ADDRESSING THE GENDER GAP IN STEM MOOCS: HOW BRIEF, IN-COURSE MESSAGES CAN INCREASE FEMALES' MOTIVATION AND ONLINE LEARNING SUCCESS

by

Alexandra D. Urban

A dissertation submitted to Johns Hopkins University in conformity with the requirements for the degree of Doctor of Education

Baltimore, Maryland

July 2022

Abstract

Women are routinely underrepresented in higher education science, technology, engineering, and mathematics (STEM). Despite lower admission barriers and more flexible schedules, massive open online courses (MOOCs) are not exempt from this trend, as female learners are less likely than males to enroll in and complete them. Using the ecological systems framework, this researcher explored factors contributing to the STEM gender gap across home country, course, and learner levels. During the mixed-methods needs assessment, female learners' self-efficacy, time constraints, and home country's level of gender inequality emerged as the three strongest factors related to the gender gap in retention and completion. The self-determination theory of intrinsic motivation was used to align the novel intervention prompts with females' sense of competency, autonomy, and relatedness. The intervention study deployed text-based messages to students in 150 STEM MOOCs to tackle the primary identified needs: boost confidence, improve planning, and emphasize individuals' values to counteract gender inequality. The Coursera platform was used to conduct a randomized controlled trial (RCT) experiment, allowing causal quantitative data analysis to assess the impact of these intervention groups on learners' persistence, skill development, and continued learning. Females' self-reported reasons for stopping before completion were also qualitatively coded by theme. This explanatory mixed-methods RCT study included 324,457 total active learners with identified gender. The four treatment groups (three variant types plus the combined treatment) each resulted in a significant increase in first-week completion rates for female learners compared with the control. The value relevance treatment group retained this significant increase, successfully eliminating the gender gap in STEM MOOC course completion. The self-efficacy treatment significantly raised the number of female course completers by 50% in the youngest age tier. Moving all active learners in this RCT from the control group to the value relevance treatment would result in approximately 1,400 additional female STEM course completers. Implications for future research and practice are explored, including the personalized deployment of the messages given differences in impact by age, gender, and geography.

Keywords: STEM, MOOC, gender gap, motivation, completion, randomized experiment

JOHNS HOPKINS
SCHOOL *of* EDUCATION

Dissertation Approval Form

Student's Name: Alexandra Urban   Date: May 25, 2022

Dissertation Title:

Addressing the Gender Gap in STEM MOOCs

The student has made all necessary revisions, and we have read, and approve this dissertation for submission to the Johns Hopkins Sheridan Libraries as partial fulfillment of the requirements for the Doctor of Education degree.

Linda Lohr		7/7/2022
Adviser	Signature	Date
Alexandra Murtaugh		5/31/2022
Committee Member	Signature	Date
Ranjini M JohnBull		5/31/2022
Committee Member	Signature	Date

Dedication

This research is dedicated to the millions of women worldwide trying to better their lives and their families' future by learning new skills online. May we bring the same passion and courage that they bring to our platform. It is an honor and a responsibility to help their dreams become their reality.

Acknowledgments

This dissertation grew from the encouragement of many individuals across time, space, and hundreds of video-conference meetings.

To the professors: My advisor, Dr. Linda Lohr, offered her continual care, support, and good humor throughout this long journey. Dr. Ranji JohnBull inspired me with her unrelenting passion for education research, and Dr. Alexandra Murtaugh lit up about the statistical analysis as much I did. These three committee members brought such kindness and thoughtful ideas, for which I am exceedingly grateful.

To the 2019 cohort of EdD students: I could not have been gifted a better group of fellow students. Thank you to my “Dream Team” of Mary Fran, Lisa, Crystal, Alex, and Tara. You five infused these past three years with more light and laughter than I could have imagined. You are the change-making soul sisters I had long desired.

To my Coursera colleagues: This research would not have been possible without the flexibility and support from Linlin and Denis, the data science expertise of Jaya, Alan, and Juhi, the legal input of Bill and Brianna, and the product expertise of Polina, Stanley, and Tyler.

To my family: Parents, you have instilled a fierce love of learning, curiosity to explore the unknown, and joy for analyzing new data. Your enthusiasm for what I want to pursue is rare and unrivaled. You are my pillars of love and strength.

Throughout this pandemic, I have been grateful for the solid roof over my head, my tiny desk where I did the majority of this writing, and our pup, Winnie, who sat at my feet during the late nights and early mornings, growing from a newborn into an adult dog during this program.

Finally, huge thanks to my fiancé, Andrew, who always knew I could do it. Your certitude strengthens my whole world.

Table of Contents

List of Tables

List of Figures

Executive Summary

Female learners are consistently less likely to pursue and persevere in higher education science, technology, engineering, and mathematics (STEM) courses (Hughes et al., 2017; Russell, 2017; Wang & Degol, 2017). Massive open online courses (MOOCs) demonstrate this same pattern, even though anyone located anywhere can participate in these courses (Allione & Stein, 2016; Grella & Meinel, 2016; Healy, 2017). This gender gap results in thousands of females on the sidelines while MOOC completers experience greater skill development, career advancement, and salary increases (Hadavand et al., 2018). In addition to individuals and their families, entire countries have decreased technical innovation and entrepreneurship activity because of the lack of females persisting in STEM careers (Beede et al., 2011; Dilli & Westerhuis, 2018).

Gender Gap in STEM MOOCs: A Problem of Practice

This work focuses on learners taking MOOCs on the Coursera platform, one of the world's largest providers of open-access online courses. With more than 100 million registered learners globally, Coursera offers a diverse environment for an improved understanding of the STEM gender gap and the deployment of automated interventions to assist learners (Coursera, 2022b). Building on others' previous experimentation in MOOCs (Kizilcec, Saltarelli, et al., 2017; Yeomans & Reich, 2017), this researcher's first goal was to reveal key drivers of females' lower persistence in STEM MOOCs on the Coursera platform before designing novel intervention prompts to combat these challenges.

At the national level, lower gender equality corresponds to a lower likelihood for females to enroll and complete STEM courses because of stronger negative stereotypes, less familial support, and limited disposable income (Guiso et al., 2008; Perez, 2019). Within the home and work spheres, full-time employment is a larger risk factor for stopping an online course for females than males (Allione & Stein, 2016). This finding is likely the result of time constraints in addition to their work schedule, such as home, family, and childcare responsibilities, all unpaid work completed most often by females in nearly all regions worldwide (Perez, 2019). Female retention in STEM courses may also be affected by the level of female representation in their instructors and fellow students, which varies depending on the course in which they enroll. This gender representation may help increase support and minimize stereotype threats, thereby helping female learners succeed (Carrell et al., 2013; Healy, 2017).

The content difficulty of a course may also influence females' likelihood of enrolling and persisting because of necessary prerequisite knowledge (Grella & Meinel, 2016) and students' varying confidence (Healy, 2017; Lambert, 2020). In addition to lower individual self-confidence (Lambert, 2020), consistently lower average self-efficacy (Handoko et al., 2019), science identity (Healy, 2017), sense of belonging (Walton et al., 2015), and goal orientation (Handoko et al., 2019) are also linked to female learners' lower persistence in STEM courses. These individual factors likely interact and are interrelated to each learner's home, work, family, and cultural context (Grella & Meinel, 2016; Guiso et al., 2008).

Needs Assessment Study

The researcher centered the analysis on a large set of quantitative data collected directly from the Coursera platform, including learners' enrollment, progression, and completion metrics, alongside available demographic information on gender and employment status. At the course level, instructor gender, average course completion time, percentage of enrolled learners who were female, and content difficulty level were all included in the analysis. The United Nations' Gender Inequality Index (UN GII), which provides a quantitative summary of each country's female-to-male inequality level, was also included. In addition, the researcher qualitatively coded female learners' reasons for stopping a course before finishing, as recorded in Coursera's pre-existing Inactivity Survey.

This needs assessment study was guided by the following research questions:

RQ1. To what extent is there a gender disparity in STEM MOOC enrollment and completion on the Coursera platform?

RQ2. How do a STEM MOOC's characteristics, such as the average time needed to complete, content difficulty level, peer representation, and instructor gender, relate to female learners' enrollment and completion rates?

RQ3. How do female learners' characteristics, such as their home country's national gender inequality level and individual employment status, relate to their enrollment and completion rates in STEM MOOCs?

RQ4. What reasons do female learners provide for dropping out of STEM MOOCs before course completion, and how often are other time demands, low confidence, and a lack of prerequisite knowledge cited?

Data were collected across 2,300 STEM MOOCs on the Coursera platform, including a sample of more than five million unique learners. Quantitative analysis was used to answer RQ1 RQ2, and RQ3, while qualitative methods were utilized to answer RQ4. Overall, a convergent parallel mixed-methods design was chosen to analyze data from numeric and textual data sources.

On the Coursera platform from March 1, 2016, to March 1, 2020, females comprised 23% of active enrollments and had a completion rate 30% lower than their male peers. This finding aligns with previous research on the gender difference in MOOC completion rates (Crues et al., 2018; Kizilcec & Cohen, 2017).

Overall, three key barriers emerged affecting females' likelihood to persist in STEM MOOCs on the Coursera platform. The time needed to complete a course ($R^2 = 0.25$) and full-time employment status as a larger barrier for females than males ($t = 499.4, p < 0.001$) demonstrated strong quantitative impacts on female course completion. Both factors highlight the challenges of home and family responsibilities, in addition to frequently having less autonomy over their free time (Perez, 2019). Moreover, "no time" was the most cited reason (21%) by female learners ($N = 174$) for why they stopped a STEM MOOC before completion. These results regarding the importance of time issues for females are consistent with previous research (Allione & Stein, 2016; Perez, 2019).

Female learners displayed higher enrollment rates in less challenging courses but were equally likely to complete courses across difficulty levels ($p = 0.54$). Plus, females cited a lack of confidence (14%) more frequently than prerequisite knowledge (9%) as their reason for not completing a course. Given these results, weaker self-efficacy, not lack of content knowledge, appears to be the second key barrier to STEM MOOC participation. This finding aligns with previous MOOC research on females' average lower self-efficacy, a diminished belief in their abilities to accomplish the task at hand (Handoko et al., 2019). The third STEM MOOC participation barrier for females was a nation's gender inequality, which strongly correlated with female learners' likelihood to enroll in ($R^2 = .051$) and complete ($R^2 = 0.31$) STEM MOOCs (countries = 153). These three main factors interact for employed females in gender-inequal societies who are often managing stronger negative stereotypes, lower self-efficacy, and a myriad of time constraints (Guiso et al., 2008; Perez, 2019).

Intervention Study

Given the needs assessment findings, the researcher examined the existing literature to inform the design of novel in-course prompts to counteract the top three STEM MOOC persistence challenges for female learners. The intervention study was a randomized control trial (RCT) experiment conducted on the Coursera platform across 150 STEM MOOCs. From December 8, 2021, to March 20, 2022, any learner newly enrolled in these participating STEM MOOCs was randomly assigned to one of five treatment groups: control, self-efficacy boost, planning support, value relevance emphasis, or a combination of these three treatments.

Using the self-determination theory (SDT) framework, the researcher anticipated these prompts would increase motivation and further engagement with the course content by fulfilling the psychological needs of competency, autonomy, and relatedness (Deci & Ryan, 2000). Female learners often display lower levels of these three psychological need areas in STEM courses, which also links to their documented lower motivation (Murphy et al., 2019; Simon et al., 2015; Stolk, Zastavker, et al., 2018). Females' lower motivation, on average, correlates with their reduced likelihood of continuing in STEM content (León et al., 2015; Murphy et al., 2019; Simon et al., 2015). Notably, SDT-based interventions have been shown to increase females' persistence and performance in STEM courses (Dell et al., 2018; Huang & Mayer, 2019).

Two process questions were selected to assess the reach and exposure of this implemented framework. After the experiment concluded, four outcome questions were used to assess the near- and long-term impact of this in-course intervention on persistence, performance, dropout reasons, and future enrollments. The six research questions for this RCT intervention study were as follows:

RQ1. To what extent did the intervention reach the target learner group?

RQ2. To what extent did learners find the prompt helpful?

RQ3. What differences in impact did each intervention have on week-one and course completions?

RQ4. What differences in impact did each intervention have on course completers' performance and skill development?

RQ5. How did the intervention affect female learners' self-reported reasons for dropping out of the STEM MOOCs for those who did not complete?

RQ6. To what extent did the intervention spark learners to continue learning in other MOOCs?

Using an explanatory mixed-methods design, the researcher emphasized quantitative analysis to analyze this intervention's implementation and impact (Teddlie & Tashakkori, 2003). The conceptual framework, RCT design, and sample of more than 324,457 active learners permitted the use of statistical analyses to summarize the data and assess the causal impact (Shadish et al., 2002). Aggregated, anonymized data were collected directly from the Coursera platform on completions, grades, and learners' demographic information, including age, gender, employment status, and country. External rankings from the UN GII represented each nation's gender inequality level. Two-way analysis of variance (ANOVA) tests and *t*-tests were used to assess the impact on outcome metrics by intervention group and gender. Qualitative coding by theme was used to analyze the females' self-reported reasons for stopping before completing, as reported in Coursera's Inactivity Survey.

The implementation displayed stronger reach and exposure metrics than expected. Females comprised 39.6% to 39.9% across each of the five groups in the final RCT sample, a more substantial proportion of females than observed in the needs assessment. The real-time question assessing learners' perceived utility of the messages indicated that 85.6% to 91.7% of learners considered the messages helpful across the four treatments. Females displayed similar or higher helpfulness percentages across these treatment groups. This affirmation demonstrated that these novel messages were successfully implemented and indicated a strong basis for investigating the outcome metrics of interest.

Despite this intervention's brief, light-touch nature, these treatment variants significantly positively impacted female learners' retention and course completion metrics. All four treatments resulted in female first-week completion rates significantly higher than the control ($p < 0.001$). The value relevance emphasis and combined treatment groups retained their impact, resulting in higher course completion rates for females than in the control group ($p < 0.001$). No impact was observed for male learners' first-week or course completion rates across any group. The value relevance treatment raised the female course completion rate to 17.65%, statistically indistinguishable from the male completion rate of 17.77% ($p = 0.076$), successfully eliminating the gender gap in STEM MOOC completion rate in this treatment. If all the active female learners in the RCT sample had received the value relevance treatment, more than 1,400 additional female learners would have completed their STEM MOOCs than in the control.

Analyzing effects by subgroups provided additional insights into the utility of these treatments. For females working full time, the value relevance emphasis was also most effective, significantly increasing their already elevated course completion rate to 18.62% ($p < 0.001, n = 12{,}282$). Across the 82 countries with sufficient females for analysis in the RCT sample, the value relevance treatment also lowered the correlation between a country's gender inequality score and its female learners' completion rate from $\beta = -0.04$ ($R^2 = 0.02$) to $\beta = -0.007$ ($R^2 = 0.0006$). The active female learner completion rate in certain countries increased by 70%

or more, demonstrating the value relevance treatment impact compared to the control; these countries included Australia ($p < 0.001, n = 850$), China ($p < 0.001, n = 340$), Ghana ($p < 0.001, n = 149$), Guatemala ($p < 0.001, n = 67$), Ireland ($p < 0.001, n = 194$), Malaysia ($p < 0.001, n = 204$), and Saudi Arabia ($p < 0.001, n = 280$). The youngest females, aged 18 to 24, benefited most from the self-efficacy boost, increasing their completion rate from 8.97% in the control group to 13.21% ($p < 0.0001, n = 315$), resulting in 50% more young female course completers.

In females' written reasons for stopping their STEM MOOC in the experiment before completion, the researcher observed that the relative prevalence of themes differed from that in the needs assessment. Content misalignment (39%) and technical issues (22%) were the most cited ($n = 41$). Two of the most prevalent themes in the needs assessment were minimally present in this smaller sample, with "No Time" comprising only 5% of responses and "Not Confident" at 7%. While only a small proportion of those who stopped before course completion submitted the Coursera Inactivity Survey, these results show encouraging changes in the areas the intervention treatments were attempting to influence.

Skill development, as indicated by course performance, was consistent across treatments, with a mere 1% difference in average final course grades between males and females. During the experiment, male learners continued enrolling in new STEM MOOCs at a ratio of 3:1 to female learners, aligning with the enrollment gender ratio seen in the initial needs assessment. This finding was expected since the intervention's design focused on retention, not future enrollments. However, female learners in the combined treatment group displayed significantly higher average grades in their future enrollments than females in the control group ($p < 0.001$), suggesting potentially improved confidence, learning momentum, or course selection in females who received all three message types.

