-efficacy variable resulting in an artificially reduced $p$ value (Grace-Martin, n.d.). Therefore, the self-efficacy respondents were treated as a discrete sample ($n = 48$). Although this low sample size reduced the generalizability of any findings, under the circumstances, it was the only option.

After attrition, the sample size for the self-efficacy measure was $n = 48$. The sample size for the change in scores over time measure was $n = 138$. And the sample size for the post-test composite scores, persistence, and course completion measures, for which attrition was not a factor, was $N = 399$.

**Descriptive Statistics**

Demographic data was collected from the pre-course survey that was sent to participants and from the demographic survey administered by Coursera. Of the 63 respondents who disclosed their sex on either the pre-course survey or the Coursera demographic survey, 25 were male and 38 were female. The majority of participants who submitted surveys selected White (75.8%) as their ethnicity. Seventeen countries were represented in the survey data, with the majority of respondents (42.4%) reporting that they live in the United States. Finally, those who submitted surveys indicated the highest level of academic study that they completed. The vast majority of the 66 respondents (87.8%) selected that they completed at least a Bachelor’s degree. See Tables 4 to 7 for a summary of demographic data.

**Table 4**

**Sex**

<table>
<thead>
<tr>
<th>Condition</th>
<th>Male</th>
<th>Female</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-test no feedback</td>
<td>7</td>
<td>9</td>
<td>16</td>
</tr>
<tr>
<td>Pre-test basic feedback</td>
<td>4</td>
<td>12</td>
<td>16</td>
</tr>
<tr>
<td>Pre-test elaborate feedback</td>
<td>6</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>Control</td>
<td>8</td>
<td>14</td>
<td>22</td>
</tr>
<tr>
<td>Total</td>
<td>25</td>
<td>38</td>
<td>63</td>
</tr>
</tbody>
</table>
Table 5

*Ethnicity*

<table>
<thead>
<tr>
<th>Ethnicity</th>
<th>Frequency</th>
<th>Percent</th>
<th>Cumulative Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>American Indian or Alaska Native</td>
<td>1</td>
<td>1.5</td>
<td>1.5</td>
</tr>
<tr>
<td>Asian</td>
<td>5</td>
<td>7.6</td>
<td>9.1</td>
</tr>
<tr>
<td>Decline to answer</td>
<td>3</td>
<td>4.5</td>
<td>13.6</td>
</tr>
<tr>
<td>Hispanic of any race</td>
<td>3</td>
<td>4.5</td>
<td>18.2</td>
</tr>
<tr>
<td>Race and ethnicity unknown</td>
<td>1</td>
<td>1.5</td>
<td>19.7</td>
</tr>
<tr>
<td>Two or more races</td>
<td>3</td>
<td>4.5</td>
<td>24.2</td>
</tr>
<tr>
<td>White</td>
<td>50</td>
<td>75.8</td>
<td>100.0</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>66</strong></td>
<td><strong>100.0</strong></td>
<td></td>
</tr>
</tbody>
</table>

Table 6

*Country of Residence*

<table>
<thead>
<tr>
<th>Country</th>
<th>Frequency</th>
<th>Percent</th>
<th>Cumulative Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bermuda</td>
<td>1</td>
<td>1.5</td>
<td>1.5</td>
</tr>
<tr>
<td>Canada</td>
<td>12</td>
<td>18.2</td>
<td>19.7</td>
</tr>
<tr>
<td>Colombia</td>
<td>1</td>
<td>1.5</td>
<td>21.2</td>
</tr>
<tr>
<td>Czech Republic</td>
<td>1</td>
<td>1.5</td>
<td>22.7</td>
</tr>
<tr>
<td>France</td>
<td>3</td>
<td>4.5</td>
<td>27.3</td>
</tr>
<tr>
<td>Germany</td>
<td>1</td>
<td>1.5</td>
<td>28.8</td>
</tr>
<tr>
<td>India</td>
<td>1</td>
<td>1.5</td>
<td>30.3</td>
</tr>
<tr>
<td>Indonesia</td>
<td>1</td>
<td>1.5</td>
<td>31.8</td>
</tr>
<tr>
<td>Japan</td>
<td>1</td>
<td>1.5</td>
<td>33.3</td>
</tr>
<tr>
<td>Mexico</td>
<td>3</td>
<td>4.5</td>
<td>37.9</td>
</tr>
<tr>
<td>Netherlands</td>
<td>1</td>
<td>1.5</td>
<td>39.4</td>
</tr>
<tr>
<td>Pakistan</td>
<td>1</td>
<td>1.5</td>
<td>40.9</td>
</tr>
<tr>
<td>Portugal</td>
<td>2</td>
<td>3.0</td>
<td>43.9</td>
</tr>
<tr>
<td>Spain</td>
<td>1</td>
<td>1.5</td>
<td>45.4</td>
</tr>
<tr>
<td>Switzerland</td>
<td>2</td>
<td>3.0</td>
<td>48.5</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>6</td>
<td>9.1</td>
<td>57.6</td>
</tr>
<tr>
<td>United States</td>
<td>28</td>
<td>42.4</td>
<td>100.0</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>66</strong></td>
<td><strong>100.0</strong></td>
<td></td>
</tr>
</tbody>
</table>
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 8 for the distribution of the sample among the conditions.

Table 8

<table>
<thead>
<tr>
<th>Condition</th>
<th>Frequency</th>
<th>Percent</th>
<th>Cumulative Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-test no feedback</td>
<td>98</td>
<td>24.6</td>
<td>24.6</td>
</tr>
<tr>
<td>Pre-test basic feedback</td>
<td>102</td>
<td>25.6</td>
<td>50.1</td>
</tr>
<tr>
<td>Pre-test elaborate feedback</td>
<td>96</td>
<td>24.1</td>
<td>74.2</td>
</tr>
<tr>
<td>Control</td>
<td>103</td>
<td>25.8</td>
<td>100.0</td>
</tr>
<tr>
<td>Total</td>
<td>399</td>
<td>100.0</td>
<td></td>
</tr>
</tbody>
</table>

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. The mean pre-test score for module one was .674 ($SD = .2054$) for the first treatment group (pre-test no feedback); .643 ($SD = .2084$) for the second treatment group (pre-test basic feedback); and .712 ($SD = .2161$) for the third treatment group (pre-test elaborate feedback). The mean pre-test score for module two was .436 ($SD = .2146$) for the first treatment group; .417 ($SD = .2236$) for the second treatment group; and .482 ($SD = .2956$) for the third treatment group. The mean pre-test score for module three was .476 ($SD = .2612$) for the first treatment group; .477 ($SD = .2500$) for the second treatment group; and .454 ($SD = .2324$) for the third treatment group. The mean pre-test score for module four was .755 ($SD = .2195$) for the first treatment group; .773 ($SD = .1834$) for the second treatment group; and .711 ($SD = .2470$) for the third treatment group. The mean pre-test score for module five was .488 ($SD = .2422$) for the first treatment group; .491 ($SD = .2723$) for the second treatment group; and .486 ($SD = .2382$) for the third treatment group. See Table 9 for a summary of the pre-test means and standard deviations per condition for all five modules.

Table 9

<table>
<thead>
<tr>
<th>Condition</th>
<th>Mod 1</th>
<th>Mod 2</th>
<th>Mod 3</th>
<th>Mod 4</th>
<th>Mod 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-test no feedback</td>
<td>N</td>
<td>84</td>
<td>55</td>
<td>42</td>
<td>40</td>
</tr>
<tr>
<td>Mean</td>
<td>.674</td>
<td>.436</td>
<td>.476</td>
<td>.755</td>
<td>.488</td>
</tr>
<tr>
<td>SD</td>
<td>.2054</td>
<td>.2146</td>
<td>.2612</td>
<td>.2195</td>
<td>.2422</td>
</tr>
<tr>
<td>Pre-test basic feedback</td>
<td>N</td>
<td>84</td>
<td>71</td>
<td>57</td>
<td>51</td>
</tr>
<tr>
<td>Mean</td>
<td>.643</td>
<td>.417</td>
<td>.477</td>
<td>.773</td>
<td>.491</td>
</tr>
<tr>
<td>SD</td>
<td>.2084</td>
<td>.2236</td>
<td>.2500</td>
<td>.1834</td>
<td>.2723</td>
</tr>
<tr>
<td>Pre-test elaborate feedback</td>
<td>N</td>
<td>86</td>
<td>66</td>
<td>48</td>
<td>45</td>
</tr>
<tr>
<td>Mean</td>
<td>.712</td>
<td>.482</td>
<td>.454</td>
<td>.711</td>
<td>.486</td>
</tr>
<tr>
<td>SD</td>
<td>.2161</td>
<td>.2956</td>
<td>.2324</td>
<td>.2470</td>
<td>.2382</td>
</tr>
</tbody>
</table>

The 399 participants in the sample submitted 1200 post-tests across all five modules, with 355 post-test submissions in module one, 272 post-test submissions in module two, 211 post-test
submissions in module three, 193 post-test submissions in module four, and 169 post-test submissions in module five. The mean post-test score for module one was \( .708 (SD = .2061) \) for the first treatment group (pre-test no feedback); \( .756 (SD = .2008) \) for the second treatment group (pre-test basic feedback); \( .742 (SD = .1925) \) for the third treatment group (pre-test elaborate feedback); and \( .700 (SD = .2000) \) for the control group. The mean post-test score for module two was \( .715 (SD = .1615) \) for the first treatment group; \( .701 (SD = .1785) \) for the second treatment group; \( .697 (SD = .1797) \) for the third treatment group; and \( .642 (SD = .1847) \) for the control group. The mean post-test score for module three was \( .710 (SD = .1750) \) for the first treatment group; \( .726 (SD = .1876) \) for the second treatment group; \( .733 (SD = .1883) \) for the third treatment group; and \( .716 (SD = .2132) \) for the control group. The mean post-test score for module four was \( .765 (SD = .2497) \) for the first treatment group; \( .743 (SD = .1652) \) for the second treatment group; \( .760 (SD = .2005) \) for the third treatment group; and \( .718 (SD = .2522) \) for the control group. The mean post-test score for module five was \( .765 (SD = .1555) \) for the first treatment group; \( .746 (SD = .1573) \) for the second treatment group; \( .725 (SD = .2031) \) for the third treatment group; and \( .687 (SD = .2179) \) for the control group.

In addition to module-level post-test data, post-test composite scores were generated for each participant. The post-test composite score is the mean of all post-test scores for each participant. The mean post-test composite score was \( .698 (SD = .1822) \) for the first treatment group (pre-test no feedback); \( .729 (SD = .1560) \) for the second treatment group (pre-test basic feedback); \( .706 (SD = .1664) \) for the third treatment group (pre-test elaborate feedback); and \( .692 (SD = .1676) \) for the control group. See Table 10 for a summary of the post-test means and post-test composite scores for each condition.
Table 10

Post-test Means

<table>
<thead>
<tr>
<th>Condition</th>
<th>Mod 1</th>
<th>Mod 2</th>
<th>Mod 3</th>
<th>Mod 4</th>
<th>Mod 5</th>
<th>Post-test Composite Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-test no feedback</td>
<td>N</td>
<td>84</td>
<td>55</td>
<td>42</td>
<td>40</td>
<td>34</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>.708</td>
<td>.715</td>
<td>.710</td>
<td>.765</td>
<td>.765</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>.2061</td>
<td>.1615</td>
<td>.1750</td>
<td>.2497</td>
<td>.1555</td>
</tr>
<tr>
<td>Pre-test basic feedback</td>
<td>N</td>
<td>84</td>
<td>71</td>
<td>57</td>
<td>51</td>
<td>46</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>.756</td>
<td>.701</td>
<td>.726</td>
<td>.743</td>
<td>.746</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>.2008</td>
<td>.1785</td>
<td>.1876</td>
<td>.1652</td>
<td>.1573</td>
</tr>
<tr>
<td>Pre-test elaborate feedback</td>
<td>N</td>
<td>86</td>
<td>66</td>
<td>48</td>
<td>45</td>
<td>37</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>.742</td>
<td>.697</td>
<td>.733</td>
<td>.760</td>
<td>.725</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>.1925</td>
<td>.1797</td>
<td>.1883</td>
<td>.2005</td>
<td>.2031</td>
</tr>
<tr>
<td>Control</td>
<td>N</td>
<td>101</td>
<td>80</td>
<td>64</td>
<td>57</td>
<td>52</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>.700</td>
<td>.642</td>
<td>.716</td>
<td>.718</td>
<td>.687</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>.2000</td>
<td>.1847</td>
<td>.2132</td>
<td>.2522</td>
<td>.2179</td>
</tr>
</tbody>
</table>

For this study, course persistence was measured by the number of modules completed by participants. In a traditional online course, persistence might be measured by participants’ progress from module one to module two, from module two to module three, etc. This is not a useful measure of persistence in a MOOC because course participation is highly variable. For example, research has shown that the majority of students who enroll in a MOOC complete the course material out of order (Kalkanis, 2019). Additionally, persistence in MOOCs is measured by Coursera and AMNH in terms of the number of active participants within a given module, without regard for the order in which the modules are completed. In the Our Earth’s Future sample, 44 participants (11%) chose not to begin the course with module one. Their level of persistence in the course should not be discounted simply because they did not begin the course with the first module. For these reasons, course persistence is measured not by the linear or sequential completion of each module, but rather by the total number of modules completed by each participant.
The mean number of modules completed by all participants ($N = 399$) was 3.01 ($SD = 1.699$). The mean number of modules completed by the first treatment group (pre-test no feedback) was 2.60 ($SD = 1.654$); 3.03 ($SD = 1.656$) by the second treatment group (pre-test basic feedback); 2.94 ($SD = 1.666$) by the third treatment group (pre-test elaborate feedback); and 3.44 ($SD = 1.736$) by the control group. See Table 11 for a summary of the course persistence data.