Implications for Research and Practice

Greater personalization and intentional deployment would amplify the impact of these interventions. Beyond the RCT, researchers and practitioners could match learners with the types of in-course messages that would provide them the greatest benefit. Given the successful increase in female first-week completion rates across all four treatments, extending the messages into later weeks of the course may also be fruitful. Finally, while not possible within this intervention study, more resource-heavy interventions, such as creating personalized schedules given learners' time demands, would be useful to explore.

In conclusion, the intervention successfully improved retention and course completion, erasing the gender gap in STEM MOOC completion rate in the value relevance treatment group. Different subgroups demonstrated the nonuniform impact of these treatments, with the self-efficacy boost most benefiting the youngest females in the sample. MOOC completers experience increases in job opportunities and salary (Hadavand et al., 2018). In the few months of this intervention, hundreds more females completed their STEM MOOC than would have if all were in the control, increasing these learners' potential for career gains and mobility. At the national level, countries observe increased technical innovations and entrepreneurship activity when more females work in STEM fields (Beede et al., 2011; Dilli & Westerhuis, 2018), highlighting the cascading benefits of helping learners complete STEM MOOCs on Coursera. Empowering females to be more successful in online science and technology courses benefits these individuals, their families, and the surrounding communities.

Chapter 1: Factors Influencing the Problem

Female scientists remain rare. Despite global advances in development and efforts to increase gender equality, women are significantly less likely to engage with and persist in science, technology, engineering, and mathematics (STEM) higher education (Buffington et al., 2016; Charles & Bradley, 2009; Hughes et al., 2017; Russell, 2017). For example, in the United States, only 36% of STEM undergraduate degrees are awarded to females, even though more women than men earn undergraduate degrees overall (National Center for Education Statistics, 2019). This discrepancy widens as graduates enter the workforce, with women holding less than 25% of STEM jobs in the United States (Beede et al., 2011). STEM fields are at the center of modern economies and produce the innovation necessary to raise living standards (OECD, 2015). Plus, diversity in STEM workplace teams increases the productivity and success of these industries (Blackburn, 2017). Thus, the persistent gender gap not only creates inequality in these professions but also impacts creativity and entrepreneurship at the national level (Dilli & Westerhuis, 2018). Successfully empowering female learners to develop their STEM skills enables individuals and economies to reach their full potential (OECD, 2015, 2019; Riegle-Crumb et al., 2012).

This chapter aims to synthesize the research on the persistent gender gap in STEM higher education. The first section outlines this problem within the context of online learning on the Coursera platform. The second section explains the theoretical framework used to structure this literature synthesis into a tiered analysis of the many relevant factors. Finally, the bulk of the chapter explores the specific factors underlying females’ decreased likelihood to engage with STEM higher education content. This overall synthesis provides the necessary foundation to investigate the specifics of this problem in the Coursera context and ultimately support female STEM learners more effectively.

Problem of Practice

Women are less likely to engage and persist in massive open online courses (MOOCs) in STEM subject areas (Alario-Hoyos et al., 2017; Crues et al., 2018; Grella & Meinel, 2016; Ihsen et al., 2015). Although not the only educational setting with persistent gender gaps in STEM, online courses offer unique advantages: they reach learners on a large scale, tailor education to different students' needs, and provide successful interventions at a lower cost than on-campus programs (Bernacki et al., 2019; Chirikov et al., 2020; Kizilcec, Saltarelli, et al., 2017; Lambert, 2020). However, before gender-related solutions can be proposed, this issue must be understood within the context of Coursera, one of the largest MOOC providers in the world.

With more than 100 million learners across 200 countries, Coursera provides diverse learners and a breadth of data to investigate this STEM gender gap on a global scale. Online learning behavior reflects the same challenges observed in on-campus programs: females are consistently less likely to start, continue, and complete STEM MOOCs on the Coursera platform than their male peers (Allione & Stein, 2016; Crues et al., 2018; Hickey et al., 2018). This problem has many facets, with enrollment, retention, and performance metrics revealing a significant gender disparity in STEM MOOCs (Alario-Hoyos et al., 2017; Brooks et al., 2018; Kizilcec & Saltarelli, 2019). A thorough investigation of the persistent STEM gender gap necessitates the parallel examination of these various outcome metrics and their underlying factors. The most prevalent factors associated with the STEM MOOC gender gap are national inequality (Kizilcec, Saltarelli, et al., 2017), female role models (Ertl et al., 2017), content difficulty (Grella & Meinel, 2016), other time demands (Eriksson et al., 2017), and learners' confidence in their abilities (Handoko et al., 2019).

Theoretical Framework

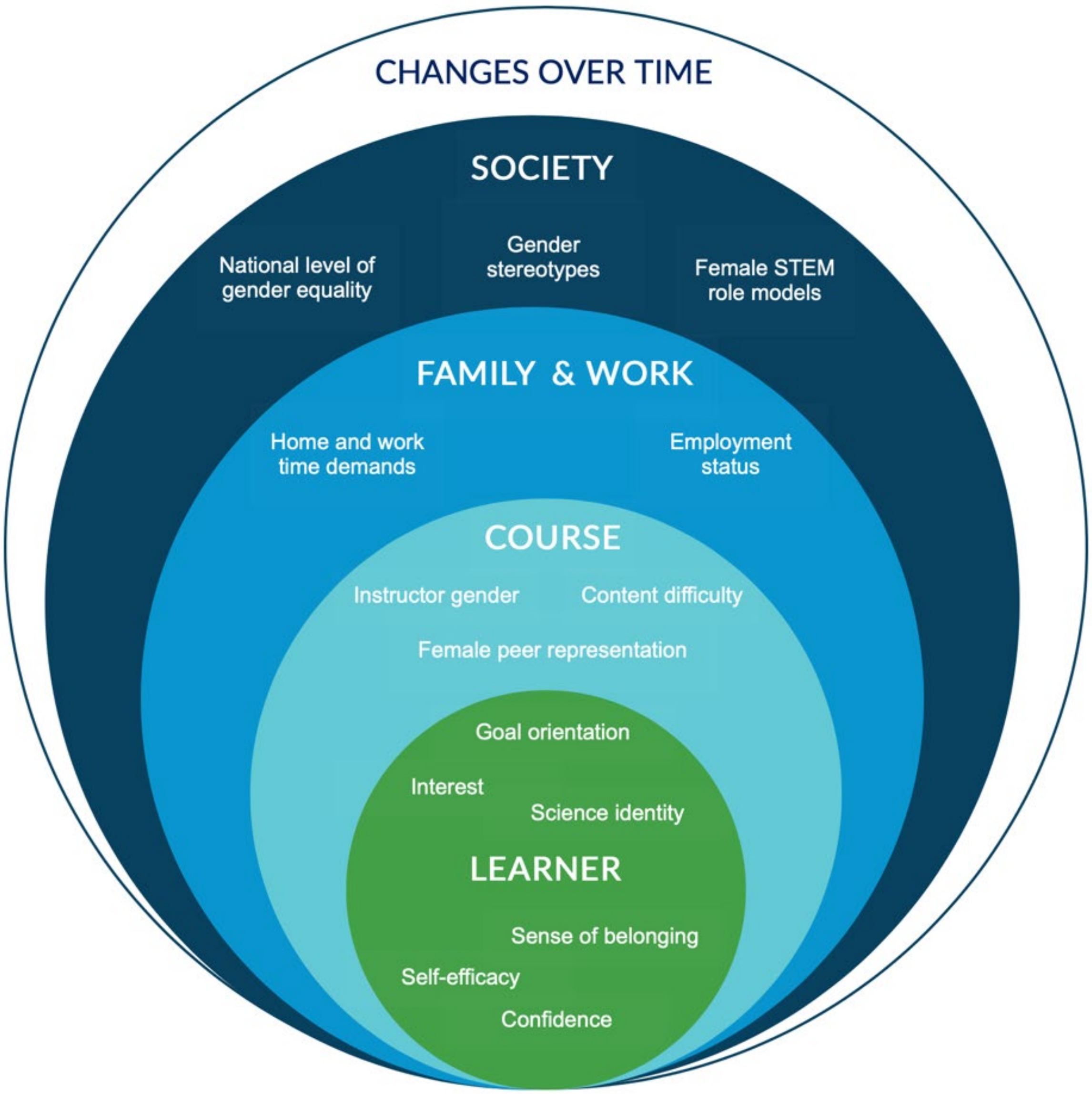
The ecological systems theory (EST), as outlined by Bronfenbrenner (1979), offers a valuable structure when considering the broad body of research on the gender gap in STEM. Several researchers have utilized a similar systems framework to guide their exploration of females' decreased likelihood to pursue STEM, with some explicitly citing Bronfenbrenner's theory (Ertl et al., 2017) and others using a nested systems approach to organize relevant factors (Blackburn, 2017; Sax et al., 2017). Across the literature, a clear pattern of factors emerges, from broad country-level variables to individual student characteristics, with each system nested inside the larger one.

While not explicitly designed to study online learning, EST is readily applicable to the MOOC environment. Just as other MOOC researchers have emphasized, the millions of learners joining these platforms represent a vast array of geographic, socioeconomic, and cultural backgrounds (Jiang et al., 2018; Kizilcec, Davis, et al., 2017). Many studies have emphasized the value of a learner-centered systems approach to investigating MOOC behavior (Wiebe et al., 2015). Some studies have even utilized the EST framework specifically to analyze learning patterns in MOOCs (Renzel & Klamma, 2013).

EST can be adapted into a structure for the levels most relevant to this research investigating STEM MOOCs. The first level explored in the literature review is the chronosystem, which examines changes over time. Second, the macrosystem encapsulates the national-level trends that may affect the student. The next level focuses on each learner's work and family life. Further nested, the next layer focuses on the course experience, examining the characteristics of the instructors, peers, and content. Both latter levels, the home life and in-course experiences, would be contained within Bronfenbrenner's (1979) microsystem; however, for clarity, it is useful to identify these sets separately when considering the diversity of MOOC learners. Finally, this literature synthesis concludes with an exploration of influences within each learner, including their self-efficacy and sense of belonging. The following sub-sections reflect the chronosystem, macrosystem, and microsystem of Bronfenbrenner's (1979) original systems theory while using titles reflecting the study's specific context. Figure 1 shows the EST model adapted to this problem of practice. Applying an adopted EST framework organizes the factors underlying the STEM gender gap to enable a more in-depth investigation and functional grouping of influence categories.

Figure 1

Theoretical Framework of Levels and Factors for the Gender Gap in STEM MOOCs



Factors Influencing the STEM Gender Gap

Changes Over Time

Although some assume the pervasive gender gap problem in STEM is improving, research has revealed a more nuanced landscape concerning trends from specific STEM disciplines. For example, the gender gap observed in the 1980s enrollment numbers for computer science (CS) degree programs in the United States was much smaller than in the early 2000s (Sax et al., 2017). Even though the absolute number of women who were attending college increased steadily from the 1980s to the early 2000s (Riegle-Crumb et al., 2012), the percentage of those beginning CS majors who are female declined from 44% in 1980 to 15% in 2011 (Sax et al., 2017). In addition, over the last 20 years, the percentage of women receiving bachelor's degrees in mathematics and statistics has declined (National Science Foundation, 2019) despite women slowly gaining a larger share of the overall college enrollment (National Center for Education Statistics, 2019).

These historical changes indicate how multiple STEM fields were, in fact, more gender-balanced several decades ago. The larger percentage of women previously attracted to and completing these STEM degree programs suggests there is no reason to believe an inherent male advantage exists. Rather, societal and personal forces have likely influenced these lower female engagement numbers over time.

Societal Factors

National Gender Inequality

Country-level influences, such as national levels of gender equality, play an essential role in STEM attainment for female learners. The research literature presents mixed effects of economic development and improved gender equality, two societal-level factors that typically increase in tandem for any given country. While some studies highlight increased STEM engagement in less-developed, less gender-equal nations (Charles & Bradley, 2009; Jiang et al., 2018), other research has also emphasized lower persistence and performance for female STEM learners in these less-developed and less gender-equal countries (Guiso et al., 2008; Kizilcec, Saltarelli, et al., 2017). Although these ideas are seemingly contradictory, these less-developed nations may be attracting more females to STEM for economic advantages but then seeing higher female dropout rates, with the degree programs and online courses unable to support these students fully. Additionally, there may be more significant barriers to achieving STEM success for women in these less-developed and less gender-equal nations, including lower disposable incomes and support from their families (Guiso et al., 2008; Kizilcec, Saltarelli, et al., 2017). This broader economic and societal context could help explain the higher female enrollments but lower retention and performance of women in these more gender-imbalanced countries.

Given the varied findings on the impact of national economic development and gender equality on female STEM engagement, it will be particularly interesting to explore these factors within courses on the Coursera platform. This analysis can draw on the variables and insights other researchers have utilized to investigate these societal factors of female STEM engagement, including economic indicators and human development metrics from the United Nations (Charles & Bradley, 2009; Kizilcec, Saltarelli, et al., 2017) as well as gender gap and equality measures from the World Economic Forum (Guiso et al., 2008; Jiang et al., 2018). These quantitative indicators may be helpful variables during the upcoming needs assessment when examining STEM MOOC enrollment and completion patterns on Coursera.

Female Role Models

Having female role models as STEM leaders in the learners' communities can help bolster young women's likelihood to persist in the sciences. Quantitative studies have shown the positive relationship between women's representation in the local STEM workforce and female science course enrollment in secondary school (Riegle-Crumb & Moore, 2014), highlighting the potential benefits of female scientist role models. Even simply reading about female scientist role models has been shown to improve individual factors for female university STEM students, such as their sense of belonging, self-efficacy, and interest in these subjects (Ertl et al., 2017; Shin et al., 2016). The visibility of women in STEM careers also relates to the overall gender stereotypes of that community and is an indicator of the broader barriers hindering females from pursuing scientific or technical fields.

Gender Stereotypes

At the national level, biases and negative stereotypes can also influence individuals' education and career choices, resulting in differences by geography and gender. Countries with more robust "science as male" stereotypes show a larger gender gap in STEM performance and persistence (OECD, 2015; Wang & Degol, 2017). This trend extends to learners in MOOCs, who are less likely to persist if joining from a country with lower levels of overall development and higher negative stereotyping for females (Kizilcec & Halawa, 2015). In the United States, these "science as male" stereotypes are a pervasive barrier to women pursuing STEM degrees (Blackburn, 2017; Hughes et al., 2017). Many countries with higher overall equality still have strong gender-STEM stereotypes, especially when males dominate the science workforce of that country (Miller et al., 2015). Thus, bias at the national level and how society views scientific roles can directly influence women's engagement in these disciplines.

These stereotypes also may help explain why different STEM sub-fields have varied female representation. Specifically, biology has much higher female student enrollment than physics and engineering, which tend to be viewed as more traditionally male fields (Cheryan et al., 2017; Master & Meltzoff, 2016). At the classroom level, female students in the most male-dominated science fields, such as engineering, have been shown to display weaker gender-STEM stereotypes than both male students and females studying the humanities, as seen in survey responses (Smeding, 2012; Smyth & Nosek, 2015). This finding suggests that women with lower individual levels of "science as male" bias relate to an increased likelihood of pursuing science themselves. In addition to broader-held beliefs about hypothetical male advantages in STEM, logistical challenges, such as greater household responsibilities often in addition to full-time jobs, also impact whether women have enough time to engage fully in online courses.

Family and Work Factors

Employment Status

One of the main reasons learners drop out of MOOCs is a lack of time (Eriksson et al., 2017; Gütl et al., 2014; Loizzo & Ertmer, 2016). Between home and job demands, women have busy schedules with limited opportunities to dive into challenging STEM curricula. This time issue is especially problematic for female learners since women do most unpaid care worldwide, impacting their ability and flexibility to spend time on personal tasks (Perez, 2019). Furthermore, full-time employment has emerged as a more significant risk factor for women dropping out of online courses than for males (Allione & Stein, 2016; Gütl et al., 2014). If a woman is already balancing a full-time job and a larger share of the unpaid housework, it is not surprising that she may ultimately drop out of the online MOOC before the course is completed.

Time Demands

Given that learners taking online courses often have full-time jobs, the addition of housework and childcare may make women unable to carve out this volunteer studying time (Allione & Stein, 2016; Perez, 2019). Even though these home and unpaid work demands likely play a significant role in women's participation in online higher education courses, they will be challenging to investigate and subsequently to change. Given the scope of this study, the researcher will identify more easily examined and actionable factors. For example, specific content characteristics, such as the average time required to complete a course, may be relevant factors to explore. Although Coursera as a platform cannot directly alter the gender imbalance of unpaid work, specific course characteristics could be designed to fit better into women's already busy lives. For example, actionable planning recommendations might help females maximize what little free time they do have.

Course Factors

Content Difficulty

Characteristics of the instructors, other students, and content design may affect how many women enroll in and complete particular STEM MOOCs. First, prerequisite knowledge, the needed skills to succeed in the current material, likely plays a role in which courses females enroll in and how likely they are to persist. Recent research has highlighted how this factor explains at least a small portion of the overall gender gap in STEM higher education (Riegle-Crumb et al., 2012; Sax et al., 2017), suggesting that prerequisites may play an essential but more minor role than sometimes believed. Researchers have noted greater female representation in beginner than intermediate and advanced online STEM MOOCs, suggesting the lack of prerequisites for introductory courses may help attract higher female enrollment (Grella &

Meinel, 2016). These enrollment patterns may also reflect learners' self-confidence and self-selection patterns, addressed later in this chapter.

Although prerequisite knowledge likely impacts the overarching gender disparity witnessed in STEM, its influence is much harder to tackle without months- or years-long interventions on remedial skills, requiring subject matter experts, additional coursework, and time from the learners themselves. However, another way to approach this issue of prerequisites is to focus on the introductory STEM MOOCs, where female learners appear more likely to congregate, and design our interventions for those courses to support the most female STEM learners. On the Coursera platform, it will be essential to validate whether these enrollment patterns by content-difficulty level are maintained. Then, these enrollment trends will help determine which courses would be best to focus on for later interventions to bolster females' participation in STEM. Overall, the skills needed to start a course likely affect female enrollment, influencing the makeup of students in different content, which, in turn, may impact female student persistence.

Female Representation

The presence of female peers may impact women's likelihood to continue within a course. Recent studies have highlighted how fellow females in the same course can motivate and support others in STEM learning (Charleston et al., 2014; Healy, 2017). Using random assignment within an undergraduate engineering program, researchers showed how female-majority groups could increase women's participation more than sex-parity and female-minority project groups (Dasgupta et al., 2015). Beyond the traditional classroom, informal learning experiences among female peers in STEM, such as hallway conversations, can mitigate attrition issues in higher education science programs (Gayles & Ampaw, 2016). Informal peer conversations can be applied to MOOCs, where discussion forums grouping women together can help them share with fellow female learners (Grella & Meinel, 2016).

Furthermore, online group work can increase women's attraction to and persistence in online courses (Bayeck, 2016). Increasing the representation of other females in a peer cohort can strengthen women's sense of belonging and science identity, which are individual-level factors explored later in this chapter (Allione & Stein, 2016; Healy, 2017; Russell, 2017). These findings demonstrate how female representation can help attract and retain women in STEM learning environments. In addition to peers, female authority figures can pave the way for more women entering and persevering in science fields.

Instructor Gender

Specifically, in the science MOOC setting, female instructors may play a valuable role in encouraging female participation. In an experimental research study, increased exposure to female instructors improved female students' academic performance and the number of credits completed within a higher education STEM degree program (Russell, 2017). Greater exposure to female instructors increased female learners' engagement in activities such as discussion posting, as confirmed through recent experimental tests on the Coursera platform (Brooks et al., 2018). Using an economic model, other researchers have found that, when high-performing female college students have their first math and science courses taught by female professors, the gender gap in performance and persistence in STEM majors disappears (Carrell et al., 2013). These findings highlight how course-level factors can directly affect relevant metrics, such as learners' progression through the material. These course-level characteristics can also indirectly affect these retention outcomes through learner-level factors, including females' self-efficacy and interest in the topic.

Learner Factors

The female learners themselves and their interests, beliefs, and actions are at the core of this study. As Wiebe and colleagues (2015) stress, new insights arise when researchers focus on the students and not only the aggregate trends of learners at the institution or course level. In this vein, this section explores the student factors influencing females' likelihood of enrolling, progressing, and completing STEM MOOCs.

Interest

Before a learner can finish an online course, she must first decide to start it, a decision often tied to learners' interests. Thus, it is imperative to explore if females are as interested in STEM content as their male peers. One hypothesis on the significantly lower enrollment numbers in STEM MOOCs for women than men is that females are simply less attracted to these more technical subject areas (Crues et al., 2018; Lambert, 2020). Despite these perceptions, white males may not be more interested in STEM subjects than females and other races (Riegle-Crumb & King, 2010). Additionally, students' interest in a topic is often influenced by other learner-level factors, such as their previous success and perceived abilities in the domain (OECD, 2015; White & Massiha, 2016).

These insights underscore how factors can indirectly affect females' participation by influencing other factors. In this case, increasing students' self-efficacy may increase their interest, likely increasing their STEM engagement and motivation to continue. Thus, the other learner-level factors explored throughout this chapter, including self-efficacy, may be even more critical for enhancing women's interest and ultimate success in STEM compared to only examining women's interest in isolation.

Science Identity

Learners' sense of their compatibility with STEM fields can also influence their likelihood of enrolling and progressing through more technical content. This concept of science identity aims to capture the extent to which an individual student's identity, including their race, gender, and age, can affect how connected they feel to a particular domain (Jones et al., 2013). Studies on this topic consistently find lower science identity for women, which often impacts the probability of their sticking with STEM throughout their higher education experience (Healy, 2017; Jones et al., 2013). This pattern of diminished science identity for females is likely intertwined with broader biases and stereotypes propagated at the societal level, as explored earlier in this chapter.

Beyond this trend of women's average lower science identity, it is equally important to examine differences among women for this factor. For example, successful women in higher education STEM degree programs often display a strong science identity (Carlone & Johnson, 2007). Others have also confirmed science identity at the individual level as a significant predictor of women's likelihood to persist in STEM higher education programs (Jones et al., 2013; Williams & George-Jackson, 2014). This variation among females' science identity levels demonstrates the importance of this factor and its positive correlation with females' retention in STEM programs. Although unclear if increasing a female's science identity may increase her likelihood of persisting in STEM content, this possibility may be worth exploring in the later exploration of intervention research literature.

Sense of Belonging

Like science identity, a learner's sense of belonging can influence her likelihood of engaging with and persisting in science-related higher education courses. More specifically, a sense of belonging is grounded in the feeling of fitting in and being accepted by the community in question (Good et al., 2012). This complex construct addresses students' perceived value by their peers and to what extent they feel a respected part of the academic group. A weakened sense of belonging is one of the largest challenges women face in STEM university programs and one of the strongest predictors of their persistence in higher education settings (Charleston et al., 2014; Cheryan et al., 2017; Good et al., 2012; Parson, 2018; Walton et al., 2015). This same predictive power for a sense of belonging on persistence has been documented in STEM MOOC settings (Kizilcec & Halawa, 2015).

One method for investigating learners' comfort and belonging in STEM MOOCs is evaluating their participation in the discussion forums. Several studies have found that females' increased activity in forums is predictive of an increased likelihood to complete the MOOC (Crues et al., 2018; Pursel et al., 2016; Qiu et al., 2016). With this act of participating in the online discussions likely relating to their comfort sharing with peers and being a valued member of the course community, these contributions can serve as a broad proxy for these learners' sense of belonging. While the specifics of this posting activity will be difficult to analyze at scale, others' previous analysis of forum behavior may indicate useful intervention paths to pursue.

Goal Orientation

In addition to females' identity alignment with the sciences and comfort in the field, learners' goal setting and active striving toward their goals can influence their likelihood of persisting in STEM MOOCs. Goal orientation, rooted in a student's mindset, comprises an individual's objectives and their alignment of actions with those goals (Wang & Baker, 2018). This construct will be beneficial to explore since it combines the setting of goals with how learners plan to achieve them. In MOOCs, more substantial goal setting at the individual level correlates with a significantly higher likelihood of completing the course (Crues et al., 2018; Wang & Baker, 2018). Other MOOC research has found that learners who set goals are more likely to finish the course successfully and achieve their own goals in applying skills gained to settings beyond the course (Handoko et al., 2019; Kizilcec, Pérez-Sanagustín, et al., 2017). This predictive power of goal setting on course completion and applying the concepts outside the course material highlights the potential benefits of having students focus on clear objectives to keep their motivation high and progress forward.

More specifically, goal orientation relates to fewer females than males persisting in scientific and technical courses online. Recent research has found that males may naturally be more likely to display goal-orientation tendencies in STEM MOOCs, further reinforcing the gender gap in course completions (Kizilcec, Pérez-Sanagustín, et al., 2017). Thus, although lower goal orientation appears to be a barrier for STEM persistence across genders, it is likely an even more vital factor for female learners. Crucially, individuals' goal orientation may be malleable.

These findings indicate how goal orientation will be a helpful construct to consider during the intervention research stage. For example, Handoko and colleagues (2019) suggest how scaffolding the process of setting goals may increase learners' likelihood to achieve their objectives, improving students' goal orientation. The Coursera platform currently features a new pilot initiative to help learners set a goal at the beginning of a course and track their progress toward meeting it, which has shown promising early results (Urban, 2019). Since women appear less likely to focus on concrete learning goals, this goal setting and tracking may be especially beneficial to female learners in STEM online content. Depending on the results of the needs assessment, the researcher may return to goal setting as a potential path for intervention.

Self-Efficacy

Another broad barrier more prevalent in female learners is weakened self-efficacy. Specifically, the multifaceted construct of self-efficacy encompasses a learner's perceived ability to execute what is necessary to succeed on a given task (Bandura, 1977b). Self-efficacy plays a crucial role in the overarching framework of triadic reciprocal determinism in Bandura's (1986) social cognitive theory, in which learning is driven by dynamic interactions among the person, their environment, and their behavior. Bandura (1977a) outlines four main sources of self-efficacy: performance accomplishments (success or failures on tasks), vicarious experience (witnessing others' task performance), verbal persuasion (encouragement or discouragement), and emotional arousal (interpreting physiological states). A student's self-efficacy will likely change as she interacts with online course work, receives feedback on her performance, and progresses successfully or not from week to week (Huang & Mayer, 2019).

Importantly, self-efficacy may not always reflect actual ability level. A meta-analysis of 1.6 million students found that females, on average, earn higher grades than males across all subjects, even in STEM, despite lower levels of reported self-efficacy (O'Dea et al., 2018). These higher grades earned by female students suggest actual ability in STEM is not likely a relevant factor of the observed gender gap and further underscore how self-efficacy may not be strictly tied to true capability. Further evidence of this disconnect between perceived and actual abilities may originate from females' stronger likelihood to attribute failure to a lack of personal ability (Murphy et al., 2019).

Females' average lower self-efficacy is a widely studied topic, especially in the sciences. Diminished self-efficacy is linked to lower enrollment, persistence, and performance for females in STEM higher education compared with their male peers (Jones et al., 2013; Macphee et al., 2013; Williams & George-Jackson, 2014). In online learning research, strong self-efficacy is a significant predictor of sustained motivation and completion in STEM MOOCs (Handoko et al., 2019; Sujatha & Kavitha, 2018; Wang & Baker, 2015). Thus, the learner-level construct of self-efficacy is likely contributing to this research's problem of practice by affecting female learners' persistence in STEM MOOCs.

Confidence

Confidence can also affect learners' persistence in more technical material and aligns closely with the construct of self-efficacy (Gayles & Ampaw, 2016; Handoko et al., 2019). Bandura (1977b) highlights how self-efficacy relates directly to learners' confidence in applying their skills. Contemporary research also makes this same explicit connection between learners' self-efficacy and confidence levels when exploring how both factors correlate with females' diminished likelihood to continue in STEM (Handoko et al., 2019; Sax et al., 2017). In a meta-analysis of 46 MOOC studies, Lambert (2020) found learners with lower confidence in their skills as one of the groups most needing assistance in STEM content and that they would most benefit from further intervention. This insight from Lambert (2020) aligns with a broader trend in secondary school STEM performance: after controlling for self-confidence in math, researchers at the OECD (2015) found the observed high school gender gap in mathematics standardized test scores disappears. In STEM MOOCs, when female students fail an assignment, they are more likely to drop out of the course than males earning the same grade, suggesting lower confidence for these female learners in their abilities (Healy, 2017; Hickey et al., 2018). Given these implications of confidence levels on female learners' likelihood to continue and succeed in STEM higher education, this factor should be further explored in the needs assessment study.

With all these factors potentially playing a role in the gender gap within STEM MOOC engagement, a conceptual framework can help reorganize these influences around the most central themes. The following section introduces a unifying conceptual framework for how these factors apply to the Coursera online learning context.

Conceptual Framework

The factors emerging from this literature synthesis reveal the crucial role of motivation when determining who engages with and persists in voluntary STEM MOOCs. Motivation is defined as the process guiding the start of and continued engagement in goal-directed actions (Schunk et al., 2020). More specifically, intrinsic motivation is deeply rooted in individuals' goals and their beliefs that the task at hand will help them reach those goals (Deci & Ryan, 2000; Sujatha & Kavitha, 2018). With MOOCs attracting learners in their free time and rarely providing academic credit, intrinsic motivation becomes a fundamental driver of learner success (Alario-Hoyos et al., 2017; Wiebe et al., 2015). Goal setting, self-efficacy, content difficulty, peer interactions, work-home balance, and stereotypes can all affect learners' motivation to engage with STEM courses (Bandura, 1977a; Ertl et al., 2017; Kizilcec & Saltarelli, 2019; OECD, 2015; Wang & Degol, 2017). These varied factors driving motivation align with the factors identified in this literature review and span the individual, household, and societal levels. Thus, learners' intrinsic motivation becomes valuable for organizing the key elements to explore during the needs assessment study.

Given the importance of intrinsic motivation for this problem of female learner success in STEM MOOCs, the self-determination theory (SDT) provides a beneficial model (Deci & Ryan, 2000). SDT can reframe these factors and identify valuable relationships between the far-ranging influences at play. Briefly, Deci and Ryan (2000) summarized intrinsic motivation by fulfilling three crucial psychological needs: competence, autonomy, and relatedness. Building on social cognitive theory (Bandura, 1977a, 1986), SDT extends the concept of self-efficacy into the broader idea of competence (Deci & Ryan, 2000). Competence highlights the importance of experiencing mastery and receiving positive feedback. Autonomy emphasizes an individual having enough freedom to make their own choices and guide their behavior. Finally, relatedness centers on gaining security from connection with others and the environment.

Examining how other researchers have implemented the SDT framework helps determine its utility in the Coursera context. First, measures of competence, autonomy, and relatedness, based on the SDT model, are positively correlated with student motivation and achievement levels in authentic educational settings, including secondary and tertiary schools (León et al., 2015; Liu et al., 2014; Murphy et al., 2019). Additionally, this model has already been explicitly applied to MOOCs (Martin et al., 2018) and used to explain the broader gender gap in STEM (Murphy et al., 2019; Simon et al., 2015; Stolk, Zastavker, et al., 2018). The SDT model's three main variables of competence, autonomy, and relatedness have been shown to change over time in direct relationship to the pedagogical methods and course design used, thereby demonstrating the malleable nature of intrinsic motivation (Stolk, Jacobs, et al., 2018; Vennix et al., 2018). Plus, SDT-based interventions have been successfully implemented to improve women's retention in STEM secondary and higher education programs (Dell et al., 2018). Thus, the SDT model has been used to capture learner motivation in a similar context, illuminate causes of the broader STEM gender gap, and improve learning outcomes through interventions increasing student motivation. These previous findings highlight how the SDT model can provide a solid conceptual framework for the remainder of this research study.