Table 11

Course Persistence: Mean Number of Modules Completed

<table>
<thead>
<tr>
<th>Condition</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-test no feedback</td>
<td>98</td>
<td>2.60</td>
<td>1.654</td>
</tr>
<tr>
<td>Pre-test basic feedback</td>
<td>102</td>
<td>3.03</td>
<td>1.656</td>
</tr>
<tr>
<td>Pre-test elaborate feedback</td>
<td>96</td>
<td>2.94</td>
<td>1.666</td>
</tr>
<tr>
<td>Control</td>
<td>103</td>
<td>3.44</td>
<td>1.736</td>
</tr>
<tr>
<td>Total</td>
<td>399</td>
<td>3.01</td>
<td>1.699</td>
</tr>
</tbody>
</table>

Course completion was measured by the presence of a post-test submission for each of the five modules. In the first treatment group (pre-test no feedback), 22 participants (22.4%) completed the course. In the second treatment group (pre-test basic feedback), 34 participants (33.3%) completed the course. In the third treatment group (pre-test elaborate feedback), 30 participants (31.3%) completed the course. And in the control group, 51 participants (49.5%) completed the course. See Table 12 for course completion data for the four conditions.

Table 12

Course Completion

<table>
<thead>
<tr>
<th></th>
<th>Completed</th>
<th>Percent</th>
<th>Did Not Complete</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-test no feedback</td>
<td>22</td>
<td>22.4</td>
<td>76</td>
<td>77.6</td>
</tr>
<tr>
<td>Pre-test basic feedback</td>
<td>34</td>
<td>33.3</td>
<td>68</td>
<td>66.7</td>
</tr>
<tr>
<td>Pre-test elaborate feedback</td>
<td>30</td>
<td>31.3</td>
<td>66</td>
<td>68.7</td>
</tr>
<tr>
<td>Control</td>
<td>52</td>
<td>50.5</td>
<td>51</td>
<td>49.5</td>
</tr>
</tbody>
</table>
In addition to pre- and post-test results, participant demographic and self-efficacy data were also collected. The demographic data came from Coursera and AMNH surveys and the self-efficacy data came from just the AMNH survey. Unfortunately, the response rate for both surveys was very low. Of the 399 participants in the study, only 66 (16.5%) answered the Coursera survey questions, 48 (12.03%) of whom also submitted the pre-course self-efficacy survey. Therefore, as noted previously, for the research questions about self-efficacy, \( n = 48 \).

For the self-efficacy survey results, as noted earlier, a composite self-efficacy variable was created using regression factor scores for the self-efficacy scale. Included in the factor analysis were the responses to the four self-efficacy indicators on the pre-course survey (\( n = 48 \)). The scale was 1 (disagree strongly) to 5 (agree strongly). The mean was centered on 0 (\( SD = 1.0 \)) with a minimum score of -3.25 and a maximum score of 1.15.

Analysis

The purpose of this study was to contribute to the literature about pre-tests and feedback by identifying the effects of pre-tests, feedback, and self-efficacy on performance, persistence, and course completion in an online science course for adults. The research questions herein examined were categorized as (1) the effects of pre-tests and feedback on post-test scores; (2) the effects of pre-tests and feedback on course persistence; and (3) the effects of pre-tests and feedback on course completion.

Effects on post-test scores. The first category had three original research questions designed to identify the effects of pre-tests, feedback, and self-efficacy on post-test scores. During the course of analysis, seven additional research questions were examined to better understand the various ways in which features of the intervention might have affected learning outcomes.
RQ1a. Were there differences in post-test scores among conditions? A one-way ANOVA with the Bonferroni post-hoc test was used to answer this question. Results indicate that there were no statistically significant differences among the group means, $F(3,395) = .939, p = .422$. Among the entire sample, pre-tests and feedback type did not affect post-test composite scores.

To fully explore this question, five additional analyses were performed: one to determine if there were any group differences by condition on the post-test scores for each module; one to determine if there were any group differences on the post-test composite scores between the groups that took a pre-test (the treatment group) and the control group; one to determine if there were any group differences on the post-test scores for each module between the treatment group and the control group; one to determine if there were any group differences by condition on the post-test composite scores for those who completed all five modules; and one to determine if there were any group differences on the post-test composite scores between the treatment group and the control group for those who completed all five modules.

The first additional research question was: *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 indicate 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, $F(3,207) = .145, p = .933$; 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 indicates that there were no statistically significant differences among the group means for module four, $F(3,99.723) = .391,$
Among the entire sample, and across all five modules, pre-tests and feedback did not affect post-test scores.

The second additional research question was: *Were there differences in post-test composite scores between the treatment group and the control group?* This question was answered by creating a new variable called treatment, in which 1 = treatment group (for all three pre-test groups) and 0 = control. Another one-way ANOVA was performed to compare the post-test composite scores of the treatment group to the control group. The result indicates that, among the entire sample, there were no statistically significant differences between the group means, \( F(1,397) = .944, p = .332 \). Among the entire sample, pre-tests did not affect post-test composite scores.

The third additional research question was: *Were there differences in post-test scores between the treatment group and the control group within the five separate modules?* This question was answered by re-running the one-way ANOVA to compare the treatment group to the control group, using the individual module post-test scores instead of the post-test composite scores as the dependent variable. Results indicate that there were no statistically significant differences between the group means for module one, \( F(1,353) = 2.268, p = .133 \); module four, \( F(1,191) = 1.187, p = .277 \); or module five, \( F(1,167) = 3.867, p = .051 \). The Levene test of homogeneity of variance for module three was significant \( (p = .028) \), indicating that the assumption had been violated. Therefore, an ANOVA with the Welch statistic was conducted. The result indicates that there were no statistically significant differences between the group means for module three, \( F(1,105.365) = .071, p = .790 \). Lastly, for module two, the difference between the group means was statistically significant, \( F(1,270) = 6.752, p = .010, d = .34 \). Among the entire sample, pre-tests did not affect post-test scores in four of the five modules. In
module two, however, there was a significant difference between the groups, with a mean post-test score of .643 for the control group and .704 for the treatment group.

The fourth additional research question was: *Were there differences in post-test composite scores among conditions for the sample that submitted a post-test in all five modules?* The analysis for this question included only the sample of participants that completed all five modules \((n = 138)\). The ANOVA with the Bonferroni post-hoc test was re-run, using the post-test composite scores as the dependent variable. Results indicate that there were statistically significant differences among the group means, \(F(3,134) = 3.005, p = .033, \eta_p^2 = .063\). However, results of the Bonferroni post-hoc test did not find any significant pairwise \(p\) values. This may be due to the low sample size, the fact that the Bonferroni test is conservative, or it might be a Type I error.

The fifth and final additional research question was: *Were there differences in post-test composite scores between the treatment group and the control group for the sample that completed all five modules?* This question was answered by re-running the one-way ANOVA using the post-test composite scores for the sub-sample \((n = 138)\) as the dependent variable. Results indicate that there was a statistically significant difference between the group means, \(F(1,136) = 7.619, p = .007, d = .459\). Among those participants who completed the course, pre-tests, regardless of feedback type, positively affected post-test scores. See Figure 3 for mean post-test composite scores by group for the sub-sample that completed the course.
RQ1b. Controlling for condition, did self-efficacy explain variability in post-test scores? Prior to answering this question, reliability analysis and principal component analysis were conducted on the self-efficacy indicators to ensure appropriate psychometric qualities. Internal consistency reliability statistics indicated high internal consistency; Cronbach’s alpha for the four self-efficacy indicators was .875. As noted previously, the sample size for participants who responded to the self-efficacy questions was greatly reduced by attrition. Therefore, the sample size for this (and subsequent analyses that include self-efficacy) is \( n = 48 \).

To answer this research question, an ANCOVA was conducted by re-running the initial model for RQ1a and adding self-efficacy as a covariate. Results indicate that there was no statistically significant difference in post-test scores among the conditions with self-efficacy as the covariate, \( F(3, 43) = .600, p = .619 \). This may be due to the low sample size that was used for this outcome measure.
RQ1c. Was there an effect of condition on post-test scores for participants who took all five post-tests? This question was answered by conducting a repeated measures mixed ANOVA using the post-test scores from the 138 participants who completed all five modules. Mauchly’s Test of Sphericity indicated that the assumption of sphericity had been violated, $\chi^2(9) = 20.634, p = .014$. Therefore, the Greenhouse-Geisser correction was used.

Results indicate that there was a statistically significant main effect of the number of modules completed on post-test scores, $F(3.758, 503.628) = 5.619, p < .001, \eta^2_p = .04$. Pairwise comparisons indicate that there was a negative effect on post-test scores between module one and module two ($p = .003$). The estimated mean for module one post-test scores was .776. This decreased to .707 for module two. However, the pairwise comparisons also indicate that there was a positive effect on post-test scores between modules four and two ($p = .005$). The estimated mean for module two post-test scores was .707. This increased to .773 for module four. There was no positive, sustained, significant effect of module completion throughout the length of the course.

Results of the overall model also indicate that there was a statistically significant main effect of condition on post-test scores; $F(1,3) = 3.005, p = .033, \eta^2_p = .063$. However, this effect was not observed in the Bonferroni post-hoc results. This may be due to the fact that the Bonferroni test is conservative, or it may be a Type I error. This is a duplication of the finding for the fourth additional analysis conducted as part of RQ1a. Lastly, there was no statistically significant interaction effect of condition and number of modules completed on post-test scores, $F(11.275, 503.628) = 1.354, p = .189$.

Similarly to RQ1a, several additional analyses were conducted to better understand the effect of persistence on post-test scores. The first additional analysis answered the following
question: *Was there an effect of treatment on post-test scores for participants who took all five post-tests?* The previous repeated measures mixed ANOVA identified the effects of the different types of feedback. This additional repeated measures mixed ANOVA identified the effects of pre-tests, regardless of feedback type.

Once again, results of the Greenhouse-Geisser model indicate that there was a statistically significant main effect of the number of modules completed on post-test scores, $F(3.758,511.132) = 5.046, p = .001, \eta^2_p = .036$. Pairwise comparisons indicate that there was a negative effect on post-test scores between module one and module two ($p = .017$). The estimated mean for module one post-test scores was .742. This decreased to .683 for module two. Pairwise comparisons also indicate that there was a positive effect on post-test scores between module two and module three ($p = .039$) and between module two and module four ($p = .002$). The estimated mean for module two post-test scores was .683. This increased to .734 for module three and to .751 for module four. Similar to the previous research question, these findings suggest that participants who completed the course had lower post-test scores after completing the second module, but higher post-test scores after completing the third and fourth modules. There was no positive, sustained, significant effect of module completion throughout the length of the course.

Results of this analysis also indicate that there was a statistically significant main effect of treatment on post-test scores for participants who took all five post-tests, $F(1,136) = 7.619, p = .007, \eta^2_p = .053$. This finding duplicates the one found for the fifth additional analysis in RQ1a and similarly suggests that pre-tests positively affected post-test scores for those who completed the course. Lastly, unlike the repeated measures mixed ANOVA for condition, there was a statistically significant interaction effect of treatment and number of modules completed on post-
test scores, $F(3.758,511.132) = 2.905, p = .024, \eta^2 = .021$. See Table 13 for mean post-test scores by module for the treatment and control groups.

Table 13

<table>
<thead>
<tr>
<th></th>
<th>Control</th>
<th>Treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td>module 1 post-test score</td>
<td>.679</td>
<td>.805</td>
</tr>
<tr>
<td>module 2 post-test score</td>
<td>.640</td>
<td>.726</td>
</tr>
<tr>
<td>module 3 post-test score</td>
<td>.727</td>
<td>.741</td>
</tr>
<tr>
<td>module 4 post-test score</td>
<td>.713</td>
<td>.788</td>
</tr>
<tr>
<td>module 5 post-test score</td>
<td>.687</td>
<td>.755</td>
</tr>
</tbody>
</table>

The second additional analysis answered the following question: Was there an effect of the number of modules completed on post-test composite scores? This question was answered by conducting an ANOVA using the post-test composite score as the dependent variable and the number of modules completed as the independent variable. The analysis included scores from all 399 participants. Results indicate that there were no statistically significant group differences, $F(4,394) = 1.945, p = .102$. Among the entire sample, persistence (module completion) did not affect post-test composite scores.

This question was further examined by re-running the ANOVA on the sub-sample of participants who were in the three treatment groups ($n = 296$). The Levene test of homogeneity of variance was significant ($p = .020$), indicating that the assumption had been violated. Therefore, an ANOVA with the Welch statistic was conducted. Results indicate that there were statistically significant group differences, $F(4,102.9) = 4.037, p = .004$. The Bonferroni post-hoc test indicates that this effect existed between module one and module five ($p = .003, d = .538$). The mean post-test composite score for participants who completed only module one was .674 and the mean post-test composite score for participants who completed all five modules was
This strongly suggests that there is a positive cumulative effect of module completion on post-test scores for those who take pre-tests. See Figure 4 for the mean post-test composite score by number of modules completed for participants in the three pre-test groups.

Figure 4. Mean post-test composite score based on number of modules completed for participants in the treatment (pre-test) groups (n = 296)

Effects on course persistence. The second category had two original research questions designed to identify the effects of pre-tests, feedback, and self-efficacy on course persistence (module completion). During the course of analysis, one additional research question was examined to better understand the effect of the intervention on course persistence.