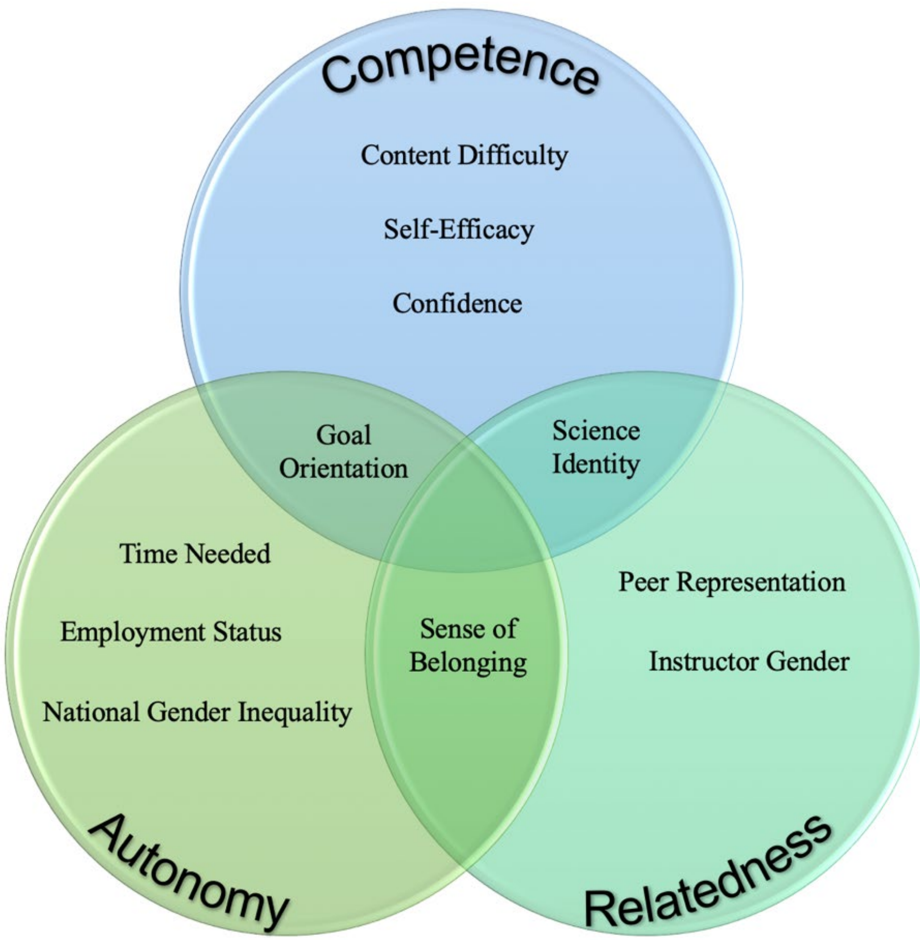
All the relevant factors identified from the literature synthesis fit into these three variables outlined by SDT, with some falling into the overlap between categories. The area of competence encompasses the factors of content difficulty, learners' confidence, and self-efficacy, as these influences each connect back to an individuals' belief in their abilities. Bandura (1977a) also explored how self-efficacy links directly with an individual's perception of their own competence in a particular domain. Autonomy captures the national-level gender inequality, employment status, and time needed to finish these courses since these are each constraining factors impacting females' ability to have the resources to engage in STEM MOOC learning. Lastly, role models can improve connection within educational settings, so female peer representation and instructor gender fall under the relatedness category.

Several of the factors explored pertain to more than one of these psychological needs. For example, goal orientation combines a learner's belief in their own ability to achieve planned milestones (competence) and the necessary resources to work toward their goals (autonomy). Similarly, science identity includes an analogous need for self-confidence in one's STEM abilities (competence) while also feeling a part of and accepted by the scientific community (relatedness). Finally, a sense of belonging requires a connection with others in your learning community or field (relatedness) and the freedom to engage and become a valued group member, and this latter process is often constrained by societal stereotypes and inequality (autonomy).

By reorganizing these influences under the broader umbrellas of competency, autonomy, and relatedness, new groupings surface regarding how these factors relate to one another and the broader construct of student motivation. When considering the needs assessment study, the SDT model can be refined to include only the factors most readily examined in the Coursera learning context. Figure 2 provides an overview of this conceptual framework and the pertinent factors relating to these three psychological needs.

Figure 2

Conceptual Framework Applied to the Problem of Practice



Note. A visualization of the SDT conceptual framework with relevant factors from the literature synthesis arranged by their connection to the three psychological need categories.

Conclusion

Many factors influence the pervasive and persistent gender gap witnessed in STEM higher education. These factors range from national-level patterns of bias and inequality to individual differences in confidence and sense of belonging. Given that millions of female learners already enroll in STEM MOOCs (Grella & Meinel, 2016; Jiang et al., 2016), the remainder of this research will focus on better supporting these women once they start, opposed to how to get more females interested and enrolled in these courses.

The remainder of this research will likely benefit the most by investigating factors related to in-course learning experience, using the SDT model to focus on drivers of intrinsic motivation. Even after this narrowing, the list of potential factors remains expansive and includes instructor, peer, content, and learner characteristics. Additionally, this narrowing of focus does not mean the macrosystem trends should be ignored during the needs assessment. Analyzing country-level trends will still offer valuable insights into learners' behavior patterns in STEM MOOCs and help inform subsequent interventions at the course level. However, research has highlighted how local and individual factors often have a more substantial effect than broad-scale policy changes (OECD, 2015; Sax et al., 2017; Walton et al., 2015). Finally, this narrowed scope on learners' motivation within the course experience aligns with where this researcher has the most control to make changes, given her position at Coursera.

Chapter 2: The Needs Assessment Study

This chapter presents a preliminary study examining the gender gap in STEM courses on the Coursera platform. This investigation aimed to answer research questions on the extent to which this STEM gender gap exists on Coursera and how the characteristics of MOOCs and the learners themselves relate to female engagement. The researcher also examined women's self-reported reasons for stopping before completing STEM MOOCs. After a description of this study's purpose, an outline of the methodology is provided, including descriptions of the sample, measures, data collection, and data analysis used. Finally, the findings are summarized by question and category to contextualize the factors identified in Chapter 1 within the Coursera learning environment.

Purpose of the Study

This needs assessment focuses on the gender gap in enrollments and completions of STEM MOOCs on the Coursera platform. Despite this well-known gender gap and substantial MOOC research already conducted, a recent systematic review of the top MOOC providers, including Coursera, showed no interventions targeting women in STEM had been implemented (Lambert, 2020). Thus, further research is needed to clarify the problem and inspire tailored interventions for these female online learners.

For this investigation, the researcher used data directly from the Coursera platform, surveying the entire catalog of STEM MOOCs and a subset of these courses when time-intensive coding methods were required. The central unit of analysis was the individual learners on Coursera. However, individual students' metrics were also grouped according to each factor. Specifically, for the nation-level gender inequality trends, the unit of analysis became the country so that learners' behaviors could be viewed in aggregate. For instructor gender and content difficulty, the unit of analysis became the course. Lastly, for confidence and existing time demands, the unit of analysis returned to each learner since individuals' survey responses were explored.

These various units of analysis each aligned most closely with different methods. Regression models offered a valuable technique for assessing the connection between a country's inequality level and females' enrollment or completion rates for the nation-level comparisons. A significant difference between the two samples was assessed using *t*-tests to compare the means and *z*-tests for proportions (Lochmiller & Lester, 2017). Lastly, for the written survey answers, qualitative coding methods for identifying themes and grouping answers by category were crucial to gain insights from these open-ended responses. This combination of quantitative and qualitative methods helped to compensate for the weaknesses of each approach, creating strengthened and more nuanced findings (Lochmiller & Lester, 2017). As outlined by Creswell and Plano Clark (2018), this study used a convergent parallel design to analyze numeric and text data during the same research stage, before then merging and synthesizing the results.

With this mixed-method approach, the researcher offered evidence of the extent to which there is a gender gap in STEM MOOCs and which factors had significant relationships with the decreased engagement of female learners. As one of the largest online global education platforms, Coursera provided the opportunity, and a rich trove of data, to gain a better understanding of online learning trends more generally. Through a systematic study of these data, this needs assessment allowed the researcher to document how and why female learners engage in and complete STEM MOOCs less frequently. The purpose of this preliminary research was to identify which factors most contribute, which allows the creation of future interventions focused on narrowing this persistent gender gap.

Methods

Research Questions to Guide Needs Assessment

For this needs assessment study, the researcher focused on the following questions, each grounded in key factors identified during the literature synthesis. The first question established the existence and magnitude of the STEM gender gap on the Coursera platform. The remaining questions were framed using EST and grouped by factors at the course and learner levels. The SDT model was also utilized, such that these questions addressed factors driving competence, autonomy, and relatedness for the learner. For example, question two focused on the course level within the EST microsystem and explored time needed (autonomy), content difficulty (competency), and instructor gender (relatedness). Using theoretical and conceptual frameworks to inform these questions' design ensured the crucial system levels and domain areas were reflected in the research that followed.

RQ1. To what extent is there a gender disparity in STEM MOOC enrollment and completion on the Coursera platform?

RQ2. How do the characteristics of STEM MOOCs, such as the average time needed to complete, content difficulty level, peer representation, and instructor gender, relate to female learners' enrollment and completion rates?

RQ3. How do female learners' characteristics, such as their home country's national gender inequality level and individual employment status, relate to their enrollment and completion rates in STEM MOOCs?

RQ4. What reasons do female learners give for dropping out of STEM MOOCs before completing, and how often are other time demands, low confidence, and lack of prerequisite knowledge cited?

Procedure

Before these questions could be answered, the researcher needed a clear plan of action to align the research goals with the context, sample, and methods available (Onwuegbuzie & Leech, 2006). This section explores the overall sample, measures, data collection, and analysis plans for the needs assessment study.

Sample

At the start of 2020, the Coursera platform offered more than 2,300 STEM MOOCs. Between March 1, 2016, and March 1, 2020, these STEM courses had more than 1.9 million female and 4.6 million male unique learners enrolled, creating a sizable population from which to draw for this needs assessment study. The various sections of this study utilized different subsets of this large MOOC and learner sample, as aligned with the methods of each research question (Johnson & Onwuegbuzie, 2004; Onwuegbuzie & Leech, 2006). Appendix A lists the subset of courses used for the instructor gender and female representation analysis.

Measures and Instrumentation

Table 1 provides a summary of the constructs and measures for this study. The remainder of this section offers richer context on the rationale for each measure and how each one is defined in relation to the Coursera platform.

Table 1

Summary of Needs Assessment Constructs and Measures

Construct	Operational Definition	Indicator
National Gender Inequality	Gender inequality level by country, considering health, education, and employment of females in that nation	United Nations' Gender Inequality Index (UN GII)
Employment Status	Whether the learner is working full or part time in addition to taking courses	Learners' self-reported full- or part-time employment on their Coursera learner profile
Content Difficulty	Instructor's rating of the course content difficulty level	On Coursera, the label of the content as Beginner, Intermediate, or Advanced
Female Representation	The ratio of males to females who are in the learner cohort by course	The percentage of females out of total enrolled learners in each course
Instructor Gender	The gender of the instructor(s)	The percentage of instructors for each course who are female
Time Needed	Learning time needed to complete the course	Total number of hours used to complete each course, averaged across course completers
Reason for Not Completing	Learners' self-reported reason for stopping a course before completing, as collected in Coursera's automated Inactivity Survey sent via email	Quotes from females on why they stopped before completing. Reasons may include low confidence, prerequisites, time demands, or meeting their goal before completing

Enrollments. The absolute number and percentage of female enrollments in STEM MOOCs can be used to measure learners' attraction, interest, and comfort to engage. On

Coursera, with hundreds of courses in the STEM subject areas, these enrollments can also be used as a metric to compare across courses and assess relative interest levels.

Completions. To assess learners' persistence through the content, the researcher can calculate the percentage of active learners who complete the entire course. The percentage who complete is calculated at Coursera as those who successfully finish all graded assessments divided by the total number of active learners. An active learner is defined as any student who has engaged with at least one learning item of the course, such as a video, reading, or quiz. The number of active learners is used as the denominator instead of total enrollments because many learners enroll without engaging with any learning items in the course; thus, active learners provide a more accurate portrayal of which learners intend to make meaningful progress through the course. Additionally, a unique learner can be active in more than one course, creating the possibility of one individual having multiple active enrollments.

When considering completions across the broader Coursera platform, it is useful to examine the ratio of active enrollments to total completions. The completion rate, both at the course and catalog level, is a valuable indicator of learners' likelihood of persisting through the material and finishing the full content as designed. This completion metric also becomes useful when comparing across gender (i.e., the completion rate for females vs. males) and groups of courses (i.e., introductory vs. intermediate courses).

Time Needed. Other time demands are among the most reported reasons for stopping a MOOC before finishing (Eriksson et al., 2017; Gütl et al., 2014; Loizzo & Ertmer, 2016). With women's limited time due to housework and unpaid childcare responsibilities (Perez, 2019), the time needed to complete the course becomes especially crucial for female learners. The Coursera platform uses clickstream data collection to record the total amount of time each learner spends engaging in any given course, including watching videos, answering quiz questions, and writing code within in-browser programming activities. This total engagement time per learner can then be averaged across all learners who completed each MOOC. Coursera's top STEM MOOCs were used to identify any broader trends between the average needed completion time and females' completion rates for this overall analysis.

Female Representation. The presence of female peers has been shown to increase women's participation and persistence in STEM content (Charleston et al., 2014; Healy, 2017). To investigate this factor, the researcher calculated the percentage of females out of the total active enrollments in Coursera's top STEM MOOCs. The total count of active learners determined the top courses.

Instructor Gender. Given that the presence of female instructors can increase female learners' engagement in STEM MOOCs (Brooks et al., 2018), this factor required investigation during the needs assessment. Focusing on Coursera's top STEM MOOCs by active learners, the researcher tagged each instructor as male or female, as determined by the self-reported gender in their Coursera profile and photo provided on the course homepage. Since many courses on the Coursera platform have more than one instructor, this metric could not simply be a binary indicator. Instead, a percentage was calculated for each course, representing the portion of all instructors for the course who were female. This standardized metric allowed for the analysis of how having a female instructor may affect female enrollments and completions among students in the investigated STEM courses.

Content Difficulty. Given the evidence of women's increased likelihood to enroll in introductory STEM content instead of intermediate or advanced courses (Grella & Meinel, 2016), an essential factor to consider for female engagement in STEM is the content difficulty level. On the Coursera platform, instructors and course staff label their course as Introductory, Intermediate, or Advanced. This variable can then be used to group courses by difficulty and analyze women's enrollment and completion rates across these content groups.

National Gender Inequality. To investigate this broad construct, other researchers have used different United Nations and World Economic Forum indicators to summarize each countries' development and gender equality levels; empirical studies then use these variables to examine correlations with female engagement in STEM (Charles & Bradley, 2009; Guiso et al., 2008; Jiang et al., 2018; Kizilcec, Saltarelli, et al., 2017). After examining these different composite measures more closely, the researcher decided to use the United Nations' Gender Inequality Index (UN GII). This single score for each country encapsulates the average level of health, education, employment, and empowerment of females in that nation. While not a perfect indicator of females' power and autonomy in each diverse nation, this GII score provides a broad summary of women's average experiences compared to men's in that country, resulting from national-level trends. This measure, encompassing a nation's high-level female rights and representation, became the indicator of gender equality to highlight the possible links between this factor and female STEM MOOC engagement.

Employment Status. Women are often tasked with more home and family unpaid work (Perez, 2019), so female learners' time may be limited. Extending previous research on how full-time employment can hinder women's ability to engage with MOOC learning more so than for males (Allione & Stein, 2016; Gütl et al., 2014), the researcher used employment status as a valuable indicator of female learners' likelihood to engage with STEM courses. Each user can indicate their current employment status on the learners' profile within the Coursera platform, including working full or part time, self-employed, student, etc. While not all learners complete their full profiles, thousands of learners have indicated this self-reported employment status. This additional layer of demographic data offers crucial insight into how employment may relate to learners' likelihood to complete MOOCs.

Inactivity Survey. To investigate why female learners stop participating in STEM MOOCs before completion, the researcher read women's self-reported reasons. If a learner has not been active in a course for three weeks, Coursera sends a brief survey to their email to understand why they have not returned to the course material. Even if enrolled in and not active in multiple courses, each unique learner receives this survey once. From these survey responses, the researcher could identify themes explaining why these female learners did not finish. In particular, this analysis provided an opportunity to validate themes found in the research literature, including low confidence (Bandura, 1977b; Handoko et al., 2019; Sujatha & Kavitha, 2018; Wang & Baker, 2015), lack of prerequisite knowledge (Grella & Meinel, 2016; Riegle-Crumb et al., 2012), other time demands (Eriksson et al., 2017; Loizzo & Ertmer, 2016), and meeting their goals without needing to complete (Crues et al., 2018; Kizilcec, Pérez-Sanagustín, et al., 2017; Wang & Baker, 2018). Written responses from the survey data were grouped to create quantitative frequency metrics, with the percentage of female learners reporting each reason for inactivity. Additionally, this survey offered valuable qualitative data in the form of quotations directly from the learners to better understand their course experience in their own words.

Data Collection

To answer these questions, the researcher analyzed de-identified data collected by the Coursera online education platform. The analysis fell under the Johns Hopkins University Institutional Review Board (IRB) blanket protocol for the needs assessment study, alongside the explicit permission from Coursera's legal and data engineering teams. Specifically, the researcher collected aggregate numbers on learners' enrollment, progression, gender, and employment status, plus course-level data on average completion time, instructor gender, and content difficulty level. Learners' self-reported reasons for dropping out of a course, as collected through an automated email survey, were also examined. These data had already been collected through the Coursera platform without human involvement, mitigating any possible influence or coercion. Leveraging existing data also meant learners did not need to be recruited or contacted during this needs assessment, ensuring the comfort and safety of the participants. This procedure aligns with Smith's (2008) ethical argument that participants should not be contacted if the researcher can access existing data that already answer their questions.

With extensive data available, the researcher used differently sized datasets to answer the various research questions. The large datasets leverage the power of Coursera's global platform to identify quantitative relationships. Furthermore, the qualitative tagging of written survey answers and course-level analysis required manual data tagging, so a subset of the total population data was curated and utilized. For the dropout survey, filtering by gender (females only), by course domain (STEM only), and by time period (in the last six months) narrowed the written-in responses to 174 single-line answers, a manageable amount to review individually. For the course-level examination, Coursera's top 100 STEM MOOCs by active enrollment offered a promising subset and ensured the most popular content determined these findings. Upon further investigation, one of the top 100 STEM MOOCs did not provide Coursera's typical learning design of graded assessments, meaning that the completion metrics for that course did not carry the same meaning. Thus, that course was dropped from all quantitative analyses, and only the top 99 STEM MOOCs were used.

Data Analysis

This needs assessment employed a mixed-methods approach, using a convergent parallel design to combine insights from numeric and text data (Creswell & Plano Clark, 2018). All statistical tests were conducted in SPSS (Version 26.0) and Microsoft Excel (Version 16.5). Specifically, regression models were used to identify any significant correlations between the factor of interest and the relevant outcome metric, either enrollment or completion (Lochmiller & Lester, 2017). For example, a regression model proved useful when analyzing the relationship between the average time to complete a course and females' likelihood to complete that course, as examined across the top STEM MOOCs on Coursera. Residual plots were analyzed to test the four assumptions of linearity, homoscedasticity, independence, and normality to determine if a linear regression model would suit the data (Knapp, 2018). When these assumptions did not hold, non-linear models were tested to ensure accurate results.

In addition to regression models, other quantitative methods were utilized. A two-sample *z*-test was used to assess differences in proportions when comparing female and male completion rates across STEM MOOCs. Additionally, two-sample *t*-tests were used to identify a statistically significant difference in means between two groups of interest. This method enabled the researcher to assess if females comprised a larger portion of active enrollments, on average, in courses with any female instructors compared to only male instructors.

Finally, qualitative coding was used for open-ended survey responses to identify themes and assess if female learners' answers aligned with different potential dropout factors from the research literature. As Hsieh and Shannon (2005) describe, a directed content analysis approach was used to incorporate prior research conducted on this topic. A priori codes were primarily used to determine the relevance of themes that emerged from the research literature when examining the context of STEM MOOCs on Coursera (Hsieh & Shannon, 2005; Lochmiller & Lester, 2017). Any text that did not fit under an a priori code was assigned a new code. For the specific naming of each theme, in vivo codes were used to reflect the exact language of learners from their written responses when possible (Elliott, 2018; Lochmiller & Lester, 2017; Miles et al., 2014). Triangulation was implemented by combining findings across different data sources and methodological approaches, all examining the same phenomenon (Guba, 1981; Small, 2011). Using a convergent mixed-methods design could enhance the accuracy and trustworthiness of this study's conclusions (Shenton, 2004; Small, 2011).

Results

Research Question 1: Gender Disparity in STEM MOOCs on Coursera

Before investigating factors, the researcher assessed the extent to which a gender gap exists in STEM MOOCs on the Coursera platform. Enrollment, activity, and completion data were used to examine this gender gap. From March 1, 2016, to March 1, 2020, male learners had 10,888,589 total active enrollments in STEM MOOCs on Coursera and 1,854,981 completions, creating an effective completion rate of 17%. Over the same period, female learners had 3,312,666 active enrollments and 430,221 completions, a completion rate of 13%. This difference of 30% in completion rate between female and male learners is consistent with previous STEM MOOC studies examining this gender discrepancy (Crues et al., 2018; Kizilcec & Cohen, 2017).

The active enrollments by gender indicate that males engaged with STEM MOOCs at a rate three times higher than their female peers. Even after controlling for their lower enrollment numbers, female learners were still less likely to complete a course: 13% of female active enrollments completed the course, compared with 17% of males, a statistically significant difference ($z = 176; p < 0.0001$). If female learners were achieving the same completion rate as their male peers during this timeframe, more than 130,000 additional course completions would have been realized for women taking STEM MOOCs on Coursera.

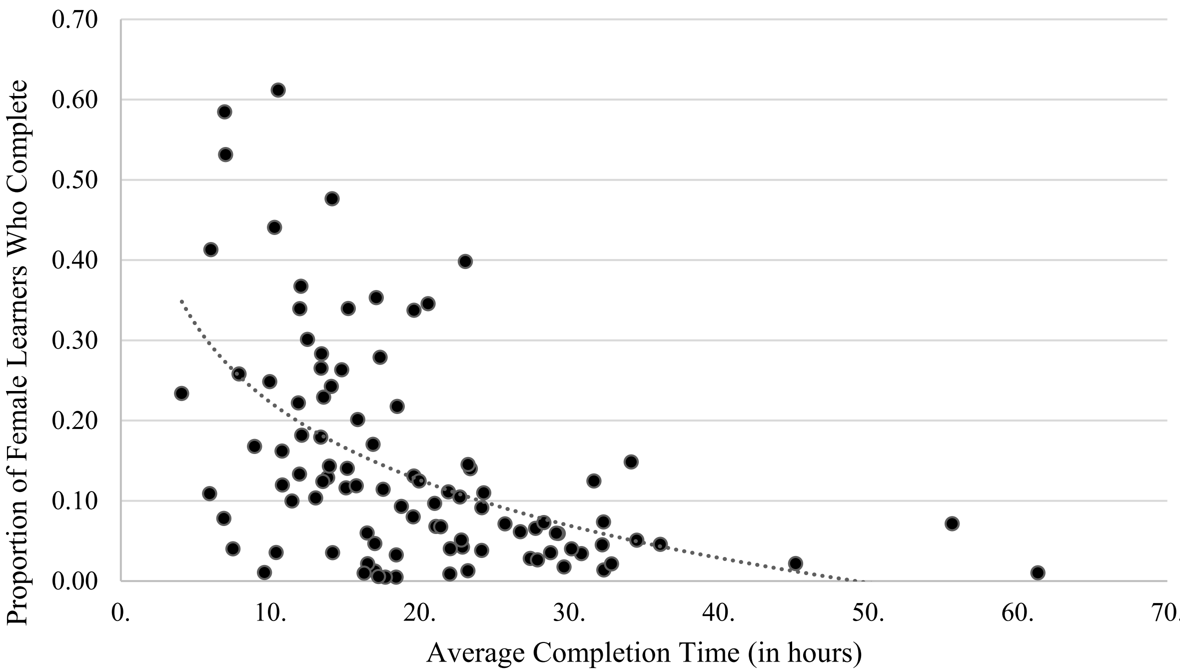
Research Question 2: Course Characteristics Affecting Female Engagement

Time Needed

The researcher used a regression model to assess the correlation between the “average time to complete” on learners’ likelihood to complete. After observing the scatter plots (see Figures 3 and 4) and residual plots (see Appendix B), the researcher performed a logarithmic transformation of the dependent variable to meet the assumptions of linearity and homoscedasticity of the residuals. The resulting logarithmic regression model shows how, for both men and women, the longer the average time needed to complete a STEM MOOC, the lower, on average, the completion rate. The researcher found the average completion time had a more substantial predictive power for females ($R^2 = 0.25$) than males ($R^2 = 0.21$). This correlation aligns with intuition: more content will usually be harder for everyone to finish successfully.

Figure 3

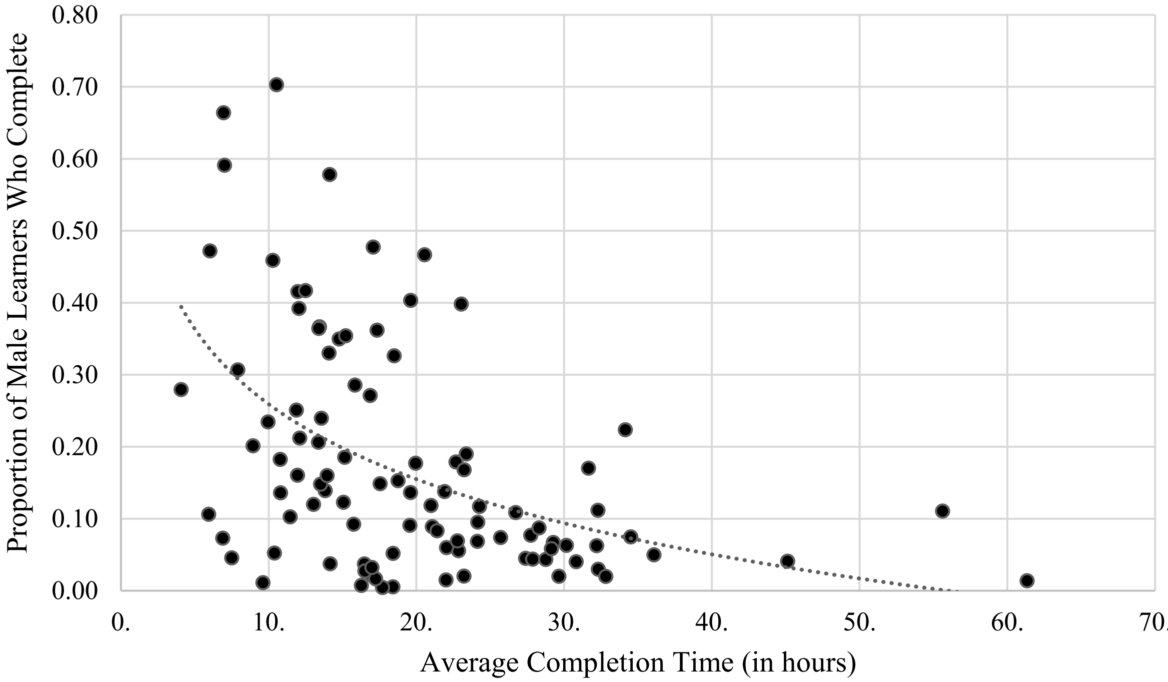
How Average Completion Time Relates to Female Learner Completion



Note. The scatter plot and fitted logarithmic regression model demonstrate the relationship between the average amount of time needed to complete and the proportion of enrolled active female learners who completed each of the top STEM MOOCs on Coursera.

Figure 4

How Average Completion Time Relates to Male Learner Completion



Note. The scatter plot and fitted logarithmic regression model demonstrate the relationship between the average amount of time needed to complete and the proportion of enrolled active male learners who completed each of the top STEM MOOCs on Coursera.

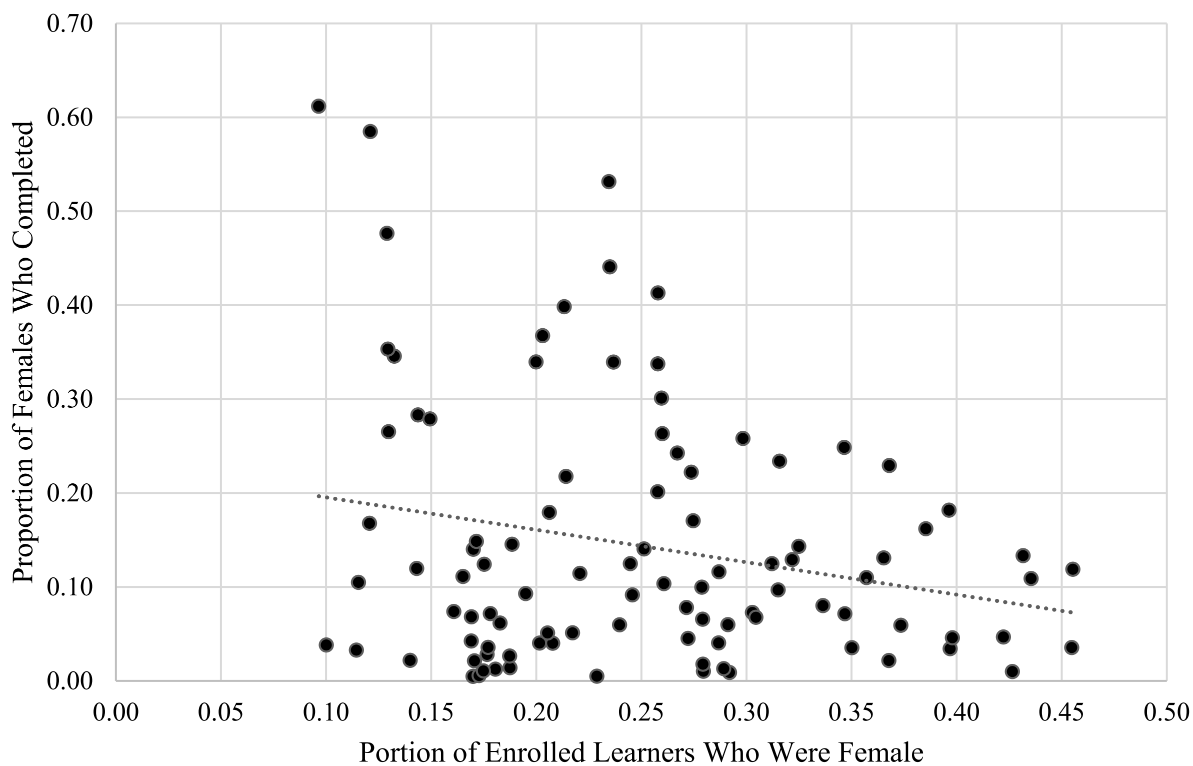
Female Representation

To explore how the percentage of learners who are female enrolled in each STEM course affected females' likelihood of completing the course, the researcher used linear regression. This method revealed the lack of a meaningful relationship between these variables of interest (R^2 = 0.05). The scatter plot supports this evidence, with no apparent pattern visible in the graph (see

Figure 5). Table 2 summarizes the correlations found across the time needed to complete and female representation independent variables.

Figure 5

How Female Peer Representation Relates to Female Completion



Note. The scatter plot and fitted linear regression model demonstrate the relationship between the percentage of enrolled active learners who were female and the percentage of these females who completed the course for each of the top STEM MOOCs on Coursera.

Table 2

Course Characteristics' Correlations with Completion

Factor	Beta	R^2
Time to Complete (for females)	-0.14 ln(x)	0.25
Time to Complete (for males)	-0.15 ln(x)	0.21
Female Learner Representation (for females)	-0.35(x)	0.05

Note. Courses, n = 99

Instructor Gender

The researcher compared courses with and without any female instructors to assess how the instructors' gender may influence female learners' enrollment and completion patterns. Since most of the top STEM MOOCs on Coursera have zero female instructors, all courses with any female instructors were grouped together for this quantitative analysis. In courses with at least one female instructor, female learners appeared more likely to enroll, with females comprising a significantly greater percentage of the learner cohorts: 30%, on average, for the courses with at least one female instructor compared to 23% for courses with all male instructors ($p < 0.001$). However, having at least one female instructor did not appear to influence the percentage of female learners in the course who eventually completed: 11% of females completed courses with at least one female instructor, compared to 16% of females in courses with no female instructors, ($p = 0.10$). Appendix A provides the full list of courses analyzed with the percentage of the instructors who were female, female enrollments, and female completion metrics. Table 3 summarizes these quantitative findings related to instructor gender.

Table 3

Instructor Gender's Effect on Female Enrollments and Completions

	% Female Enrollments	% Completion by Females
No Female Instructors	0.23	0.16
One or More Female Instructors	0.30	0.11
p-value	0.001	0.10

Note. Courses, n = 99

Content Difficulty

After grouping courses by content difficulty, i.e., "Beginner" and "Intermediate," the researcher was able to investigate how these levels affected enrollment and completion by gender. Courses within larger learning programs do not have individually tagged content difficulty levels and thus had to be omitted for this analysis. This narrowing left 49 courses with course-level difficulty tagging from the list in Appendix A. Females comprised 28% of the learner cohorts in beginner STEM MOOCs compared to 19% in intermediate courses ($p < 0.001$). However, content difficulty level did not appear to have a statistically significant relationship with the percentage of enrolled females who eventually completed the course: 18% of females completed beginner courses compared to 15% in intermediate courses, ($p = 0.54$). Overall, this factor showed a similar pattern to that observed for instructor gender. Table 4 presents these quantitative findings related to content difficulty and female course engagement.