RQ2a. Were there group differences, by condition, in the level of course persistence, as indicated by the number of modules that were completed? This question was answered using ordinal logistic regression analysis. The assumption of proportional odds was met, as determined by a full likelihood ratio test that compared the fit of the proportional odds model to a model with varying location parameters, $\chi^2(9) = 11.430, p = .247$. Results of the analysis indicate that the final model statistically significantly predicted the dependent variable over and above the
intercept-only model, $\chi^2(3) = 14.460$, $p = .002$. The odds ratio of the pre-test no feedback group completing fewer modules than the control group was $0.372$, 95% CI $[0.223, 0.620]$, a statistically significant effect, $\chi^2(1) = 14.442$, $p < .001$. The odds ratio of the pre-test elaborate feedback group completing fewer modules than the control group was $0.559$, 95% CI $[0.337, 0.929]$, a statistically significant effect, $\chi^2(1) = 5.038$, $p = .025$. There was no statistically significant effect for the pre-test basic feedback group, $\chi^2(1) = 3.620$, $p = .057$. This finding suggests that participants in the control group are likely to complete more modules than participants in the pre-test no feedback group and the pre-test elaborate feedback group. See Figure 5 for course persistence as indicated by the number of modules completed by participants in each of the four conditions and Figure 6 for the decline in engagement over time as indicated by the post-test submissions per module by participants in each of the four conditions.

![Course Persistence (Number of Modules Completed)](image)

*Figure 5.* The number of modules completed by participants in all four conditions ($N = 399$)
To fully explore the question of persistence, one additional analysis was conducted to answer the following question: *Were there group differences, by treatment (pre-tests), in the level of course persistence, as indicated by the number of modules that were completed?* This question was answered by re-running the ordinal logistic regression analysis. Results of the analysis indicate that the final model statistically significantly predicted the dependent variable 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 those in the control group were more likely to persist through the course than those in the treatment groups. In other words, pre-tests significantly and negatively affected course persistence. See Figure 7
for course persistence as indicated by the percentage of participants in the treatment and control groups who completed the course modules.

![Course Persistence by Module](image.png)

**Figure 7.** The percentage of participants in the treatment and control groups who completed the course modules ($N = 399$)

RQ2b. Controlling for condition, did self-efficacy predict the level of course persistence? This question was answered by re-running the ordinal logistic regression on the sub-sample ($n = 48$) and adding self-efficacy as a covariate. The assumption of proportional odds was not met, as determined by a full likelihood ratio test that compared the fit of the proportional odds model to a model with varying location parameters, $\chi^2(12) = 35.111, p < .001$. The two options for continuing the analysis were either a multinomial logistic regression or interpreting the results from separate binomial logistic regression analyses. Multinomial regression analysis would have necessitated the loss of the ordinality of the dependent variable. With multiple binomial logistic regression analyses, the small sample affected the results. Therefore, the results of the initial ordinal logistic regression are reported, acknowledging on the one hand that the current sample
size does not meet the assumptions of the test whereas on the other hand that it is the best of three bad options. The final model did not statistically significantly predict the dependent variable over and above the intercept-only model, $\chi^2(4) = 2.590, p = .629$.

**Effects on course completion.** The third and final category had two research questions designed to identify the effects of pre-tests, feedback, and self-efficacy on course completion.

RQ3a. Were there group differences, by condition, in the likelihood of course completion as indicated by attempting all five module post-tests? This question was answered using logistic regression analysis. 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 the pre-test no feedback group was .284, 95% CI [.154, .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 [.279, .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 [.250, .796], a statistically significant effect, $\chi^2(1) = 7.474, p = .006$. All three conditions were statistically significantly less likely than the control group to complete the course. See Figure 8 for a chart of course completion by condition.

![Course Completion](image)

*Figure 8. Course completion by condition (N = 399)*
RQ3b. Controlling for condition, did self-efficacy predict the likelihood of course completion? This question was answered using logistic regression analysis with self-efficacy as a covariate on the sub-sample of $n = 48$. The logistic regression model was not statistically significant, $\chi^2(4) = 2.830, p = .587$. 
CHAPTER V

Discussion

This study attempted to identify whether or not the positive effects of pre-tests and feedback that are observed in research with traditional students in traditional classes are also observed in a non-traditional population of adult learners in an informal online science course. More specifically, we sought to understand: (1) the effects of pre-tests and feedback types on post-test scores; (2) the effects of pre-tests and feedback types on course persistence; and (3) the effects of pre-tests and feedback types on course completion. Self-efficacy was considered as a covariate within these three types of effects. The multivariate experimental design of the study allowed for the identification of causal connections among the conditions and the outcomes. Conducting the study in a MOOC provided the opportunity to share findings with a professional community that is in a position to immediately modify its practices and continue to build upon this research.

This study uniquely contributes to the work that has been done previously in the areas of assessment and feedback. The strongest and most compelling findings are 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 on achievement. Although pre-tests were effective for these participants, they had a negative effect on course persistence and completion among the larger sample that includes 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. Self-efficacy was not a significant covariate, possibly due to the low response rate for the pre-course survey. Unexpectedly, feedback type did not affect learning outcomes.
This chapter will discuss (a) explanations for the somewhat counter-intuitive feedback findings in this study; (b) the effects of pre-tests on post-test scores, persistence, and course completion; (c) the contextualization of this study and its findings within the MOOC landscape; (d) the practical implications of this study for future MOOC development; (e) challenges and limitations of this study; and (f) directions for future research.

Feedback Findings in the Present Study

Based on the literature related to feedback, it was expected that feedback type would affect learning outcomes in the present study. This expectation was reinforced by a MOOC study that was conducted in 2015. In an analysis of student ratings of MOOCs on a course review website (www.coursetalk.com), researchers found that students reported that they appreciate immediate feedback on automated tests so that they can identify gaps in what they have learned. Furthermore, when feedback on tests is provided, students reported that they prefer specific feedback that indicates why their answer was wrong and which answer is correct (Floratos, Guasch, & Espasa, 2015). Unfortunately, these anecdotal preferences for feedback on tests were not supported in the current study with quantitative evidence that this type of feedback leads to higher learning outcomes.

Despite abundant evidence in the literature about the positive effects of feedback, findings from this work suggest that, within MOOCs, pre-test feedback has no effect on learning outcomes. Possible explanations for this include: limited prior knowledge; structural elements of the tests; participants’ receptivity and responses to feedback; and the inability to generalize feedback findings from traditional student populations to MOOC student populations. Let us examine each of these possibilities in turn.
**Prior knowledge.** One explanation for the ineffectiveness of pre-test feedback in the present study is limited prior knowledge. Several authors note that the effectiveness of feedback can sometimes 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 (Hattie & Timperley, 2007). Without sufficient prior knowledge, making these connections may not be possible. Relatedly, Bangert-Drowns et al. (1991) acknowledge the importance of prior knowledge in their five-stage model of “the state of the learner” (p. 217) that is derived from an analysis of a variety of feedback studies. The first category in the model includes, among other factors, the student’s prior knowledge.

Research findings confirm the importance of prior knowledge on the effectiveness of feedback. In one study, researchers sought to understand the effects of the type and timing of feedback in a genetics web-based learning platform. The participants in the study were secondary education students in the Netherlands, and the instrument was multiple-choice questions. The authors hypothesized that students with low prior knowledge would benefit most from immediate elaborate feedback. However, this was not the case. Results of the study found no significant effects of immediate elaborate feedback for students with low prior knowledge (Smits, Boon, Sluijsmans, & van Gog, 2008). The same may be true of the present study: participants with low prior knowledge may not have benefitted from the pre-test feedback. In fact, an examination of participants’ pre-test scores reveals limited prior knowledge in three of the five modules. The mean pre-test score for all three treatment groups was < .50 for module two, module three, and module five. (See Table 9 on page 54 for pre-test mean scores by module for the treatment groups.) 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).

**Structural elements of the test.** The structure of the tests provides a second possible explanation for the ineffectiveness of feedback. This may have had an effect in one of four ways. First, the pre-test questions may have been too dissimilar from the post-test questions. This finding was observed in a study of the effects of different types of computer-based feedback on learning outcomes with high school science students. In that experiment, researchers found that feedback was more effective when the questions that appeared during instruction were identical to the questions that appeared after instruction. As the similarity between the instructional tests and post-tests decreased, so too did the effects of feedback (Clariana, Ross, & Morrison, 1991). A similar finding was observed in a study of feedback in multiple-choice tests conducted by Butler and Roediger (2008), further supporting the idea that feedback is more effective when pre- and post-test questions are identical. The post-test questions used in the present study were different from the pre-test questions, possibly affecting feedback-related learning outcomes.

The second structural element of the tests that may have affected the usefulness of feedback is that none of the feedback groups received the correct response to the pre-test questions. In one study about this topic, researchers examined whether or not feedback could be used to enhance the positive effects of multiple-choice tests and diminish the negative effects of multiple-choice tests. They concluded that the most important piece of feedback provided to students after a multiple-choice test is not verification feedback or elaborate feedback; rather, the most important feedback to provide is the correct answer, which gives students an opportunity to encode the correct response for future retrieval attempts (Butler & Roediger, 2008).
In the present study, participants in the pre-test no feedback group received a test score without an indication of which questions were answered correctly/incorrectly. Participants in the pre-test basic feedback group received a test score with an indication of which questions were answered correctly/incorrectly. And participants in the pre-test elaborate feedback group received a test score with an indication of which questions were answered correctly/incorrectly along with elaborate feedback explaining why their incorrect answers were wrong. None of the treatment groups received feedback that indicated which answer was correct. This is the second component of the test structure that may have limited the effectiveness of the feedback.

The third structural element of the tests that may have affected the usefulness of feedback is time between tests. It is possible that too much time may have elapsed between when participants took the pre-tests and the post-tests. Unfortunately, there is no way to impose minimum or maximum time-between-tests upon participants in a MOOC. However, the Coursera platform itself could be used to track this data for a future study. Analyzing additional data such as time between tests is one of the benefits of digital assessments (Parshall, 1995) and would help to shed light on our understanding of the effectiveness of feedback types in automated assessments for MOOC students.

The fourth and final test element that may have affected the feedback outcomes is that participants may have treated the pre-tests like an open-book test. Students in a MOOC can easily do this by opening the pre-test in one browser tab and opening the course content (or any other web site) in another browser tab. Pre-tests are not timed, so participants could have toggled between the pre-test and the course content as many times 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 decreases the effectiveness of feedback (Bangert-Drowns et al., 1991).
We know that pre-test mean scores per condition were low in three of the five modules, suggesting that participants did not take advantage of the ability to pre-search for test answers. Nevertheless, this may have been a factor for the other two modules.

**Participants’ receptivity and responses to feedback.** We can look to participants’ receptivity and responses to feedback as a third possible explanation for the ineffectiveness of pre-test feedback in the present study. We know that context matters when delivering feedback, and student receptivity is a critical context to consider (Brookhart, 2018; Lipnevich et al., 2017). Individual student factors that affect receptivity to feedback include ability, history with feedback, positive or negative dispositions toward feedback, and feedback preferences (Lipnevich et al., 2017). Participants with low receptivity to feedback in the present study may not have spent enough time reviewing feedback, or they may have ignored it completely. Reading ability may have negatively affected receptivity to feedback and participants’ dispositions toward feedback may have affected their motivation to utilize it. Any one of these factors – or a combination of them – may have contributed to the feedback findings in the present study.

The first possibility is that participants with low receptivity to feedback may not have spent much time reviewing the feedback that they received. This is problematic because taking the time to review feedback is especially important after completing a multiple-choice test. We saw in the literature review that exposing students to incorrect information in the form of lures on a multiple-choice test can cause them to acquire false knowledge (Butler & Roediger, 2008; Marsh et al., 2007; Roediger & Marsh, 2005). These studies found that students who selected a lure on one multiple-choice test were likely to select the same lure on future versions of the test.
(Butler & Roediger, 2008; Roediger & Marsh, 2005). This finding demonstrates that lures can reinforce false knowledge.

One way to limit the transmission of false knowledge on multiple-choice tests is to provide students with detailed feedback so they can correct and modify their misunderstandings (Butler & Roediger, 2008; Marsh et al., 2007). In one study examining this strategy, Butler and Roediger (2008) found that students answered more questions correctly on a post-test when the pre-test questions included feedback. This suggests that one possible reason for the ineffectiveness of feedback in the present study is that students did not spend enough time reviewing it. And if the participants did not spend enough time reviewing the feedback they received, then it is possible that the act of incorrectly answering multiple-choice questions on the pre-tests led to the encoding of misinformation that was activated again during the post-tests. Clearly, the relations among automated multiple-choice tests in MOOCs, feedback, and post-test scores represent a valuable area of focus for future feedback research.

If it is possible that students did not spend enough time reviewing feedback, it is also possible that they ignored it completely. Recall that feedback is only effective if students use it (Brookhart, 2018; Lipnevich et al., 2017; Lipnevich & Smith, 2009a). Indeed, if we return once again to the taxonomy of feedback articulated by Bangert-Drowns et al. (1991), we see that two of the five stages in the model relate to what students do with the feedback that they receive. In stage four, they evaluate their own response in light of the feedback provided to them. And in stage five, they modify their knowledge accordingly. However, others have acknowledged that inaction is also a potential response to feedback (Lipnevich et al., 2017). In the present study, it is possible that participants who received elaborate feedback and were not receptive to it simply chose not to read it.
Closely related to student receptivity to feedback is student motivation. It has been found that both intrinsic and extrinsic motivation significantly predict student engagement with course content in a MOOC (Xiong, Li, Kornhaber, Pursel, & Goins, 2015). 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. In the present study, participant motivation may have been affected in two ways.