Table 4

Content Difficulty's Effect on Female Enrollments and Completions

	% Female Enrollments	% Completion by Females
Beginner	0.28	0.18
Intermediate	0.19	0.15
p-value	0.001	0.54

Note. Courses, n = 49

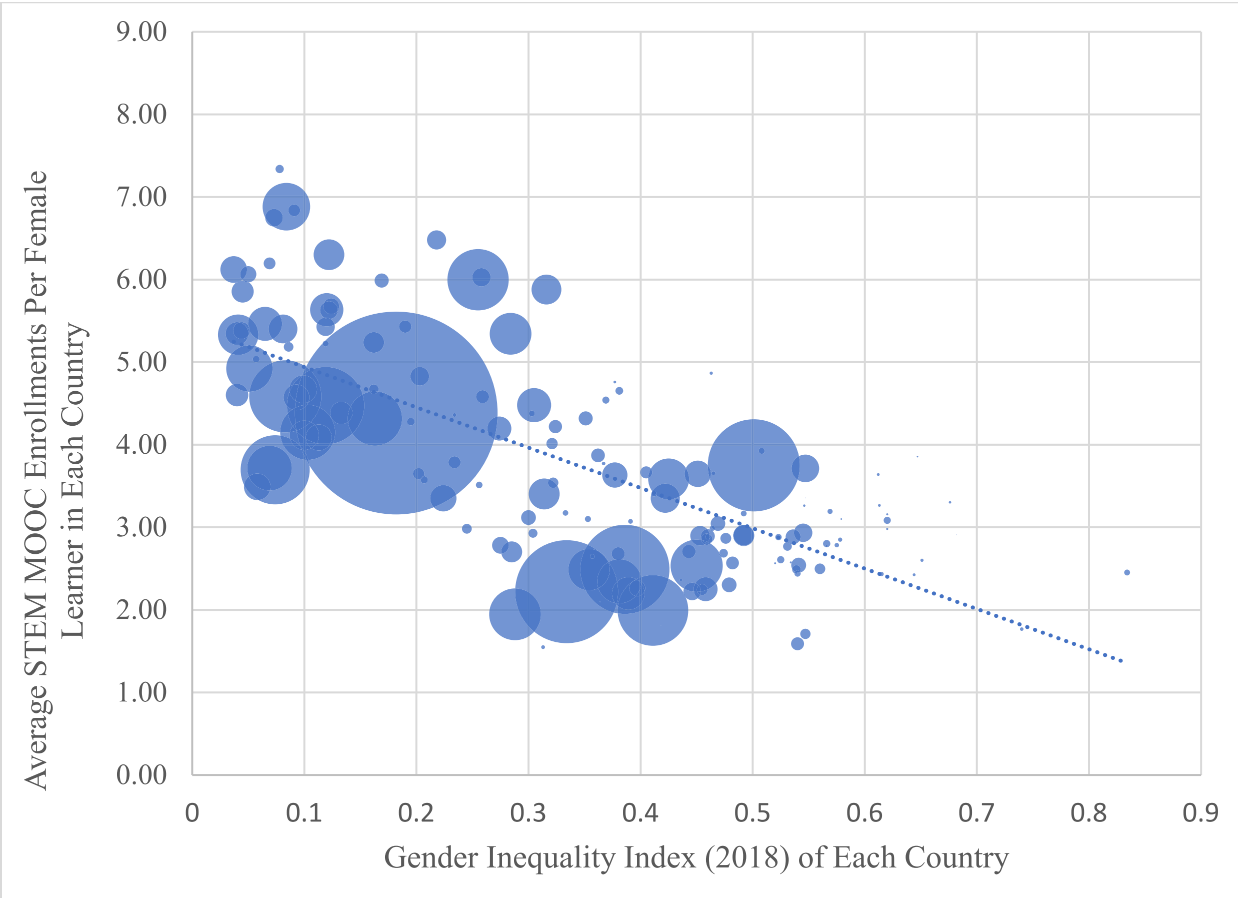
Research Question 3: Learner Characteristics Affecting Female Engagement

National Gender Inequality

The researcher used linear regression to assess the relationships between national gender inequality and female learners' enrollment and completion metrics. The UN GII from 2018 was used as the indicator of each country's female-to-male equality level, summarizing each nation's health, labor market, and female empowerment trends. Using linear regression, the researcher found the UN GII scores strongly correlated with female learners' lower enrollments in Coursera courses ($R^2 = 0.51$), meaning females from countries with more gender-equal societies were more likely to enroll in STEM MOOCs on the Coursera platform (see Figure 6). Figure 7 shows a weaker but still significant correlation between each country's UN GII and the average number of completed STEM MOOCs per female learner by country ($R^2 = 0.31$).

Figure 6

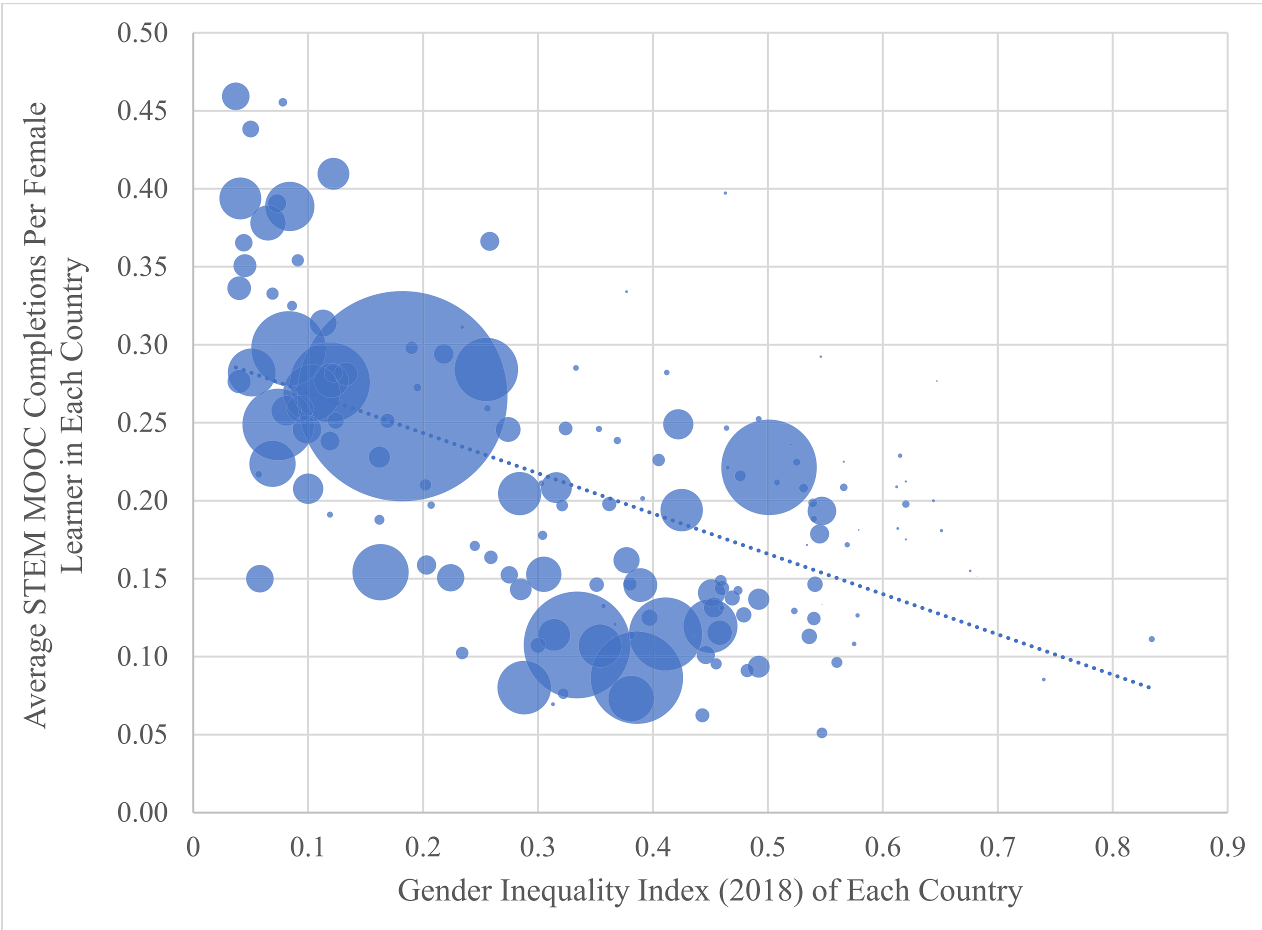
How National Gender Inequality Relates to Female Enrollments in STEM MOOCs



Note. The scatter plot and fitted linear regression model demonstrate the relationship between national gender inequality and the average number of STEM MOOC enrollments per female learner for each country with greater than 30 female enrollments in STEM MOOCs on Coursera. Each bubble represents a single country, with the width of the bubble displaying the relative number of female learners joining STEM MOOCs on Coursera from that nation.

Figure 7

How National Gender Inequality Relates to Female Completions in STEM MOOCs



Note. The scatter plot and fitted linear regression model demonstrate the relationship between national gender inequality and the average number of STEM MOOC completions per female learner for each country with greater than 30 female enrollments in STEM MOOCs on Coursera. Each bubble represents a single country, with the width of the bubble displaying the relative number of female learners joining STEM MOOCs on Coursera from that nation.

Table 5

National Gender Inequality's Effect on Female Enrollments and Completions

Theme	Beta	R^2
National Gender Inequality and Female Enrollments	-4.88(x)	0.51
National Gender Inequality and Female Completions	-0.26(x)	0.31

Note. Countries, n = 153

Employment Status

To investigate how full-time jobs may affect course persistence by gender, the researcher split STEM enrollments by employment status and compared completion metrics. Males with self-reported full-time jobs enrolling in STEM MOOCs (N = 402,451) had an average of 1.8 course completions per learner. Females with full-time jobs enrolling in STEM MOOCs (N = 186,062) had an average of 1.1 course completions. An unpaired t-test demonstrated a significant difference between genders' average course completions (t = 499.4, p < 0.001). Women with full-time jobs completed fewer STEM MOOCs on average than their full-time employed male peers. Furthermore, these male learners had an average of 5.9 course enrollments for every course completion, whereas these female learners had an average of 7.7 course enrollments. This finding highlights how full-time employed females enroll in more courses than full-time employed males, on average, before eventually completing one.

Research Question 4: Females' Self-Reported Reasons for Dropping Out

The researcher used qualitative tagging to identify themes from the open-ended responses provided in the Inactivity Survey of female learners who stopped STEM MOOCs before completing. All a priori themes (low confidence, other time demands, needing prerequisite knowledge, and meeting their goal before completing) appeared in the written responses. Additionally, the theme of payment surfaced during the coding process. In vivo codes were used when possible and are indicated with quotations and italics in the full list of categories identified, alongside their frequencies of prevalence (see Table 6).

The most prevalent categories reported by women as reasons they stopped before finishing included insufficient time, an issue with the content, lack of interest in the assignments, low confidence in their abilities, missing prerequisite knowledge, and not wanting to pay. This list includes the three primary reasons women do not complete courses suggested by the literature synthesis. The researcher highlighted these reasons in research question four: time demands, low confidence, and a lack of prerequisite knowledge. A subset of female learners within the “not interested in the assignments” group, 13 out of 30, reported they had already gained what they wanted from the course. This subset demonstrates the importance of goal orientation and how learners begin MOOCs to achieve various goals, not all of which require course completion.

Additionally, a handful of responses (22 out of 174) did not fit into any category. These miscellaneous answers spanned a broad range of topics, including learners’ career changes, technical difficulties, and mistakenly thinking they had already completed the course.

Table 6

Female Learners' Self-Reported Reasons for Stopping a STEM MOOC Before Completion

Theme	Prevalence
"No time"	21%
Content itself	20%
"Not interested in assignments"	17%
"Not confident"	14%
Prerequisite knowledge	9%
"Don't want to pay"	6%
Other	13%

Note. Learners, n = 174

In addition to trends, exploring females' longer quotations and specific language for dropping out of a STEM MOOC helps elucidate the gender gap in greater depth. For example, falling under the prevailing category of insufficient time, seven unique learners included the exact phrase "I don't have time" as part of their response. Other females in this same theme cited "time constraints" or "[I] need to dedicate more time to this last item and could not yet," which both highlight how life can get in the way of fully engaging with the course material. Also included in the theme of "no time," one learner responded, "I don't have a lot of free time to write a paper. It's hard to write a paper with interruptions." Not only does this learner comment on the lack of availability in her schedule to devote to online learning, but she also brings up the idea of "interruptions," repeating the theme of how women often are not in control of their free time because of home and childcare responsibilities (Perez, 2019). From this same "no time" theme, another learner wrote, "These courses are fascinating. I have learnt so much from Coursera and would like to thank all the team for making this knowledge available to us. However I often have time and family constraints." This longer explanation highlighted her enjoyment and appreciation of the course materials, indicating her interest in the subject area and the quality of the content are not issues. Despite her desire to spend more time on the course, this learner clearly expressed that family duties restricted her ability to engage fully. Her gratitude appears as genuine as her busy schedule.

In the “not confident” bucket, female learners used consistent language to express a lack of faith in their abilities. Different learners explained feeling “stuck,” “helpless,” “not sure,” “not confident,” and “not good enough.” This negative language around their competence in the course exemplifies the well-documented lower confidence and self-efficacy levels seen for women in STEM throughout the literature (Handoko et al., 2019; Jones et al., 2013; Sax et al., 2017; Williams & George-Jackson, 2014). One learner explained, “I would have failed the test,” indicating how she assumed she would fail without even attempting the final assignment of this course. This quotation epitomizes how low confidence often prohibits females from continuing in STEM even when they have the same ability level as their male peers, a finding supported by research (Blackburn, 2017; OECD, 2015). Another learner noted, “I don’t want to share my answers publicly or with other learners.” Her explanation highlights the insecurity and anxiety often present for learners who stop before finishing a MOOC (Gütl et al., 2014; Sujatha & Kavitha, 2018). Exploring learners’ full written reasons for stopping before completing uncovers nuances within each coded theme and was useful to consider when designing interventions for these females.

Synthesizing Findings Across Research Questions

The well-documented gender gap in STEM higher education extends into the Coursera MOOC context. Not only are male learners enrolling at rates more than three times higher in STEM MOOCs, they are also significantly more likely to complete the courses than female learners even after controlling for enrollment numbers. These quantitative findings highlight the vast gender disparity and opportunity for interventions tailored to female learners' needs and preferences in this male-dominated educational setting.

Beyond the existence of the gender gap itself, several factors surfaced as meaningful contributors to females' lower engagement with STEM MOOCs on Coursera. National gender inequality showed significant predictive power on females' likelihood of starting and finishing STEM MOOCs. Content difficulty and instructor gender displayed similar patterns of impact on female participation in STEM MOOCs: a significant relationship with women's enrollment numbers but a non-significant relationship with their completion metrics. Female peer representation in the learner cohort also did not exhibit a meaningful relationship with females' completion rates at the course level. Lastly, time to complete the course linked to stopping a STEM course before completing it and full-time employment appeared to be a larger barrier for females than for males. The stronger correlation seen between time needed to complete and female completion rate—opposed to male completion rate—relates to Perez's (2019) exploration of home and family responsibilities often falling more on females.

Combining insights across the quantitative and qualitative analyses provided further evidence of the most salient factors. For example, the regression model for the average time needed to complete presented a strong relationship for females. Plus, in their open-ended survey responses, female learners referenced a lack of time more than any other reason for why they were unable to complete a course, suggesting the importance of this factor and the broader power of the autonomy domain in the SDT model. Additionally, a larger proportion of female learners self-selecting to enroll in beginner STEM MOOCs than their male peers suggests they have lower confidence in their abilities or simply less prerequisite knowledge in these topic areas.

Before completing, female learners who stopped a STEM MOOC cited a lack of confidence in their abilities slightly more often than referencing holes in their topic knowledge.

Given that these two rationales, lower self-efficacy and missing prerequisite knowledge, appeared in a significant subset of learners' written responses, we can hypothesize that both factors play a role. The research literature on this issue indicates that these females' self-efficacy is more likely to hold them back from progressing than their actual STEM expertise (Murphy et al., 2019; O'Dea et al., 2018). These findings on self-efficacy, confidence, and content difficulty emphasize the significant role of the competence domain in the SDT model for this problem of practice. Overall, female students' burden of additional time demands (Perez, 2019) and weaker self-efficacy (OECD, 2015; Wang & Baker, 2015) are both contributing to this complex problem.

Conclusion

Insights from the needs assessment study clarify nuances of the gender gap problem in STEM MOOCs on Coursera and help define the next steps for exploring potential solutions. In preparation for the intervention, it will be most beneficial to focus on retention instead of gaining new female enrollments. This focus is in response to the immense number of women already enrolling in STEM MOOCs (Grella & Meinel, 2016; Jiang et al., 2016) and this researcher's purview of influence at Coursera. The intervention study will focus on better supporting these learners who are starting but not yet completing STEM MOOCs on the Coursera platform.

Within these courses, the most critical factors to address will be female learners' time constraints, often lower self-efficacy, and challenges of societal gender inequality. While learners' home, work, and country ecosystems cannot be altered directly, the intervention design can focus on transforming STEM MOOCs to align with women's already demanding lives while boosting confidence in their abilities throughout their learning journeys. The following chapter explores other researchers' findings when attempting to support learners with similar challenges and how interventions based on SDT can lead to successful results.

Chapter 3: Review of the Intervention Literature

Women continue to complete MOOCs at lower rates than their male peers, and this effect is exaggerated in technical subject areas (Alario-Hoyos et al., 2017; Crues et al., 2018; Grella & Meinel, 2016; Ihsen et al., 2015). While hundreds of thousands of women are already enrolled in STEM MOOCs, most do not complete those courses (Allione & Stein, 2016; Kizilcec & Halawa, 2015). As one of the world's largest MOOC platforms, Coursera has demonstrated this same pattern in a recent needs assessment, with female learners significantly less likely to complete STEM MOOCs (13% vs. 17%) even after controlling for their drastically lower enrollment numbers compared with males. Leaders at Coursera have an opportunity to shrink this persistent gender gap by better supporting female learners in STEM content.

From Problem to Potential Solutions

Currently, MOOCs on Coursera are presented without personalization to the specific individual, which creates a more suitable learning environment for some learners than for others (Allione & Stein, 2016; Eriksson et al., 2017). While the current design works for many users, and thousands already complete these online experiences, the structure does not equally assist all learners. Updates would likely raise the tide, most helping those currently falling behind their peers (Allione & Stein, 2016; Grella & Meinel, 2016). Evidence suggests the current MOOC structure most benefits male learners and those with previous experience in the subject domain area (Gütl et al., 2014; Ihsen et al., 2015).

Female learners often face additional barriers to completing STEM courses. Modern researchers have investigated Bandura's (1977a) concept of self-efficacy in the MOOC context. They have found that women, on average, have significantly lower self-efficacy, especially in STEM material, which drives lower retention and performance (Sujatha & Kavitha, 2018; Wang

& Baker, 2015). Thus, helping learners augment their self-efficacy may particularly aid women's retention in STEM MOOC content (Handoko et al., 2019; Sujatha & Kavitha, 2018). Additionally, women worldwide face more substantial time constraints than their male peers because of family and household responsibilities (Perez, 2019). These extra time burdens make women with full-time jobs less likely to complete MOOCs than their employed male peers (Allione & Stein, 2016; Gütl et al., 2014). Time constraints are among the most frequently cited reasons for stopping a MOOC before completion (Eriksson et al., 2017; Loizzo & Ertmer, 2016). Women's time constraints are often compounded in countries where females have diminished autonomy and lower gender equality (Guiso et al., 2008; Kizilcec, Saltarelli, et al., 2017; Perez, 2019). During the needs assessment study on Coursera, learners' self-efficacy, time constraints, and home country's gender inequality level emerged as the three strongest factors related to reduced female retention in STEM MOOCs. These findings suggest how women often have steeper barriers to completing courses and may benefit from greater attention to their needs, especially in these three identified areas.

Conceptual Framework to Guide Solutions

Returning to learners' motivation provides a useful starting point for investigating potential solutions. As discussed in Chapter 1, self-determination theory (SDT) posits that intrinsic motivation stems from an individual's psychological need for competence, autonomy, and relatedness (Deci & Ryan, 2000). With its internal focus on the learner, SDT emphasizes the microsystem level of the ecological systems theory (Bronfenbrenner, 1986), which is likely where this researcher can have the greatest impact. The seminal motivation theory of SDT by Deci and Ryan (2000) provided a solid conceptual model on which to build a literature review for possible interventions.

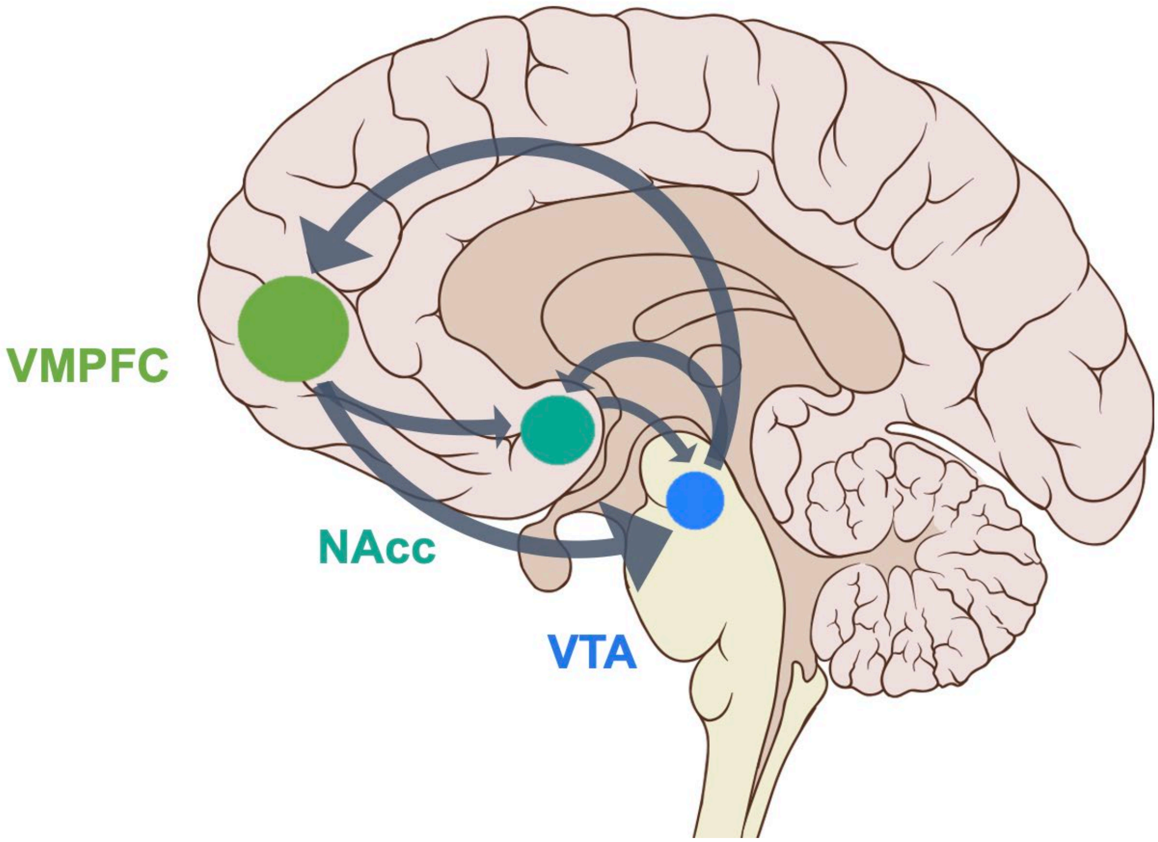
Connecting SDT to Cognitive Neuroscience Research

Recent neuroscientific findings demonstrate how an individual's seeking system and reward circuit mimic the intrinsic motivation constructs of the SDT model. Cognitive neuroscientists often use the seeking system to explore intrinsic motivation since this circuity underlies much of humans' curiosity, desire, and persistence (Di Domenico & Ryan, 2017). Dopamine, a neurotransmitter linked to reward and delight, drives the neuronal connections in this seeking system. While complex behaviors such as motivation result from varied neural circuits working in tandem, researchers have found activation from three core brain regions especially relevant (Bouarab et al., 2019; Di Domenico & Ryan, 2017; Mannella et al., 2013). Specifically, the ventral tegmental area (VTA), the nucleus accumbens (NAcc), and the ventromedial prefrontal cortex (VMPFC) are thought to comprise the centers of an individual's seeking system (Di Domenico & Ryan, 2017). Figure 8 shows these brain areas and the dopaminergic projections among these regions.

Notably, the VTA is most known for responding to the reward of completing a new task and the stress when abilities are tested (Bouarab et al., 2019; Diederen et al., 2016), suggesting a relationship with SDT's competence domain. The VMPFC is responsible mainly for decision-making and risk-reward analysis of new choices (Hiser & Koenigs, 2018; Murayama et al., 2015). This relationship with choice and planning links to SDT's autonomy category. Finally, the NAcc has been implicated in goal-directed behavior, specifically linking individuals' values and goals (Mannella et al., 2013). This focus connects to SDT's relatedness construct, highlighting how the alignment between content utility and an individual's values is represented in the brain. While each neurological region is complex and multipurpose, connections to the SDT model are evident across this neurological seeking system (Di Domenico & Ryan, 2017; Murayama et al., 2015). The SDT model can help situate possible interventions within the neurobiological system of learners' intrinsic motivations through precise alignment with how the human brain processes and responds to new situations.

Figure 8

Core Brain Regions and Connections of Individual's Seeking System Mirroring SDT



Note. This sagittal view of the human brain highlights the core regions and connections that comprise an individual's seeking and reward system, including the ventromedial prefrontal cortex (VMPFC), the nucleus accumbens (NAcc), and the ventral tegmental area (VTA). The colored dots highlight these three brain regions, and the arrows represent the synaptic connections between regions. This figure is adapted from Dubuc's (2020) open-access neuroscience resources.

Aligning SDT with Proposed Interventions

The theory-based conceptual framework of SDT aligns with the purpose and goals of this study. Using surveys, researchers have found that lower intrinsic motivation in females most strongly predicts the gender gap in university STEM majors (Stolk, Jacobs, et al., 2018). Females, on average, demonstrate lower levels of self-efficacy (affecting their perceived competence), feel less choice and control (interrupting autonomy), and exhibit lower task value and fewer supportive connections (hindering relatedness) across studies analyzing secondary and tertiary STEM progress (Murphy et al., 2019; Simon et al., 2015; Stolk, Zastavker, et al., 2018). These lower levels of competence, autonomy, and relatedness link to lower intrinsic motivation levels for females in STEM content (OECD, 2015; Simon et al., 2015; Stolk, Zastavker, et al., 2018). Furthermore, women's diminished motivation directly connects to their reduced likelihood of continuing with STEM content (León et al., 2015; Murphy et al., 2019; Simon et al., 2015).

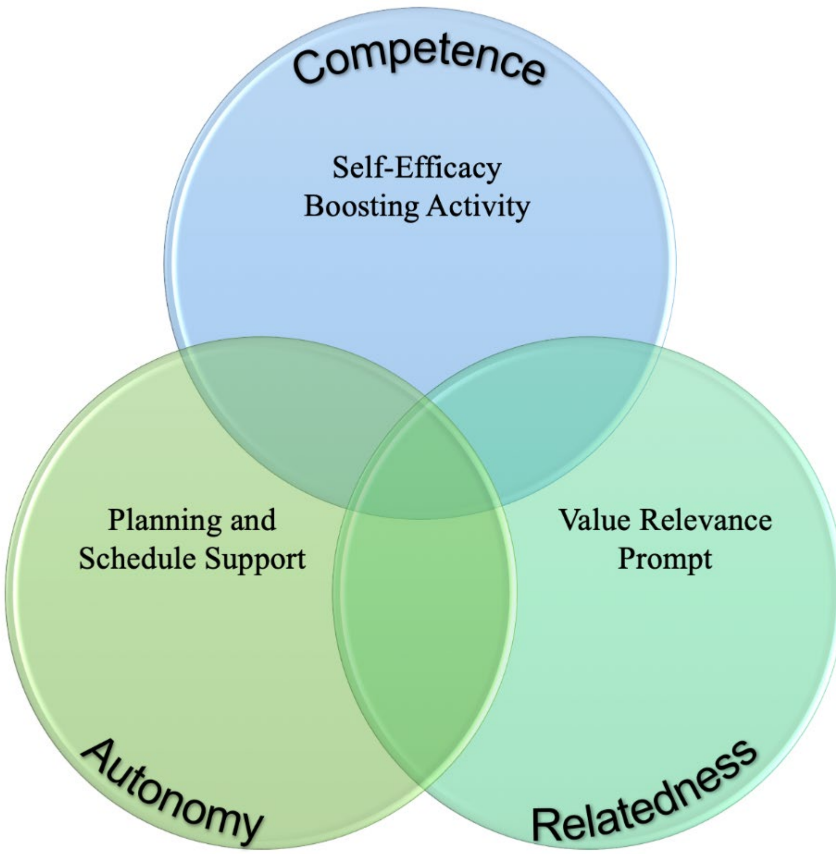
Also, the SDT framework aligns with this research's context and design. Other authors have previously applied the SDT model to MOOCs (Martin et al., 2018) and design interventions to better support females in higher education STEM courses by increasing intrinsic motivation (Dell et al., 2018). Thus, the SDT framework is especially relevant to the Coursera platform and this solution-oriented gender gap research. Just as the factors explored in Chapter 1 affecting females' reduced participation and persistence in STEM MOOCs map onto the three areas of SDT, the categories for these potential interventions parallel the pillars of this framework, as shown in Figure 9's Venn diagram.

This literature review investigates the benefits and challenges of implementing an intervention to address each of the three areas identified in the needs assessment by fulfilling learners' need for competence, autonomy, and relatedness. First, boosting self-efficacy may help counteract female learners' often low confidence in their abilities to succeed and directly relates to the SDT area of competence. Helping learners fit the course content into their already busy schedules may allow greater autonomy and ultimately greater success. Finally, increasing the salience of learners' values within the content may help them feel greater relatedness in the course and persist in the content. Thus, this value-focused intervention links back to the relatedness category.

Figure 9 shows the central relationship between each psychological need and proposed intervention, even though these three interventions each arguably draw on more than one of the three SDT pillars. These interventions will aim to fulfill the main areas of psychological need as outlined in SDT to increase intrinsic motivation and ultimately boost persistence in the course. Figure 10 summarizes this mechanistic framework, including the mediating variables and outcome metrics of interest.

Figure 9

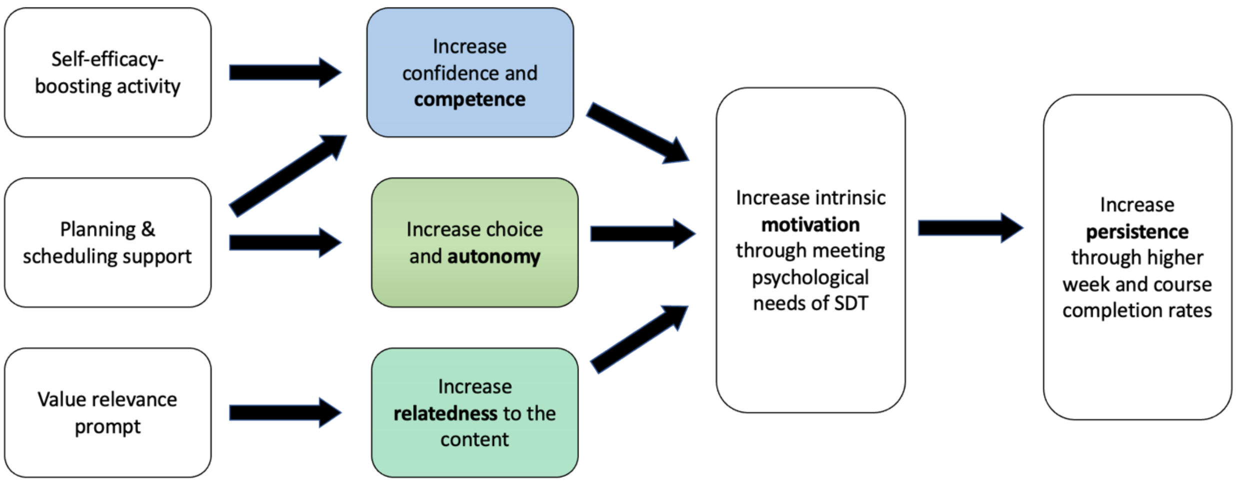
Conceptual Framework of the Proposed Intervention Areas



Note. This simplified visualization of the SDT framework, as presented in Chapter 1, shows each proposed intervention area appropriately aligned to one of the three main psychological categories of Deci and Ryan's (2000) intrinsic motivation theory.

Figure 10

Mechanistic Summary of Conceptual Framework for Proposed Interventions



Note. This view into how each intervention can increase course persistence through the mediating factors and psychological needs outlined in SDT provides a mechanistic framework summarizing the change processes behind the desired outcomes.

Intervention Literature Review

Previous research has outlined practical methods for assisting female learners through tailored interventions. MOOCs attract thousands of learners to the same course, providing an ideal environment to pilot new interventions (Aguilar, 2018; Kay et al., 2013). The dynamic MOOC context enables automation and assessment of the utility of personalized messages at scale (Kizilcec et al., 2020). With an unprecedented volume of recorded data and the possibility for random-assignment experiments, MOOCs also offer unique spaces to further researchers' understanding of self-regulation and motivation strategies, including individual differences of what works best when and for whom (Kizilcec & Brooks, 2017; Lodge & Corrin, 2017; Yeomans & Reich, 2017). Many instructional designers already apply findings from the science of learning to MOOC content, including using multiple modalities to explain concepts and spaced retrieval practice to test understanding (Mayer, 2019; Moulton, 2014). Building on others' applications, this researcher can strengthen support for female learners in STEM MOOCs on the Coursera platform. Thus, this section explores the existing literature related to three distinct intervention options for the in-course experience on Coursera: augmenting learners' self-efficacy, offering concrete planning and scheduling support, and increasing the content's value relevance.