First, we must consider the feedback itself. 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. Second, it is possible that participants were not extrinsically 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, extrinsic motivation – or both – certainly would 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, even among traditional students in traditional classes. It has recently been identified by Lipnevich and Smith (2018) as an important area for future research.
Finally, reading level may have affected participants’ receptivity to feedback by decreasing their willingness or ability to read and process the pre-test feedback. The American population is not largely comprised of advanced readers. As of 2003, 29% of adults were reading at a basic level and 14% of adults were reading at a below basic level (National Assessment of Adult Literacy, 2003). When creating public-facing web pages, Nielsen (2005) recommends that developers write text at a 6th-to-8th grade reading level. The text presented in the *Our Earth’s Future* course was modified from a graduate-level climate change course. As such, the essays and test questions were written for an adult audience capable of understanding scientific terms and complex sentence structures. If reading comprehension was a challenge for the participants of the study, then it may have been a reason they did not read the feedback that they received.

In summary, there are several potential explanations for the non-significant feedback findings in this study. These include limited prior knowledge, elements of the test structure, and participants’ receptivity and responses to feedback. One last possible explanation for the feedback findings in the present study is that feedback findings with traditional students simply do not generalize to this non-traditional population of learners. Many scholars posit that the MOOC experience is fundamentally different from other kinds of learning and therefore requires special research considerations. This idea will be fully explored later in this chapter.

**Effects of Pre-tests on Post-test Scores**

Although feedback was not found to significantly affect post-test scores, the same is not true of pre-tests. Analysis of the original research questions along with subsequent questions that were generated during the data review process provided a thorough understanding of the relations between the treatment (pre-test) group and the control group. Of the tests that were conducted, several findings were statistically significant.
First, among the sub-sample of people who completed all five modules, there was a significant effect of pre-tests (without considering feedback) on post-test composite scores. This finding suggests that in a general MOOC population with low stakes and high attrition, neither pre-tests nor feedback type affect learning outcomes. However, among the small percentage of MOOC participants that completes the entire course, taking module-level pre-tests, regardless of feedback, positively affects post-test scores. Among this population, pre-tests can be used as instruction. With an effect size of .459, this was one of the strongest findings in the study.

Second, the repeated measures analyses revealed two main effects and one interaction effect. The first main effect confirms the initial finding that pre-tests positively affected post-test scores for participants who completed the course. The second was a main effect of the number of modules completed on post-test composite scores. Interestingly, this finding was both positive and negative: Among the sub-sample that completed the entire course, participants’ post-test scores decreased after module two and increased after modules three and four. There was no overall significant increase from module one to module five. This suggests that there is no sustained, cumulative effect of persistence on achievement among course completers. Finally, there was an interaction effect of treatment and number of modules completed. Across all five modules, participants in the treatment group who completed the course had higher post-test scores than participants in the control group who completed the course.

These findings are generally consistent with the body of literature related to pre-tests. We know from this work that students who take a pre-test before a unit of instruction achieve higher post-test scores than those who do not take a pre-test (Beckman, 2008). We also know that even those students who answer pre-test questions incorrectly benefit from the activity as evidenced
by their post-test scores (Kornell et al., 2009), provided they received feedback after the pre-test (Butler & Roediger, 2008).

Importantly, these pre-test findings exist 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, indicates that pre-tests do not significantly affect learning outcomes. It is only for those who completed the entire course that pre-tests affected achievement. There is a parallel, then, between MOOC students who completed the course, and traditional students. It seems that pre-test findings with traditional students do not generalize to MOOC students who do not complete all of the coursework, whereas pre-test findings with traditional student do generalize to MOOC students who do complete courses.

Third, 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, cumulative 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. See Figure 4 on page 66 for a summary of this data.

This finding suggests that there is a benefit to taking pre-tests multiple times throughout a course, prior to the start of each instructional unit. With an effect size of .538, 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 cumulative effect of that distributed practice.
Finally, several findings were inconclusive because they were outliers or because the $p$ values were significant but the post-hoc tests were not. For example, among the entire sample, those who took a pre-test, regardless of feedback, had higher post-test scores than the control group on the module two post-test. However, because this finding was only present in one of the five modules, it cannot be generalized. The finding itself may be an outlier, or it may be an effect of the material on the test, or there may be some other confounding variable present in this particular module that does not exist in the other modules. This could be explored in future studies.

Additionally, there may be an effect of condition (feedback) on post-test composite scores among the sub-sample of people who completed all five modules. Although the overall model was significant, the Bonferroni post-hoc test was not. Thus, it was not possible to identify the specific feedback type that may have been effective. This inconclusive finding may be due to the sample size, the fact that the Bonferroni test is conservative, or a Type I error. This finding, like the previous one, represents an opportunity for future research.

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. Several inconclusive findings lend themselves to additional exploration in future studies.

**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. Results from the analysis of persistence indicate that there is a negative relation between condition and the number of modules completed. Specifically, participants in the pre-test no feedback group and the pre-test
elaborate feedback group were less likely than participants in the control group to complete additional course modules. Additional analyses expanded on this finding to reveal that those in the three treatment groups were more likely than those in the control group to drop out of the course after modules one, three, and four.

Course persistence is closely related to course completion. As exposure to pre-tests significantly negatively affected persistence, so too did pre-tests significantly and negatively affect course completion. Analyses of the effects of condition on course completion indicate that participants in the three treatment groups were significantly less likely to complete the course than participants in the control group. Thus, exposure to pre-tests had a negative effect on both persistence and completion.

These findings were not 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.

Another possible reason for the absence of this finding in the related literature is that MOOC research is still in its infancy. Though it is a popular medium in which to conduct experimental studies, it is still only seven years old (Pappano, 2012). Given the limited number of MOOC studies that exist compared to general education studies, it is not surprising that this issue has not yet been studied. Therefore, this particular finding represents both a unique contribution to this domain and an opportunity for further exploration.
Effects of Self-efficacy as a Covariate

The three research questions that incorporated self-efficacy as a covariate did not result in significant findings. This may be due to the low sample size and/or the inability to statistically replace the missing values in the dataset. Of the two notable changes in attrition in this study, the one concerning the self-efficacy outcomes had the greatest impact because it restricted the ability to generate conclusive findings. For this reason, this area in particular is of interest for future studies.

Contextualizing Findings in the MOOC Literature

Conducting research in MOOCs is uniquely complex and rewarding. The platform is robust and well-designed for multivariate educational experiments. Because the population is “massive” but not well understood, the potential for research is enormous. Courses run repeatedly, so research findings can be applied immediately. But with these opportunities come challenges, some of which were mentioned previously in this chapter. In the following section, we will continue to explore this landscape.

One way to understand the findings of the present study is by situating them within the body of research that deals specifically with MOOCs. Since their popularization in 2012, MOOCs have generated vast amounts of data that have been used in educational studies. Though there are no empirical MOOC studies that directly examine automated pre-testing with feedback, some findings from other areas are relevant to this study.

For example, researchers recently conducted a study of MOOCs offered by the University of Texas at Austin and hosted on the edX platform. Their goal was to try to understand MOOC persistence and engagement relative to individual student goals at the time of enrollment. Using pre-course surveys and course activity data, they found a negative relation
between students who enrolled in STEM courses to learn about MOOCs/online learning and their engagement in the courses. The authors suggest that students who had this goal when they enrolled in the STEM courses may have had lower engagement because they left the courses after finding the information they needed, or they may have never intended to complete the courses at all. This finding was not observed in non-STEM courses (Williams et al., 2018). As *Our Earth’s Future* is a STEM course about climate science, this effect of student goal (to learn more about online learning) on course engagement may have been a factor in the attrition of many and the engagement of few.

Another factor that may have affected the present study’s outcomes is the self-guided nature of the course. In *Our Earth’s Future* (and all other AMNH MOOCs, and many other non-degree MOOCs), the presence of the instructor is limited to her appearance in course lecture videos. Dr. Tillinger does not greet students or answer questions in the discussion forums, provide individual feedback on students’ progress, or congratulate them on their achievements. This is typical of MOOCs, a consequence of being too “massive” to provide instructor presence in an individual and traditional way.

Unfortunately, this lack of interpersonal connectivity in a MOOC may affect students’ engagement with the material and persistence in the course. In a study about student engagement in MOOCs, Jung and Lee (2018) found a relation between instructor presence and persistence: Attrition increased when there was limited instructor presence in a course, an effect that was also found in a study by Hone and Said (2016). Similar findings were also observed in a recent study that identified relations among student engagement, persistence, course completion and instructor presence (Gregori, Zhang, Galván-Fernández, & Fernández-Navarro, 2018). It seems that in MOOCs, like in traditional formal education settings, the teacher matters. The more
present a teacher is, the more active the students will be. However, in a course with many thousands of students, this poses a unique challenge. For some institutional MOOC developers with significant financial resources, paying for a small army of instructional staff to provide support to students is possible. For others, however, hiring enough instructional staff to monitor MOOCs and provide support to students is cost-prohibitive. Unfortunately, creating MOOCs that are purely self-guided is often the price of doing business in this space. The benefit of scalability comes at the cost of instructor presence. This lack of instructor presence in *Our Earth’s Future* may have affected attrition, engagement, persistence, and completion in the present study.

Much has been researched and written about persistence in MOOCs, but not all of the findings are observed in the present study. For example, Evans, Baker, and Dee (2015) found that completing a pre-course survey was among the biggest predictors of course completion, with students who completed the survey three times more likely to complete the MOOC than those who did not complete the survey. Moreover, these findings were observed in a STEM MOOC. An additional regression analysis conducted on the dataset for the present research study did not replicate this finding\(^\text{12}\). This underscores the fact that, even within the domain of MOOC research, samples and courses are varied and inconsistencies abound.

Finally, and perhaps most importantly, there is a debate among researchers about how best to conceptualize and study MOOCs. Some people suggest that findings from MOOC research can be applied to traditional campus-based college courses because the format of a MOOC closely resembles that of a face-to-face course (Seaton, Bergner, Chuang, Mitros, & Pritchard, 2014). Others, however, posit that the MOOC experience is so different from traditional courses that reconceptualization of key variables is necessary in order to accurately

---

\(^{12}\)  \(R^2 = .008, F(1,397) = 3.059, p = .081\)
and effectively study learning in this space (DeBoer, Ho, Stump, & Breslow, 2014). See Figure 9 for a helpful illustrated comparison of the structures of traditional courses and MOOCs.

**Figure 9.** The universal, linear structure of a traditional course compared to the unique and varied structure of a MOOC. Reprinted from *Changing "course": Reconceptualizing educational variables for massive open online courses*, by J. DeBoer, A. D. Ho, G. S. Stump, and L. Breslow, 2014, *Educational Researcher, 43*(2), p. 76. Copyright 2014 SAGE Publications.

The assertion that new conceptualizations are required for MOOC research was first made in 2014, just several years after the rise of MOOC popularity. DeBoer, Ho, Stump, and Breslow (2014) declared that, given the degree to which the personalized progression through a MOOC varies from that of a traditional course, applying existing metrics to new (MOOC) contexts risks the misinterpretation of findings. To mitigate this risk, they offered a recontextualization of four variables from traditional higher educational scholarship that are often applied to MOOC research.
The first variable is **enrollment**, which is traditionally a measurement of the number of students who register to complete a course and the deadline by which they must do so. The re-contextualized version of this variable measures differentiated user intention. For example, rather than count all of the tens of thousands of enrollments equally, it is helpful, for research purposes at least, to differentiate enrollments based upon the degree to which a student is committed to the course (enrollees who are just browsing, auditing, intending to complete, etc.).

The second variable is **participation**, which is traditionally a measurement of student activity. The re-contextualized version of this variable measures MOOC participation – in all of its forms – as an outcome rather than a predictor. This recognizes that, for some students, participation *is* the goal.

The third variable is **curriculum**, which is traditionally measured and operationalized as either a sequence of coursework or a prescribed pathway toward a degree. The re-contextualized version of this variable acknowledges that MOOC students personalize the course experience by choosing the content they will consume and the order in which they will do so. (This personalized course experience is a hallmark of the MOOC platform, and the very thing that makes MOOC research equally challenging and exciting.)

Finally, the fourth variable is **achievement**, which is traditionally measured as a final grade. This variable can be re-contextualized to measure achievement of *student* goals, rather than achievement of the instructor’s goals for the student (DeBoer et al., 2014). For example, if a student’s goal is to learn about climate modeling and they take and pass just the module three test that deals with that specific content, then that should be measured as achievement, not attrition.
Many researchers and practitioners agree with this need to re-contextualize traditional variables for MOOC research. Whereas DeBoer et al. (2014) focus on course structure, others focus on students, suggesting that MOOC students are so different from traditional students that we as researchers must treat them differently. This assertion is based upon the fact that traditional students in traditional courses are extrinsically motivated to complete assignments and succeed in the course so that they can earn a passing grade. This is not the case for MOOC students, who typically enroll in a course for a variety of informal reasons, including browsing to find a new topic of interest, exploring an area of current interest, or seeking specific information about a given subject. Most MOOC students “are more curious than serious” (Seaton et al., 2014, p. 61). In the absence of the extrinsic motivation that exists in formal education settings, attrition prevails. Because the motivation is completely different for traditional versus MOOC students, we should not apply traditional metrics and terminology to MOOC populations (Williams et al., 2018).

These perspectives begin to shed light on the surprising outcomes of this study by underscoring the fact that traditional educational findings do not often generalize to MOOCs because MOOC students and their learning experiences are vastly different from traditional formal education populations. These foundational contextual differences may explain some of the findings from the present study.

I posit that new contexts are required for some MOOC research, particularly around issues of enrollment, intention, persistence, completion, and the diverse demographic make-up of the population. Although these variables are different enough to require re-contextualization for MOOC research, other traditional variables transfer easily to MOOC investigations and do not require recontextualization. The present study exemplifies this duality.
On the one hand, note the differences. We saw during the data preparation process that the sample required segmentation by behavior patterns in order to identify the participants to include in the analysis. We also saw that the high attrition rate exceeded acceptable research standards while being completely representative of MOOCs. As noted previously, attrition within MOOCs is infamously high – approximately 92 – 97% (Hew & Cheung, 2014; Williams, Stafford, Corliss, & Reilly, 2018). By comparison, the freshmen retention rate in traditional schools of higher education is 81%, and the six-year graduation rate is 60% (Undergraduate Retention and Graduation Rates, 2018). Additionally, including course completion in the study may have been misguided given that the majority of people who enroll in a given MOOC do not intend to complete it (Gütl, Rizzardini, Chang, & Morales, 2014). These and other data suggest that students who choose to enroll in informal educational MOOCs are differently motivated than their counterparts in traditional formal education settings. Considering these issues, re-contextualizing variables is reasonable.

On the other hand, note the similarities. There are variables used in this study that transfer quite nicely from traditional courses to MOOCs, enabling accurate comparisons between the two populations. For example, if we are willing to equate the sub-sample of participants in the present study who completed all five modules with students in a face-to-face course (where attrition is generally low and completion is generally high), then we can see that new variables and metrics are not needed and that research comparisons can be made. In this particular case, the pre-test findings observed in research studies with traditional populations are also observed with the MOOC students in the present study; that is, for both traditional and online students who complete a course, those who take a pre-test before a unit of instruction learn more than those who do not. And among those who take pre-tests, there is a benefit of distributed practice
on post-test scores. These types of comparisons and analyses are not only possible, but they are also valuable in both research and practical contexts.

The present study illustrates that this need not be a debate over which metrics and/or terminology to use in MOOC research. It is not a case of either/or; rather, it is a case of both/and. Allowing for both methodologies presents us with an opportunity to focus on the exciting business of asking new questions of this fascinating population. Instead of needlessly investigating the number of students who complete a MOOC, we should investigate how to support the learning of all MOOC students – whether they participate for a day, a week, or the entirety of the course.

Despite the conversation about the need for new measurements and vocabulary, and despite the potential challenges of conducting research in MOOCs, platforms like Coursera remain a robust and compelling option for testing innovative teaching and learning strategies. Few other systems provide an opportunity for multivariate random experiments with people from around the world, allowing educational psychologists to learn more about the effects of teaching strategies, culture, language, geographic region, and international lived experience on how people learn (Williams et al., 2018). The MOOC study herein described is just one example of the research opportunities that exist in this domain.

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 how the results of the present study can be implemented by instructional designers.

Before addressing the practical implications of this study, let us reiterate the findings: (1) among all users in a MOOC, pre-tests and feedback do not affect learning outcomes; (2) the presence of pre-tests significantly and negatively affects persistence and completion, deterring some participants from progressing through the course; (3) among those who do persist and complete the course, those who take pre-tests achieve higher learning outcomes than those who do not; and (4) among those who take pre-tests, there is a positive, cumulative effect of persistence (module completion) on learning outcomes.

We have articulated possible explanations for these findings and situated them within the unique educational landscape of MOOCs. Now we turn our attention to a different question: What do these findings mean for instructional designers and MOOC developers? This section will examine the practical implications of the present study. These include: creating pre-tests without feedback; fostering motivation to increase course engagement and persistence; and experimenting with adaptive course designs.

First, until we know more about the effects of feedback on automated pre-tests with this population, it is best to focus on the benefits of pre-tests without feedback. 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.

Second, we know that pre-tests were effective for participants who completed the course, yet also negatively associated with course persistence and completion. We also know that, among those who submitted pre-tests, those who submitted all five pre-tests achieved higher learning outcomes 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 no easy task, as MOOCs are an informal, often-times inconsequential educational resource that have very low stakes. Indeed, data collected on pre-course surveys across the AMNH MOOCs indicates that just 31.4% of participants intend to complete the course in which they enrolled. Given both this 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 contradictory 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 the MOOC experience. This could be achieved by awarding digital badges for completing tests, or by implementing individual test leaderboards or a course-level leaderboard (Gené, Núñez, & Blanco, 2014). Offering micro-credentials for courses might also support student motivation.
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 *Our Earth’s Future*, pre-tests were not factored into the course grade. If they had contributed to just a small percentage of the final grade, more students might have completed them. Though these strategies might incentivize more people to progress through more of the material, they would not necessarily lead to less attrition.

The last practical consideration is larger in scope, and 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. Coursera CEO Jeff Maggioncalda recently acknowledged that “different learners need different solutions” (Maggioncalda, 2019). Support for this idea can be found in a study of data from a psychology MOOC offered on the Coursera platform (Koedinger, McLaughlin, Kim, Jia, & Bier, 2015). The goal of the study was to identify variations in course engagement patterns among groups of students in order to predict achievement outcomes. Notably, the course did not rely solely on the content capabilities of the Coursera platform, which were limited to videos, text/essays, tests, and peer-reviewed essay assignments. Rather, the course included interactive learning modules that were embedded via Learning Tools Interoperability (LTI) functionality, allowing students to enhance their learning with a series of content-related digital activities. Findings from this study indicate that the limited interactive engagement that leads to passive (less effective) learning can be combatted by embedding interactive learning modules to support the video lectures, texts, and quizzes that are the core features of MOOC platforms. Those students who engaged with
embedded activities in addition to watching the course lecture videos achieved higher learning outcomes than their peers who did not complete the interactivities (Koedinger et al., 2015).

Embedding digital activities into MOOCs is not the only way to create a more interactive and less prescriptive learning experience for students. Adaptive learning, or the use of intelligent technology to create personalized learning pathways for individual students, could be a wonderful 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 align with MOOC user engagement patterns and were present, to some degree, in the Spanish custom MOOC platform. The six areas of adaptive learning include: (1) course material that is not necessarily presented linearly, but rather becomes available to students based on their expressed interests and choices; (2) course material that is not available on a pre-determined schedule, but rather is available according to the students’ individual schedules; (3) students’ ability to select their desired level of difficulty for the course material; (4) creating 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. After completing a pre-course questionnaire, a custom version of a course could be immediately created and presented to a newly enrolled student based upon their learning preferences (which already dictate a MOOC learner’s engagement and progress in a course). Customizations could include: specific sub-topics, content types (essays vs. videos vs. audio files), assessment types (if any), and forums (if any). 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 custom and adaptive components may lead to the equity and achievement outcomes that have long been the promise of this technology.

**Limitations**

The research study described herein is not without its challenges or limitations. These are primarily due to the highly variable nature of the MOOC population. For example, the first limitation of conducting the study within a MOOC is attrition. As shown in Table 3 (page 49),
two of the measures had considerable attrition rates, particularly the measure related to student self-efficacy. This high attrition rate is an unfortunate yet predictable side-effect of conducting research in a MOOC, where the overall attrition rate is 92 – 97% (Hew & Cheung, 2014; Williams, Stafford, Corliss, & Reilly, 2018).

Related to the challenge of attrition is the low response rate on the pre-course self-efficacy survey. Only 80 participants submitted the survey from January - December of 2018, and only 50 of the 80 participants who completed the pre-course survey actually went on to submit pre- and post-tests. Of those 50 participants, two were minors whose data was excluded from the study. This means that just 12.03% of participants submitted both survey and test responses; there was no self-efficacy data for 87.97% of the sample.

The survey response rate might have been higher had the survey been embedded directly into the course instead of built in Survey Monkey and sent to participants via a link in the welcome email and shared on the course home page. However, this functionality (embedding external items into a Coursera course) was not available at the time the four experimental versions of the course were created. An alternative to embedding a survey in a course is to append the four self-efficacy measures to the end of the module one post-test. The potential decline in module one post-test submissions would likely be off-set by an exponential increase in collected self-efficacy data.

A third challenge is the variability of the experiences and environments of the participants. One goal of the study was to include adult participants from around the world. Though this goal was achieved, it introduced the possibility of confounding variables. There is so much that we do not know about MOOC learners. For example, some participants may have had to overcome a language barrier; in a course chock full of scientific terms, this could not have
been easy, and it might have affected learning outcomes, engagement, motivation, and/or persistence. Some participants may have been exposed to climate science in previous educational settings, giving them an advantage over their peers. Still other participants may have struggled to understand climate science within the context of competing cultural and/or religious interpretations.

Having a limited understanding of participants’ prior experiences was compounded by the fact that conducting educational research online is a challenge unto itself. It was not possible to identify or control the conditions in which participants engaged with the course. Some participants may have struggled with slow internet speeds, or with completing the course on a mobile device with just a cellular connection. Some may have been in a quiet place with no interruptions while they took the pre- and post-tests. Others might have been in a public place with lots of distractions. Some may have used Google or Wikipedia to support their progress in the course. The only assured consistency among the sample is that, unlike participants in a face-to-face experiment, their experiences during their participation in the study were varied and could not be controlled.

Another limitation of this study is the inability to determine whether or not participants reviewed course content prior to submitting a pre-test. As noted earlier, it is possible that some participants may have read essays or watched videos and then gone on to take 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 — the digital version of an open-book test. Educational research on a robust MOOC platform like Coursera is easy to scale because of its many millions of participants. In exchange for scale and methodology benefits
such as random assignment, researchers sacrifice the ability to more carefully manage the participants in their samples.

Related to the management of the participant experience, there was no way to control the time a user spent between the pre- and post-tests. The range of patterns is fascinating. Some participants took the post-tests 30-60 minutes after taking pre-tests, strongly suggesting that they proceeded through the course in a linear fashion. 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. And some participants tried to “game the system” by taking the post-tests first and then going back to take the pre-tests. (As previously noted, these scores were removed from the dataset.)

This inconsistency in the time spent between taking the pre- and post-tests is not something that can be controlled in the Coursera platform. Although time between tests cannot be controlled, enforcing sequence within the platform can be controlled. It is possible to require a student to review and/or complete particular course content before progressing to the next course element. This feature was not enabled for this experiment because it is not part of the AMNH Coursera instructional design plan.

Despite these challenges and limitations, this study is a valuable and unique contribution to the domains of assessment, feedback, and educational technology. The findings of this study have important implications for both research and practice. Though we now have new information about pre-tests and feedback with MOOC students, more questions remain to be explored in future studies.
Directions for Future Research

Future research can expand on this study in three ways: first, by working with the current dataset to conduct a conceptual change study and to understand the effects of multiple pre- and post-test submissions on learning outcomes, persistence, and completion; second, by modifying the research design to minimize attrition and create a more prescribed learner pathway; and third, by replicating the study in order to understand if the present findings are subject-specific, domain-specific, and/or generalizable to other populations.

Using the current dataset. While preparing the data for analysis, it quickly became clear that many users took the pre- and post-tests multiple times. The current study only included participants’ first test submissions in order to minimize any effect of taking the same pre- and post-tests multiple times (a threat to internal validity). A future study could extend the examination of persistence and achievement by analyzing whether or not there is (a) a relation between condition and number of post-test attempts and/or (b) a relation between post-test scores and number of post-test attempts.

In addition to working with data from multiple submissions of post-tests, it would also be useful to study multiple submissions of pre-tests. Analyzing the scores of participants’ second pre-test attempts by condition would provide information about which type of feedback, if any, is optimal when the same test is administered multiple times. A study like this would be especially useful for MOOC developers because MOOCs are often designed to allow for multiple test submissions.

The existing dataset could also be used to conduct a conceptual change study. To do this, the course author could identify common misconceptions that people have within the domain of climate change science. We could then use the existing pre- and post-test responses to
understand (a) the proportion of participants who believe climate science misconceptions and (b) whether or not their beliefs in those misconceptions change as they proceed through the course and correct their misunderstandings.

**Modifying the design.** There are several ways in which future research should incorporate modifications of the present study’s research design. As mentioned previously, the first way to modify the design is by enabling sequential progress throughout the course. In this way, users would be technologically forced 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 in order.

An additional modification to the research design is to adjust the administration of the pre-course surveys. First, it would be beneficial to embed the pre-course self-efficacy survey directly into Coursera to increase the response rate of participants who submit the survey. A second adjustment would be to ask participants to re-take the self-efficacy survey questions at the end of the course. By measuring the change in people’s climate change self-efficacy over time, we would begin to understand if participation in a course like *Our Earth’s Future* encourages people’s beliefs in their own abilities in a given domain. This methodology modification would treat self-efficacy as both a predictor and an outcome.

One final design modification 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 that all of the research questions – especially those related to self-efficacy – could be answered with a large sample size. Attrition would likely be much lower in a higher-stakes course. 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 populations are very different.

**Replication.** 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. 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. AMNH has an online graduate program for teachers – *Seminars on Science* – that would be an ideal candidate for this type of replication study.
In conclusion, future research should address four areas: (1) the effects of multiple pre- and post-test submissions on learning outcomes, persistence, and completion; (2) conceptual change outcomes; (3) the minimization of attrition and control of the sequence in which course content is accessed; and (4) an understanding of whether or not the present findings are subject-specific, domain-specific, and/or generalizable to other populations. These four areas can be studied by using the current dataset, modifying the research design, and replicating the study. Conducting research in any of these areas would both extend the impact of the present study and contribute to the knowledge base of best educational practices for teaching and learning with an online adult population. It is a goal of both the American Museum of Natural History and Coursera to continue to grow this body of work.

Summary

The goals of this study were: (1) to examine the effects of pre-tests and feedback on learning outcomes, persistence, and course completion for non-traditional adult students in an informal online science course; and (2) to identify links among self-efficacy and outcome measures. Conducting this study in a massive open online course on the Coursera platform afforded the ability to create a multivariate experiment with random assignment. Using a course created by the American Museum of Natural History afforded access to a highly qualified climate scientist who created the pre- and post-test measures. This study is the first of its kind regarding experimental feedback conditions with pre-tests in automated MOOC assessments.

The first goal of the study was achieved, but the findings were not anticipated. In summary: (1) among all users in a MOOC, pre-tests and feedback do not affect learning outcomes; (2) the presence of pre-tests significantly and negatively affects persistence and completion, deterring some participants from progressing through the course; (3) among those
who do persist and complete the course, those who take pre-tests achieve higher learning outcomes than those who do not; and (4) among those who take pre-tests, there is a positive, cumulative effect of persistence (module completion) on learning outcomes. Unfortunately, the second goal of the study was not satisfactorily achieved due to a low response rate on the pre-course survey.

Based on the previous research that has been conducted in these areas, it was expected that participants in the three feedback groups would have significantly higher post-test scores than the control group, and that those who received elaborate feedback would have significantly higher post-test scores than those in the other two treatment groups. The findings of the present study did not align with these expectations. Possible explanations include limited prior knowledge, elements of the test structure, participants’ receptivity and responses to feedback, and the lack of generalizability to traditional educational populations. Any one of these factors, or a combination thereof, could have accounted for these unexpected findings.

Despite the unanticipated outcomes, these research 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. As a first-of-its kind study, this research lays a solid foundation for future inquiry into the areas of pre-tests and feedback delivered online to adults. Additionally, it is my hope – and the hope of my 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.
Appendices
Appendix A

Pre-tests

Pre-test Module One

Question 1: Why are the tropics warmer than the poles all year long?
   1. The angle of sunlight is more direct in the tropics.
      a. Option 1 feedback: Correct!
   2. The equator is closer to the sun, so more sunlight hits the tropics.
      a. Option 2 feedback: The size of the Earth is so small compared to the Sun-Earth distance that there is no appreciable difference.
   3. Higher albedo at the poles means they absorb less sunlight.
      a. Option 3 feedback: This is a true statement, but doesn't answer the question. The cooling due to higher polar albedo is a smaller effect than the tilt of the Earth.
   4. The atmosphere is higher at the tropics than the poles, so there are more greenhouse gasses to trap outgoing heat.
      a. Option 4 feedback: It is true that the atmosphere is higher at the tropics, but that doesn't influence the greenhouse effect.

Question 2: How do greenhouse gasses warm the Earth?
   1. They absorb and re-radiate thermal (longwave) energy from the Earth.
      a. Option 1 feedback: Correct!
   2. They allow more energy to enter than to exit the atmosphere.
      a. Option 2 feedback: Greenhouse gasses have no effect on incoming radiation, but increase the near-surface temperature. This allows the incoming and outgoing energy fluxes to remain approximately equal.
   3. They reflect thermal radiation back to the Earth.
      a. Option 3 feedback: They don't reflect thermal radiation; they absorb and re-emit it equally in all directions (towards the Earth and out to space)
   4. They absorb and re-radiate shortwave energy from the Sun.
      a. Option 4 feedback: Greenhouse gasses do not interact with incoming shortwave radiation.

Question 3: Which of the following is a correct scientific statement about global warming?
   1. Global warming is the result of increased atmospheric greenhouse gasses.
      a. Option 1 feedback: Correct!
   2. Global warming is bad because it is causing many animals to become extinct.
      a. Option 2 feedback: The second half of the statement is true: many animals are going extinct because of global warming. But "good" and "bad" in the ethical sense are not scientific descriptors, so this is not a correct scientific statement.
   3. Global warming is caused by variations in the solar cycle.
a. Option 3 feedback: This statement is scientific, but not correct. The variability in the solar cycle does not have a large impact on climate, and the observed variability in the Sun does not match the observed variability in the climate system.

4. Global warming can't be proven true, so it should not be accepted as truth.
a. Option 4 feedback: In the strictest sense, nothing can ever be proven true. (For example, you can't prove that we're NOT living the Matrix). Proving things false, not true, is the basis of science.

Question 4: The amount of carbon dioxide in the atmosphere:
1. changes with the seasonal cycle, but is increasing overall
   a. Option 1 feedback: Correct!
2. has been steadily increasing since The Industrial Revolution
   a. Option 2 feedback: This is true of the yearly averages, but there are still seasonal cycles in the amount of atmospheric carbon dioxide.
3. changes with the seasonal cycle, but is steady overall
   a. Option 3 feedback: There is an overall increase in addition to changes in the seasonal cycle.
4. varies widely by location, so there isn't a large-scale trend
   a. Option 4 feedback: Although there is some local variability, both the seasonal cycle and the increase in the average can be seen in measurements from around the world.

Question 5: 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: 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.
7. 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.
8. 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.
Pre-test Module Two

Question 1: Which of the following contributes to sea level rise?
1. melting glaciers
   a. Option 1 feedback: Correct!
2. melting sea ice
   a. Option 2 feedback: Sea ice is just frozen seawater, so there's no change in sea level when it forms or melts. You can see the same thing with a melting ice cube in a glass of water; as the ice melts, the water level stays constant.
3. melting glaciers and melting sea ice
   a. Option 3 feedback: Melting glaciers add mass to the ocean, so they increase the sea level. However, sea ice is just frozen seawater, so there's no change in sea level when it forms or melts. You can see the same thing with a melting ice cube in a glass of water; as the ice melts, the water level stays constant.
4. neither melting glaciers nor melting sea ice
   a. Option 4 feedback: Melting glaciers add mass to the ocean, so they do increase sea level. Changes in sea ice do not.

Question 2: On average, how does water in the ocean move?
1. Warm water moves from the tropics to the poles.
   a. Option 1 feedback: Correct!
2. Cold water sinks throughout the ocean, leaving warmer water to rise.
   a. Option 2 feedback: The sinking of cold water is very important near the poles, but it doesn't happen many places in the ocean. There is much more horizontal motion than vertical motion.
3. Water is mostly moved by the wind.
   a. Option 3 feedback: Although wind-driven circulation is important near the surface, the wind can't reach deeper layers of the ocean.
4. Warm water rises, displacing colder water.
   a. Option 4 feedback: The ocean is heated from the top, so there's nowhere for the warm water to rise (other than evaporating into the atmosphere).

Question 3: Which is true of the annual cycle of Arctic sea ice?
1. Since ice melts during the summer, it reaches a minimum in early fall.
   a. Option 1 feedback: Correct!
2. The total amount of sea ice stays nearly constant throughout the year.
   a. Option 2 feedback: Sea ice grows in the winter as the surface of the ocean freezes, and retreats in the summer as it melts.
3. Sea ice grows as glaciers break off into the ocean.
a. Option 3 feedback: Although this sounds plausible, sea ice is made of frozen ocean water, and glaciers are made of snow. When glaciers break off into the ocean, they can remain as icebergs for awhile and then they melt.

4. Nearly all the ice melts each summer and refreezes in the winter.
   a. Option 4 feedback: Sea ice can survive for multiple years, so there is a mix of older and younger ice. Older sea ice is generally stronger and thicker than younger sea ice.

Question 4: Which of the following happens when water is heated?
   1. Large amounts of heat lead to small changes in temperature.
      a. Option 1 feedback: Correct!
   2. Large amounts of heat lead to large changes in temperature.
      a. Option 2 feedback: It takes a lot of heat to change the temperature of water, which is why ocean temperatures are warmer at the end of the summer than at the beginning.
   3. It evaporates quickly.
      a. Option 3 feedback: It takes a lot of heat to change the temperature of water, and even more to change its state (i.e. from liquid to gas). That's why it takes so long to boil a pot of water.
   4. It changes density rather than temperature.
      a. Option 4 feedback: Under most conditions, heating leads to changes in both temperature and density.

Question 5: Which of the following contains the most carbon?
   1. the deep ocean (below 1000 m)
      a. Option 1 feedback: Correct!
   2. the surface layer of the ocean
      a. Option 2 feedback: The surface layer of the ocean contains around 1000 gigatons of carbon; the deep ocean contains more than 30,000 gigatons.
   3. the atmosphere
      a. Option 3 feedback: The atmosphere contains around 800 gigatons of carbon. This is slightly less than the surface ocean (1000 gigatons) and a lot less than the deep ocean (30,000+ gigatons).
   4. plants and soils
      a. Option 4 feedback: Plants (500 gigatons of carbon) and soils (2000+ gigatons of carbon) combined still hold much less carbon than the deep ocean (30,000+ gigatons).
Pre-test Module Three

Question 1: Which of the following natural events leads to a decrease in temperature?

1. volcanic eruptions
   a. Option 1 feedback: Correct!
2. melting glaciers
   a. Option 2 feedback: Glaciers may melt in response to a decrease in temperature, but they don't cause it.
3. earthquakes
   a. Option 3 feedback: Earthquakes do not affect temperature.
4. increased cloud cover
   a. Option 4 feedback: Clouds can warm OR cool the atmosphere depending on when and where they occur.

Question 2: The African drought of the 1980s (made famous to many outside of the continent by the song "We Are the World" by Michael Jackson and Lionel Richie) was most directly linked to which of the following?

1. changes in ocean temperatures
   a. Option 1 feedback: Correct!
2. poor land management
   a. Option 2 feedback: Overgrazing and over-farming may have exacerbated the problem, but it was set in motion by changes in ocean temperatures.
3. changes in atmospheric temperature
   a. Option 3 feedback: Although atmospheric temperatures did change somewhat during that time, the changes in ocean temperatures were the driving force behind those changes and the drought.
4. high levels of dust in the atmosphere
   a. Option 4 feedback: Increased dust from the Sahel region of Africa was a result of the drought, not a cause.

Question 3: Which of the following is NOT a way for scientists to test the accuracy of climate models?

1. comparing model predictions with daily weather
   a. Option 1 feedback: Correct!
2. checking predictions made right after specific events such as volcanic eruptions with observed results
   a. Option 2 feedback: The eruption of Mt. Pinatubo in 1991 provided a great opportunity for climate scientists to test the accuracy of their predictions (which were quite accurate).
3. looking at large-scale structures like hurricanes that are not coded for directly by the model
a. Option 3 feedback: Climate models contain the basic rules of physics, but there is no computer code to tell them that hurricanes are round, or what the overall temperature structure of the atmosphere should look like, or where rainfall should occur.

4. comparing the results from different models, or multiple runs of the same model
   a. Option 4 feedback: It would be unwise to trust the results of just a single model run, since small changes in the model setup can have big changes in the output. Climate predictions are based on ensembles of multiple model runs.

Question 4: Which of the following is an example of a positive feedback?

1. After an unusually cold winter, larger sections of the ocean will freeze to sea ice, leading to increased reflection of sunlight and colder temperatures.
   a. Option 1 feedback: Correct!

2. Due to high levels of greenhouse gases, some people are choosing to drive less frequently.
   a. Option 2 feedback: This is a negative feedback because the initial change (higher levels of greenhouse gases) leads to changes in the opposite direction (less frequent driving will produce lower levels of greenhouse gases). In this context, "positive" and "negative" refer only to the direction of the change, and are not a value judgement.

3. A decrease in land ice leads to an increase in sea level.
   a. Option 3 feedback: This isn't a feedback, but simply cause and effect. If mass is moved from land to ocean, sea level must rise.

4. As carbon dioxide increases in the atmosphere, the ocean can absorb less of it.
   a. Option 4 feedback: This is true, but not a direct part of a feedback loop. The ocean is absorbing less carbon dioxide over time because it is becoming saturated and because warm water holds less carbon dioxide than cold water does.

Question 5: Which of the following forces act from outside the climate system and force it to respond with some internal change?

1. an increase in solar output
   a. Option 1 feedback: Correct! There is an 11-year solar cycle that forces small changes in the climate system, and some longer-term variability that is also extremely important.

2. an increase in atmospheric water vapor
   a. Option 2 feedback: This change occurs within the climate system, since the water has just moved from one place (like the ocean) to another (the atmosphere). It might be driven by an outside forcing, or it might be part of a feedback loop within the system.

3. a change in the speed of atmospheric currents
a. Option 3 feedback: This is a feedback within the system, not an external forcing. Atmospheric currents are driven by temperature differences, which are themselves the result of the way sunlight hits the Earth.

4. a decrease in ocean salinity
   a. Option 4 feedback: This change occurs within the climate system, since the salt has just moved from one place (rocks) to another (the ocean). It might be driven by an outside forcing, or it might be part of a feedback loop within the system.
Pre-test Module Four

Question 1: How has global sea level changed in the 20th and 21st centuries?
1. Sea level rose in the 20th century; it continues to rise at a higher rate in the 21st century.
   a. Option 1 feedback: Correct!
2. Sea level didn't change in the 20th century; it is now rising in the 21st century.
   a. Option 2 feedback: Sea level rose in the 20th century; it continues to rise at a higher rate in the 21st century.
3. Sea level rose in the 20th century; it is rising at a lower rate in the 21st century.
   a. Option 3 feedback: Sea level rose in the 20th century; it continues to rise at a higher rate in the 21st century.
4. Sea level fell in the 20th century; it is now rising in the 21st century.
   a. Option 4 feedback: Sea level rose in the 20th century; it continues to rise at a higher rate in the 21st century.

Question 2: How are the ice sheets covering Greenland and Antarctica changing?
1. Both Greenland and Antarctica are losing ice.
   a. Option 1 feedback: Correct!
2. Greenland is losing ice while Antarctica is gaining ice.
   a. Option 2 feedback: Both Greenland and Antarctica are losing ice, but the rate is slower for Antarctica.
3. Greenland is gaining ice while Antarctica is losing ice.
   a. Option 3 feedback: Both Greenland and Antarctica are losing ice.
4. Neither Greenland nor Antarctica are losing ice.
   a. Option 4 feedback: Both Greenland and Antarctica are losing ice.

Question 3: How can global warming impact food security?
1. Changes in large scale circulation patterns bring drought to some locations.
   a. Option 1 feedback: This is true, but it is not the only true statement in the list of answer options.
2. Warmer temperatures lead to increased evaporation, which dries out soils.
   a. Option 2 feedback: This is true, but it is not the only true statement in the list of answer options.
3. Warmer temperatures reduce the yield of grain.
   a. Option 3 feedback: This is true, but it is not the only true statement in the list of answer options.
4. All of the above.
   a. Option 4 feedback: Correct!

Question 4: How can large-scale global warming affect individual animal species' habitats?
1. Their prime habitat will move towards the poles, leaving less territory.
   a. Option 1 feedback: Correct!
2. Their prime habitat will move towards the poles, where cooler temperatures will allow them to thrive.
   a. Option 2 feedback: As animals move to follow their habitat, they are likely to find the areas of cooler temperatures either at elevations that are too high or in places already occupied by humans.
3. Their prime habitat will move towards the equator, where warmer temperatures will put them at risk.
   a. Option 3 feedback: As animals move to follow their habitats, they are likely to find the areas of cooler temperatures either at elevations that are too high or in places already occupied by humans.
4. Their prime habitat will move towards higher elevations, where cooler temperatures are available.
   a. Option 4 feedback: As animals move to follow their habitat, they are likely to find the areas of cooler temperatures either at elevations that are too high or in places already occupied by humans.

Question 5: How has the Earth's orbit changed over time?
1. The shape of the orbit, the tilt of the planet, and the direction of the tilt all go through different cycles, resulting in a complex pattern of changes.
   a. Option 1 feedback: Correct!
2. The Earth's orbit does not change significantly.
   a. Option 2 feedback: The shape of the orbit varies in a 100,000-year cycle. The tilt of the planet varies in a 23,000-year cycle. And the direction of the tilt varies in a 23,000-year cycle. These cycles combine to produce a complex pattern.
3. Changes in the Earth's orbit lead to changes in tilt as well.
   a. Option 3 feedback: The shape of the orbit varies in a 100,000-year cycle. The tilt of the planet varies in a 23,000-year cycle. And the direction of the tilt varies in a 23,000-year cycle. These cycles combine to produce a complex pattern.
4. The Earth's orbit becomes slower over time, but doesn't change shape.
   a. Option 4 feedback: The shape of the orbit varies in a 100,000-year cycle. The tilt of the planet varies in a 23,000-year cycle. And the direction of the tilt varies in a 23,000-year cycle. These cycles combine to produce a complex pattern.
Pre-test Module Five

Question 1: What is the largest contributor to uncertainty in climate models?

1. uncertainty about future greenhouse gas emissions
   a. Option 1 feedback: Correct!

2. uncertainty about the parameterization of clouds
   a. Option 2 feedback: Although there are still uncertainties in the science (especially clouds), the largest contributor to uncertainty is not knowing how humans will behave.

3. uncertainty in the ability of models to accurately respond to changes
   a. Option 3 feedback: Models have proven robust at responding to external forcings. Although some scientific uncertainty remains, the largest contributor to uncertainty is not knowing how humans will behave.

4. uncertainty about the physics of the climate system
   a. Option 4 feedback: Although there are still uncertainties in the dynamics of climate, the largest contributor to uncertainty is not knowing how humans will behave.

Question 2: Which of the following is NOT considered when calculating the risk of a large flood?

1. the last year in which a large flood occurred
   a. Option 1 feedback: Correct!

2. the probability of sea level rise in the area
   a. Option 2 feedback: Risk is probability scaled by cost, so all probabilities and costs are important. The last year in which a large flood occurs doesn't matter, although terms like "fifty-year flood" give the misleading impression that floods happen on a schedule.

3. the costs of various mitigation strategies
   a. Option 3 feedback: Risk is probability scaled by cost, so all probabilities and costs are important. The last year in which a large flood occurs doesn't matter, although terms like "fifty-year flood" give the misleading impression that floods happen on a schedule.

4. the costs associated with a particular event occurring
   a. Option 4 feedback: Risk is probability scaled by cost, so all probabilities and costs are important. The last year in which a large flood occurs doesn't matter, although terms like "fifty-year flood" give the misleading impression that floods happen on a schedule.

Question 3: What would most likely happen to the climate if all greenhouse gas emissions were stopped today?

1. The temperature would continue to rise for several decades, then stabilize.
   a. Option 1 feedback: Correct!
2. The temperature would stabilize within a year or so.
   a. Option 2 feedback: It takes more than a few years for temperature to stabilize. This is called "warming in the pipeline," and occurs because atmospheric temperature takes time to adjust to carbon dioxide levels.
3. The temperature would decrease.
   a. Option 3 feedback: Temperature would not decrease immediately because atmospheric temperature takes time to adjust to carbon dioxide levels.
4. The temperature would continue to increase because we passed a tipping point.
   a. Option 4 feedback: While this is possible, it is unlikely.

Question 4: How is global warming affecting the impact of hurricanes?
1. The impact of hurricanes is larger because global warming has led to an increase in sea levels.
   a. Option 1 feedback: Correct!
2. Hurricanes are unrelated to global warming.
   a. Option 2 feedback: The relationship between hurricanes themselves and global warming is still uncertain, but there is a definite effect on hurricane impacts. If the average water level is higher, it takes less additional water to cause flooding.
3. Hurricanes are stronger due to global warming.
   a. Option 3 feedback: The relationship between hurricanes themselves and global warming is still uncertain, but there is a definite effect on hurricane impacts. If the average water level is higher, it takes less additional water to cause flooding.
4. The impact of hurricanes is unchanged because global warming has led to a decrease in coastal populations.
   a. Option 4 feedback: Despite the risks, coastal populations are increasing. The relationship between hurricanes themselves and global warming is still uncertain, but there is a definite effect on hurricane impacts. If the average water level is higher, it takes less additional water to cause flooding.

Question 5: Which of the following is an example of adaptation to global warming?
1. restoring coastal ecosystems (i.e. oysters, coral, or mangroves) to reduce flooding
   a. Option 1 feedback: Correct!
2. decreasing greenhouse gas emissions
   a. option 2 feedback: This is a strategy for mitigating global warming, not adapting to the consequences.
3. analyzing tail risks
   a. Option 3 feedback: Although necessary, analysis itself is not an adaptation. Adaptations are changes made to reduce the impact of global warming.
4. improving storm prediction
a. Option 4 feedback: Although this might help us respond, predictions themselves are not adaptations. Adaptations are changes made to reduce the impact of global warming.
Appendix B

Post-tests

Post-test Module One

Question 1: A scientific statement is one that:
   1. can be proven to be correct by existing research
   2. has been published and widely accepted by most scientists
   3. is well defined, logical, and can be proven true
   4. can be proven to be false

Question 2: How do greenhouse gas molecules in the atmosphere warm the surface of Earth?
   1. They allow only thermal radiation from the Sun to pass through.
   2. They absorb all the thermal radiation from the Sun, then radiate that heat onto Earth’s surface.
   3. They reflect thermal radiation back into space.
   4. They absorb thermal radiation from the Earth and reemit it, warming the atmosphere.

Question 3: The Keeling Curve shows that:
   1. Carbon dioxide levels in the atmosphere rise and fall from one year to the next, but overall they are slowly increasing.
   2. Carbon dioxide levels in the atmosphere vary with the seasons, but remain consistent overall.
   3. Carbon dioxide levels in the atmosphere vary with the seasons, but are steadily increasing overall.
   4. Carbon dioxide levels in the atmosphere vary with the seasons, but are steadily decreasing overall.

Question 4: Think about the definition of a scientific statement. Which of these is not a scientific statement?
   1. The Himalayan Mountains are made entirely of granite.
   2. Isaac Newton was the most important scientist of the 17th century.
   3. All roses have thorns.
   4. No plants eat meat.

Question 5: What is the main reason that, year-round, the tropics are warmer than polar regions?
   1. The equator is closer to the sun, and since energy decreases with distance, sunlight is stronger in the tropics.
   2. The atmosphere in the tropics contains higher levels of greenhouse gases, trapping more heat energy and warming the air.
   3. Sunlight strikes Earth more directly in the tropics and at more of an oblique angle in the polar regions, so sunlight is more concentrated in the tropics.
   4. Tropical rainforests are darker and absorb lots of sunlight, or heat energy, while the bright, snow-covered polar regions reflect sunlight.
Question 6: Which of these is NOT a greenhouse gas?
1. carbon dioxide
2. nitrogen
3. methane
4. water vapor

Question 7: Which of these statements describe a location’s climate?
5. Temperatures in Portland, Oregon are expected to rise this week.
6. Last year Denver, Colorado received more snowfall in January than in February.
7. Average temperatures in July are about 15 degrees warmer in Los Angeles than they are in San Francisco.
8. The air pressure in Miami drops significantly before the arrival of a hurricane.

Question 8: When a system is in energetic equilibrium, it is:
1. staying at a constant temperature by absorbing and releasing the same amount of energy
2. staying at a constant temperature by not exchanging energy with its surroundings
3. changing temperature depending on the nature of the equilibrium
4. changing temperature to achieve a balanced state

Question 9: What is the importance of the Greenland Ice Core Project (GISP2)?
1. It provides a detailed history of the world's climate for past hundred thousand years.
2. It provides evidence that can be combined with other data sources to calculate the climate of the past hundred thousand years.
3. It proves that crystal structure is important to paleoclimate.
4. It proves that chemistry is part of climate change.

Question 10: Which statement about longwave and shortwave radiation is correct?
1. Shortwave radiation is emitted by the Sun; longwave radiation is emitted by the Earth.
2. Shortwave radiation is emitted by the Sun; longwave radiation is reflected by the Earth.
3. Shortwave radiation is primarily absorbed in the atmosphere; longwave radiation is primarily absorbed by the Earth.
4. Shortwave radiation is primarily absorbed by the Earth; longwave radiation is primarily reflected by the atmosphere.
Post-test Module Two

Question 1: How do Earth’s oceans affect the atmosphere?
1. The ocean keeps the atmosphere warm during the summer and cold in the winter.
2. As water flows towards the tropics and warms, it absorbs carbon dioxide from the atmosphere.
3. **Water flowing from the tropics to the poles heats the atmosphere above.**
4. Earth’s oceans have no effect on the atmosphere.

Question 2: How has sea ice in the Arctic Ocean changed in recent decades?
1. There are no longer extreme seasonal variations in sea ice extent.
2. **The extent of September sea ice continues to get smaller.**
3. Sea ice is becoming older, since most of the young ice is melting.
4. The ocean’s albedo has increased, so more sunlight is reflected.

Question 3: Which event has the greatest impact on rising sea levels?
1. the freezing of ocean water
2. the formation of glaciers
3. the melting of sea ice
4. **the melting of glaciers**

Question 4: Which of the following statements about the ocean and atmosphere are NOT true?
1. The ocean holds nearly all the climate system’s heat energy—far more than the atmosphere.
2. The ocean holds more carbon dioxide than the atmosphere.
3. The ocean affects the climate system on longer time scales than the atmosphere.
4. **Ocean currents move faster than air currents.**

Question 5: Which of these statements about glaciers is true?
1. **They form on land from a buildup of snow.**
2. The more they melt, the more slowly they move.
3. They form when the surface of ocean water freezes.
4. They are slightly saltier than sea ice.

Question 6: Which statement describes the large-scale movement of water in the oceans?
1. Deep ocean currents are driven by wind.
2. Surface currents are primarily driven by the changing density of water.
3. In general, warm water moves toward the tropics and cool water moves toward the poles.
4. **In general, warm water moves from the tropics to the poles.**

Question 7: How does salinity affect ocean circulation?
1. **Highly saline water helps drive deep ocean circulation.**
2. The effect of salinity depends on the water temperature.
3. Salinity has different effects in the northern and southern hemispheres.
4. Salinity is determined by ocean circulation.
Question 8: Ocean gyres are approximately circular because:

1. The path of the water is determined both by density differences and the rotation of the Earth.
2. The path of the water is primarily determined by the wind.
3. The ocean basins themselves are round.
4. The rotation of the Earth pushes fluids in a circle.

Question 9: Water has a high specific heat. This is important to the climate system because:

1. Small temperature changes are associated with large changes in heat.
2. Large temperature changes are associated with large changes in heat.
3. Heat is liberated when liquid water changes to ice.
4. The ocean is an enormous reservoir of heat.

Question 10: Which of the following is NOT true about the layered structure of the ocean?

1. The mixed layer brings carbon dioxide from the deep zone to the atmosphere.
2. The mixed layer contains less carbon than the deep zone does.
3. The mixed layer is driven by the wind while the deep zone is driven by density differences.
4. The mixed layer is functions as a boundary between the deep zone and the
Post-test Module Three

Question 1: Based on climate models and observations, which of these factors can actually cause a decrease in temperatures?

1. volcanic eruptions
2. increased solar output
3. the melting of sea ice
4. an increase in Arctic vegetation

Question 2: Besides temperature and rainfall, which of these factors can be calculated by climate models?

1. cloud distributions
2. ocean currents
3. the extent of snow cover
4. all of the above

Question 3: Earth’s climate system is dominated by positive feedbacks, which means that:

1. Small changes can result in increasingly larger effects.
2. Most changes lead to an increase in global temperatures.
3. The climate system always returns to equilibrium.
4. Most changes have a beneficial effect on life on Earth.

Question 4: How did the late-20th-century droughts in the African Sahel help scientists test their climate models?

1. Scientists used climate models to predict the length and severity of the drought before it happened.
2. Scientists were able to enter historic forcing data and simulate the event.
3. The models were not helpful.
4. Even when scientists entered data, it could not simulate the event.

Question 5: Using climate models with different historical forcings, scientists determined that the Sahel droughts were most closely linked to:

1. variations in ocean temperatures
2. overgrazing and overfarming
3. periods of extended rainfall in the 1920s and 1950s
4. Models did not reveal a strong connection between drought and a single historic forcing.

Question 6: What happens to an area of open ocean and ice if you have an unusually cold winter?

1. Temperatures will remain constant because the amount of ice and water reflecting and absorbing sunlight would stay the same.
2. Temperatures will continue to fall because more water will freeze, directly cooling the air around it.
3. Temperatures will eventually rise because more water will freeze, which will reflect more sunlight and warm the air around it.
4. Temperatures will continue to fall because more water will freeze, more sunlight will be reflected, the air will cool, and the cycle will continue.
Question 7: What is something that global climate models CANNOT do?
1. Calculate temperature, pressure, winds, humidity, and salinity of oceans.
2. Reproduce large-scale phenomena that cannot be explained by physics on a smaller scale.
3. Simulate variables of climate in different layers of the atmosphere.
4. **Forecast changes in local weather.**

Question 8: Which of the following statements about forcings is correct?
1. Human-induced forcings always have a warming influence.
2. Human-induced forcings always have a cooling influence.
3. Temperature and humidity are examples of forcings.
4. **Solar output is a variable forcing (one that can change).**

Question 9: Which of the following statements about greenhouse gases is not true?
1. Greenhouse gases are forcings that contribute to changes in global climate.
2. Greenhouse gases act as positive feedbacks, amplifying the initial changes.
3. **Carbon dioxide levels increase during the summer and decrease in the winter.**
4. Methane emissions from wetlands increase with temperature.

Question 10: Large scale features of climate are considered emergent phenomena because:
1. They arise spontaneously from the basic physics in the model.
2. They can only be modeled using parameterizations.
3. They exist on multiple spatial scales.
4. They are functions of small-scale physics.
Post-test Module Four

Question 1: Based on computer models, how do scientists expect warmer temperatures to affect wolverines in the Pacific Northwest?

1. Their prime habitat will move northward, where cooler temperatures will threaten their populations.
2. Their prime habitat will move southward, where warmer temperatures will threaten their populations.
3. Their prime habitat will shift to lower elevations, where they will come in contact with human populations.
4. Their prime habitat will move northward, and they will have less territory.

Question 2: How did the 2003 heat wave in France affect crops?

1. Decreased rainfall resulted in increased crop yield.
2. Due to the planting of heat-resistant crops, the food supply was unaffected.
3. Farms had to be shifted northward to escape rising temperatures.
4. The regional crop yield was decreased by as much as 30 percent.

Question 3: How have sea levels changed over the past century?

1. They rose and fell each decade during the twentieth century, but have only increased during the twenty-first century.
2. They have been rising at a steady rate over the past century.
3. They have been rising, but the rate has begun to accelerate in recent decades.
4. After decades of steady sea levels, they have begun to rise in recent decades.

Question 4: To protect wildlife, some scientists propose connecting national parks and other protected habitats with areas where both wildlife and people can live. What benefit would this provide for the wildlife?

1. Animals would have more space to travel safely in their habitat.
2. The animals’ prime habitat would be protected from the impact of climate change.
3. Communities would help feed and care for wildlife.
4. Humans would see first-hand the impact of climate change and work to reduce emissions.

Question 5: Which of these factors is NOT a direct cause of sea level rise in the 21st century?

1. melting of mountain glaciers and permafrost
2. thermal expansion of the ocean due to warming
3. melting of sea ice in the arctic
4. melting of Antarctic ice sheet

Question 6: Why are climate scientists interested in the Milankovitch cycles?

1. Scientists have learned that small changes in solar radiation have minimal effects on Earth’s climate due to the absence of feedback loops.
2. They are concerned that Milankovitch cycles could happen more frequently in the future, enhancing climate change already progress.
3. A Milankovitch period is expected to occur in the next century, bringing a mini ice age that could counter the global warming underway.
4. The effects of past Milankovitch cycles provide evidence that small forcings can have a large effect on the climate system if there are feedback loops.

Question 7: Which of the following is NOT true about the Greenland Ice Sheet?
1. It is expected to melt within the next century.
2. The rate of ice loss is accelerating.
3. It contains enough water to raise sea level by 7 feet.
4. It is vulnerable to small changes in water temperature.

Question 8: Why are Pacific Island communities particularly vulnerable to climate change?
1. Many of the islands are flat and nearly at sea level.
2. Saltwater inundation can make it difficult to grow crops.
3. Atolls are vulnerable to changes in ocean chemistry.
4. all of the above

Question 9: Which of the following is expected to occur regarding droughts and floods?
1. Some areas will have increased flooding, while others will have increased droughts.
2. As drought risk increases, flood risk decreases.
3. Areas with mountain glaciers are vulnerable to floods but not droughts.
4. Soil moisture will remain high.

Question 10: Why is the West Antarctic Ice Sheet an area of particular concern for climate scientists?
1. It has a large area in contact with warming ocean water.
2. It is the largest land glacier.
3. It is losing more ice than the Greenland glacier.
4. all of the above
Post-test Module Five

Question 1: Each of the following farms is located in a flood plain. If a risk assessment were done for each farm, which one would be described as having a high probability, but a lower consequence?

1. a farm in a narrow valley that floods once a century
2. **a farm on a wide valley that floods once a decade**
3. a farm in a narrow valley that floods once a decade
4. a farm on a wide valley that floods once a century

Question 2: One way to plan a strategy for the future is to apply the concept of risk. Which aspect of an event is NOT used when calculating the risk?

1. the probability of an event occurring
2. the consequences if the event occurs
3. the costs (such as financial loss or number of deaths) of the strategy to mitigate the event
4. **the need to convince the public of taking mitigating actions to prepare for an event**

Question 3: What is meant by a “tipping point”?

1. a point at which the climate transitions very rapidly, causing abrupt environmental changes
2. a point at which sudden, large-scale forcings cause sudden changes in the climate
3. a point at which the climate reverses its direction after long periods of gradual changes
4. a point at which the global mean temperature will continue to rise and cause permanent destruction to Earth’s environments

Question 4: Which of the following describes “tail risk” as it relates to the climate system?

1. “Tail risk” refers to an outcome that cannot happen.
2. “Tail risk” refers to an outcome that can only happen in the distant future.
3. **“Tail risk” refers to an extremely unlikely outcome that is characteristic of a system that has feedbacks.**
4. “Tail risk” refers to a likely but catastrophic outcome.

Question 5: Which statement describes the likelihood of global temperatures exceeding 2 degrees Celsius above pre-industrial levels before the end of the century?

1. The more we increase global emissions, the lower the probability of exceeding the 2 degrees Celsius.
2. **The less we increase global emissions, the lower the probability of exceeding 2 degrees Celsius.**
3. Global emissions do not affect the probability of global temperatures exceeding 2 degrees Celsius above preindustrial levels.
4. The more we cut global emissions, the higher the probability of exceeding 2 degrees Celsius above preindustrial levels.

Question 6: Which statement does NOT describe most scientists’ predictions based on climate models?
1. If we stopped all greenhouse gas emissions now, we would see an immediate decrease in the atmospheric temperature.
2. Small changes in sea level and temperatures will increase the frequency of extreme events.
3. If the current rate of emissions continues, sea level rise will accelerate.
4. If the current rate of emissions continues, temperature rise will accelerate.

Question 7: What lessons were learned in New York City after Hurricane Sandy?
1. The impact of hurricanes is limited to coastal communities.
2. Hurricanes are caused by climate change.
3. **Rising sea levels, due to human-induced climate change, can increase the effect of hurricanes.**
4. It is impossible to predict hurricanes, so we shouldn’t worry about them.

Question 8: What is the largest contributor to uncertainty in climate models?
1. **Uncertainty about future greenhouse gas emissions**
2. Uncertainty about the physics of the climate system
3. Uncertainty in the parameterization of clouds
4. Uncertainty in the validity of modeling

Question 9: Which of the following is a long-term adaptation to sea level rise?
1. **Restoring oysters along coastlines**
2. Using surge barriers during periods of increased flooding
3. Decreasing emissions
4. Installing flood walls

Question 10: Why are heat waves a significant concern even with only a few degrees of warming?
1. An increase in nighttime temperatures puts increased stress on the human body.
2. They are expected to increase in severity as well as frequency.
3. They increase the demand on power grids, leading to increased GHG emissions.
4. **All of the above**
Appendix C

Pre-course Survey

1. How do you feel about learning? Select the oval that corresponds to the degree with which you agree or disagree with each statement.

<table>
<thead>
<tr>
<th>Statement</th>
<th>disagree strongly</th>
<th>disagree a little</th>
<th>no opinion</th>
<th>agree a little</th>
<th>agree strongly</th>
</tr>
</thead>
<tbody>
<tr>
<td>I am confident I can complete this course.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I can explain the evidence for climate change.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I can explain the difference between climate and weather.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I can explain the effect of greenhouse gas emissions on temperature.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I can define the term &quot;scientific statement.&quot;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I like receiving feedback on my work.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I use feedback to improve my work.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I don't read feedback.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feedback on quizzes is not helpful to me.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I have difficulty getting started on a task.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I am efficient and get things done.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I am persistent, working until the task is finished.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2. Sex:
   - female
   - male
   - other
   - decline to answer

3. With which ethnic group do you identify?
   - American Indian or Alaskan Native
   - Asian
   - Black or African American
   - Hispanic of any race
   - Native Hawaiian or other Pacific Islander
   - White
   - Race and ethnicity unknown
   - Two or more races
   - Other
Decline to answer

4. In what country do you live?
   - (dropdown list of all countries)

5. What is the highest level of academic study you have completed?
   - High school diploma / US GED
   - Associate’s Degree
   - Bachelor’s Degree
   - Master’s Degree
   - Professional or Doctoral Degree (Ed.D., J.D., M.D., Ph.D.)
   - does not apply

6. Are you 18 years of age or older?
   - yes
   - no
Appendix D

SQL Queries

SQL Query to Aggregate Test Data

SELECT assessment_actions.amnh_user_id,
amnh_course_user_ids.earth_climate_change_user_id,
course_branches.authoring_course_branch_name,
course_branch_items.course_branch_item_name,
assessment_questions.assessment_question_prompt,
assessment_responses.assessment_response_score,
assessment_actions.assessment_action_ts
FROM assessment_responses,
assessment_actions,
course_branch_items,
course_branch_item_assessments,
assessment_questions,
course_branches,
amnh_course_user_ids
WHERE assessment_responses.assessment_action_id = assessment_actions.assessment_action_id
AND assessment_questions.assessment_question_id =
assessment_responses.assessment_question_id
AND course_branch_item_assessments.assessment_id = assessment_actions.assessment_id
AND (course_branch_item_assessments.course_item_id =
course_branch_items.course_item_id AND
course_branch_item_assessments.course_branch_id =
course_branch_items.course_branch_id)
AND course_branches.course_branch_id = course_branch_items.course_branch_id
AND amnh_course_user_ids.amnh_user_id = assessment_actions.amnh_user_id
AND assessment_action_ts > '2018-01-08 00:00:00.0'
SQL Query to Aggregate Coursera Demographic Data

SELECT
assessment_actions.amnh_user_id,
amnh_course_user_ids.earth_climate_change_user_id,
course_branches.authoring_course_branch_name,
demographic_questions.question_desc,
demographic_choices.choice_desc,
demographic_answers.answer_int
FROM
assessment_responses,
assessment_actions,
course_branch_items,
course_branch_item_assessments,
assessment_questions,
course_branches,
amnh_course_user_ids,
demographic_questions,
demographic_answers,
demographic_choices
WHERE
assessment_responses.assessment_action_id = assessment_actions.assessment_action_id
AND assessment_questions.assessment_question_id =
assessment_responses.assessment_question_id
AND course_branch_item_assessments.assessment_id = assessment_actions.assessment_id
AND (course_branch_item_assessments.course_item_id =
course_branch_items.course_item_id
AND course_branch_item_assessments.course_branch_id =
course_branch_items.course_branch_id)
AND course_branches.course_branch_id = course_branch_items.course_branch_id
AND amnh_course_user_ids.amnh_user_id = assessment_actions.amnh_user_id
AND assessment_actions.amnh_user_id = demographic_answers.amnh_user_id
AND demographic_questions.question_id = demographic_answers.question_id
AND demographic_answers.choice_id = demographic_choices.choice_id
AND demographic_answers.question_id = demographic_choices.question_id
AND demographic_questions.question_id = demographic_choices.question_id

AND assessment_action_ts > '2018-01-08 00:00:00.0'
References


participants’ learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest, 14*(1), 4–58.