Self-Efficacy-Boosting Activity

Learners' confidence in their skills can influence their engagement and success. Self-efficacy centers on a learner's belief in her abilities to accomplish the task at hand (Bandura, 1977a). Learners need confidence in their abilities to satisfy the competence domain area of the SDT framework (Deci & Ryan, 2000). In addition, these beliefs can significantly affect learners' task interest, motivation, retention, and performance, especially in higher education STEM settings (Bandura, 1977a; Huang & Mayer, 2019; Macphee et al., 2013; Sujatha & Kavitha, 2018; Williams & George-Jackson, 2014). During the needs assessment, the researcher found that women comprise a larger proportion of enrollment in introductory STEM courses than intermediate ones, likely signaling weaker self-efficacy than their male peers. Low confidence in their abilities also emerged as a key theme when coding females' self-reported reasons for stopping a STEM MOOC prior to completing.

While these beliefs may seem fixed, many researchers have highlighted the malleable nature of self-efficacy (Macphee et al., 2013; Williams & George-Jackson, 2014) and even boosting learners' self-efficacy in online learning settings (Huang & Mayer, 2019; Peechapol et al., 2018). Some investigators have recently shown the causal impact of a self-efficacystrengthening video and written messages in an online STEM course (Huang & Mayer, 2019). Aligning with the performance accomplishments and verbal persuasion sources of self-efficacy outlined by Bandura (1977a), the intervention for enhancing self-efficacy would aim to celebrate assignment successes and offer encouragement throughout the course. For example, a video or text-based prompt could provide motivational strategies for handling STEM anxiety and praise learners' progress through different course milestones. The following section explores the benefits and shortcomings of adding an asynchronous video to increase self-efficacy based on previous researchers' successful experiments (Huang & Mayer, 2019; Kizilcec & Cohen, 2017).

Self-efficacy is a vital area to focus on for females in STEM MOOCs. First, females tend to have lower self-efficacy in STEM than their male peers, even before secondary school (Blackburn, 2017; OECD, 2015). However, females with higher STEM self-efficacy are more likely to persist in math, science, and technical courses, making this construct a key variable for influencing the gender gap in the STEM pipeline (Blackburn, 2017; Simon et al., 2015; Williams & George-Jackson, 2014). Furthermore, higher self-efficacy in learners relates to a higher likelihood of persisting and completing MOOCs (Sujatha & Kavitha, 2018), including in majority-female samples (Handoko et al., 2019) and STEM content (Wang & Baker, 2015). Third, short text- and video-based interventions can successfully increase self-efficacy in online, asynchronous settings, especially for students with limited experience in the subject (García-Martín & García-Sánchez, 2020; Huang & Mayer, 2019; Peechapol et al., 2018; Reeves & Chiang, 2019). In fact, self-efficacy interventions have been the most effective for younger students and females, two demographic groups more likely to stop STEM MOOCs before completing (Chyung, 2007; Macphee et al., 2013; Rabin et al., 2020).

Researchers have documented many benefits of incorporating a self-efficacy-boosting activity within online learning settings. For example, Huang and Mayer (2019) randomly assigned 147 participants, a mixture of university students and adults from a crowdsourcing platform, into treatment or control groups, ensuring consistent gender ratios across units. Both groups received the same mathematical course content, and the treatment group was provided self-efficacy-boosting activities. Specifically, these researchers integrated brief verbal and written messages to introduce anxiety-coping strategies and praise students for their effort in an online statistics course. As shown by pre- and post-survey results on a previously validated scale, the treatment group's increased self-efficacy caused significant positive differences in practice, retention, and transfer of their new mathematical knowledge. This focus on in-course performance and longer-term learning gains provides valuable outcome measures to consider for the current study. Given the online setting, STEM subject matter, and majority-female sample, this study's significant results are particularly relevant to STEM MOOC gender gap research. The authors designed these interventions to be asynchronously delivered online, with clear parallels to the massive course environment on Coursera.

Other researchers have also seen success using an online self-efficacy-boosting intervention. With a final sample of 224 university students (79% female) enrolled in an undergraduate biology course, Bernacki and colleagues (2019) examined the effect of digital training. The two-hour training for the treatment group introduced science principles and corresponding self-regulation study strategies to increase learners' confidence and engagement in the course material. While not exclusively focused on self-efficacy, these researchers drew on broader research to raise learners' motivation and performance in an online setting. These researchers used statistical analysis of their random-assignment experiment to reveal that those who received the learning strategies training significantly increased their use of optional practice quizzes and had consistently higher grades on the summative exams throughout the course. Although a two-hour training is likely not possible in the MOOC setting, insights from this training remain important to consider when designing possible confidence-boosting interventions. This study's findings, combined with those from Huang and Mayer (2019), demonstrate clear quantitative advantages of using a self-efficacy-boosting activity to assist females' motivation to persist in STEM MOOCs on the Coursera platform.

In addition, scientists have investigated how the human brain's prediction of a reward can affect behavior and learning, mirroring the concept of self-efficacy. Specifically, researchers measured blood-oxygen-level-dependent (BOLD) responses from 27 participants in a functional magnetic resonance imaging (fMRI) scanner (Diederen et al., 2016). The task required participants to predict the size of an upcoming reward drawn from distributions with varying standard deviations. Participants showed greater midbrain, specifically VTA, activation through increased BOLD responses when making predictions for the rewards from smaller variability distributions. As individuals became more confident in their predictions, their VTA activation increased, which helps humans learn to make data-informed decisions. These insights into how limiting uncertainty can increase confidence and help students learn new skills can be applied when designing the timing and content of these upcoming MOOC interventions.

Returning to the SDT model, "perceived competence is necessary for any type of motivation" (Deci & Ryan, 2000). Humans use the midbrain to update reward predictions, serving as an indicator of their confidence in their abilities when starting a new task, given external reinforcement (Di Domenico & Ryan, 2017; Diederen et al., 2016). With the VTA's role in the brain's broader seeking system, increases in VTA activation correspond to heightened motivation (Bouarab et al., 2019; Di Domenico & Ryan, 2017). By exploring how greater confidence in reward predictions can increase VTA activation within the brain's broader motivational circuitry, Diederen and colleagues (2016) provided biological evidence for how increasing confidence and self-efficacy may benefit motivation at a neurological level.

However, there are potential downfalls of pursuing a self-efficacy intervention. First, attracting learners to engage with optional videos in MOOCs can be challenging, with as little as 10% of total learners completing this type of activity (Jansen et al., 2020). Although high-quality interventions might be present, the positive effects can only occur when students fully participate. Second, self-efficacy interventions often help younger students and females, with males' self-efficacy shown to be less malleable (Chyung, 2007; Macphee et al., 2013). Thus, this intervention may only help certain Coursera users, so, while beneficial to this study's gender-gap focus, it may be less desirable to implement as an overall company strategy. Finally, creating new video content requires time, money, and effort far beyond written prompts, increasing the resources needed for this intervention option (Huang & Mayer, 2019). The researcher will need to weigh the meaningful impact of this intervention with the costs of its implemention, which may not be realistic for this dissertation study. This intervention option would need to be adapted to fit the scalable nature of the MOOC setting and limited time of Coursera's online learners.

Planning and Schedule Support

Compared with traditional higher education courses, the asynchronous nature of online courses allows for more flexibility in how individuals participate and personalization in what content different learners see (Eriksson et al., 2017; Sun et al., 2015; Yu et al., 2017). This section focuses on one aspect of this flexibility: helping learners fit the course activities around their busy schedules. Autonomy-driven motivation emerged as the largest gender difference in a study of STEM university students using a previously validated SDT survey, indicating how autonomy may be the most needed psychological area in which to intervene (Stolk, Zastavker, et al., 2018).

Since time demands emerged as one of the largest interferences with women completing MOOCs during the recent needs assessment, more tailored schedules or support in planning may help these learners progress more successfully. Average course completion time more strongly predicts females' than males' likelihood of completion, suggesting time plays a more prominent role in females' inability to finish. In addition, "no time" emerged as the most prevalent qualitative theme from females' self-reported reasons for stopping STEM MOOCs before completing. Adding planning and schedule recommendations can provide viable paths for learners to work around time constraints and satisfy their need for autonomy, one of the three main SDT domain areas.

Without changing the content itself, the learning platform could recommend one of three or four different schedules for how learners could progress through the MOOC, given their demographic information, preferences, time constraints, and previous learning patterns (Sun et al., 2015; Yu et al., 2017). For example, instead of a course displaying its content over only four weeks, certain learners may benefit from that same material spread out over eight weeks. Alternatively, maintaining the original schedule but actively assisting learners through automated prompts on planning and strategies to budget time may also help them progress (Bernacki et al., 2019; Yeomans & Reich, 2017). Specifically, courses can incorporate guiding questions to help learners identify where, when, and how they would make time to progress, as well as sample study plans they could apply to their own lives (Bernacki et al., 2019; Yeomans & Reich, 2017). In a recent review of self-regulated learning interventions in MOOCs, planning emerged as one of the most successfully malleable outcome variables (Wong et al., 2019), indicating the importance of focusing on this area in future studies. Through differentiated course schedules or assistance in planning, MOOCs can offer learners the flexibility to tailor the content to fit existing time demands. This section investigates the pros and cons of providing learners with schedule options or planning support to assist them in succeeding in the course.

Providing schedule options has many potential advantages. First, learners report a lack of time or inability to fit with their existing schedule as one of the most common reasons for stopping a MOOC prior to completion (Eriksson et al., 2017; Loizzo & Ertmer, 2016; Onah et al., 2014). Furthermore, MOOC content is well-suited for personalized messages and schedules, given its asynchronous nature and modular design. Sun and colleagues (2015) explored building micro-learning pathways from existing MOOC content to fit learners' time constraints and desires. Online learning platforms can already direct learners to the specific topics they desire using machine-learning algorithms built to match skills with unique items in each course (Urban, 2019; Yu et al., 2017). For example, on the Coursera platform currently, a learner can search for "python pandas" to find a specific video explaining that topic as well as related full courses (Urban, 2019). Thus, even though Coursera has not added scheduling personalization, the course setup and platform's infrastructure align with this possible addition. While MOOCs are already asynchronous by design, building personalized schedules with varied assignment deadlines and content amounts per week for different learners would require significant resources to incorporate.

Instead of fully designed schedule options, a lighter-weight version of this intervention could involve a planning prompt that asks students to reflect on where, when, and how they will return to progress in the course. Action-oriented planning prompts have been used in MOOCs to request learners to write where and when they will make time to study (Yeomans & Reich, 2017). This brief intervention underscored the autonomy and importance of students creating a personalized plan of where and when to study. These researchers conducted the study across STEM and non-STEM MOOCs with majority-female cohorts in all three courses and a total final sample of 2,053 learners. They randomly assigned learners to the course with or without the planning prompt, keeping all other course activities the same. Despite the intervention's brief nature, they found the planning prompts resulted in a 29% increase in the treatment group's completion rate. The results of this brief intervention for females in MOOCs demonstrate immense promise for a similar solution to assist learners on the Coursera platform.

In a study with different versions of planning support offered to MOOC learners, Jansen and colleagues (2020) presented three short (four minutes or less) videos focused on goal setting, planning, and time management. The content of their intervention centered on concrete planning techniques for students, including methods for incorporating achievable study goals into already busy schedules. They randomly assigned the 1,471 active learners across STEM and non-STEM MOOCs to each course's experiment or control version. With only partial demographic data available, these researchers decided not to publish any age, gender, or geographic information of the learners in their sample. Using regression models, they found the short videos caused significant increases in course retention and completion rates for the treatment group, despite many learners not watching the full videos. Together with Yeomans and Reich's (2017) findings, these results highlight how brief written or video interventions around planning can have a meaningful impact on course persistence and completion. These positive results can even occur when learners only spend a few minutes engaging with the planning activity.

Cognitive neuroscientists have also explored the benefits of autonomy and choice on individuals' task performance. With a sample of 31 healthy adults (17 female), researchers had participants play a game while in an fMRI scanner to examine the effects of forced vs. self-determined choice (Murayama et al., 2015). They found that self-determined choice significantly improved task performance, as measured by response speed. The fMRI scans showed failure feedback led to a drop in VMPFC activation for the forced-choice but not the self-determined choice condition. Notably, the VMPFC's resilience to failure in the self-determined choice trials correlated directly with an individual's improved task performance. These researchers hypothesized that the VMPFC activation patterns demonstrated improved intrinsic motivation from the autonomy of choice and that even failure feedback can be viewed as helpful information when motivation is high. These neuroscientific studies elucidate how autonomy-based interventions rooted in the pillars of SDT can affect neural-activation patterns related to motivation and improved persistence.

Additionally, providing more structure and greater autonomy over a MOOC's schedule can help boost self-efficacy, another major factor linked to female learners' stopping STEM MOOCs during the recent needs assessment. For example, reducing cognitive load by providing structure through specific task-ordering and scheduling recommendations can increase self-efficacy and motivation to persevere for undergraduates in a higher education STEM course (Feldon et al., 2018). Furthermore, by analyzing successful learners' habits and techniques in MOOCs, researchers have identified how explicit scheduling and planning activities can increase learners' self-efficacy and likelihood to complete (Lung-Guang, 2019; Sambe et al., 2017). When enrolled in a self-paced instead of a strict-deadline MOOC, learners reported higher satisfaction and perceived likelihood to succeed, reflecting a positive increase in self-efficacy from this greater schedule control (Watson et al., 2018). Providing automated but still personalized schedule recommendations increased students' performance on the final exam of an online course, with improvements in self-efficacy accounting for 24% of the variance in their final learning achievement (Xu et al., 2014). Adapting to individual learners' needs, including presenting a personalized schedule, can also boost learners' motivation to persist in optional online learning experiences (Alario-Hoyos et al., 2015). The demonstrated impact on learners' motivation, retention, and performance highlights useful outcome metrics for this future study while emphasizing the broader utility of helping learners create a schedule plan that works for them.

However, there are downsides to the potential integration of these personalized schedules and planning prompts. While successful in smaller pilot settings of learners (Huang & Mayer, 2019; Kizilcec & Cohen, 2017; Yeomans & Reich, 2017), a recent large-scale MOOC study found that simply surfacing these messages to massive learner cohorts nullifies the long-term positive impact (Kizilcec et al., 2020). Instead of purposefully selecting which intervention appeared, the authors randomly chose one of their messages to display to each learner, which may have led to the nullification. More specifically, the planning prompt yielded only marginal increases in learner activity for the first few weeks of the course and no meaningful increases in course completion rates (Kizilcec et al., 2020; Young, 2020). This recent finding suggests that a single planning prompt may not be sufficiently powerful to affect the large gender gap in course completions currently witnessed on the Coursera platform. Plus, tailoring schedule options to different subsets of learners, instead of a single message or activity provided to all, requires significantly more engineering and product resources for a one-time, up-front investment (Alario-Hoyos et al., 2015; Sun et al., 2015). These greater resources required for potentially less return on that investment may make a schedule-planning intervention less realistic for this dissertation intervention study.

Value Relevance Prompt

To help students underrepresented by race, gender, or social status feel more connected to the course, researchers have trialed the use of value relevance activities (Kizilcec, Saltarelli, et al., 2017; Walton et al., 2015). In the recent needs assessment, the gender inequality level of female learners' home countries had a negative relationship to STEM MOOC completion rates. This finding suggests how female learners from developing countries or regions with robust gender biases may experience additional barriers to progressing in online courses.

Value beliefs, focusing on the reasons and extent to which students view a task as beneficial to them (de Barba et al., 2016), build on the SDT area of relatedness and can counteract negative stereotypes (Walton et al., 2015). A value relevance activity requests students to select the value category most meaningful to them (such as family, career, or learning) and then reflect on how participating in the course could further this personal value in their lives (de Barba et al., 2016; Kizilcec, Saltarelli, et al., 2017; Peters et al., 2017). With learners explicitly making these connections to their personal goals and priorities, they often feel greater motivation and purpose in the course, increasing their sense of relatedness to the content. This increased motivation leads to higher retention rates for groups previously most likely to drop out before completing, such as female learners and those from developing countries (Kizilcec et al., 2020; Kizilcec, Saltarelli, et al., 2017).

Incorporating a value relevance activity within STEM content can bring tangible benefits for learners. Given that women from developing countries are less likely to complete STEM MOOCs on Coursera than their female peers from other countries, a value-based intervention shown to assist both females in STEM (Walton et al., 2015) and those from developing countries (Kizilcec et al., 2020; Kizilcec, Saltarelli, et al., 2017) would be beneficial. In a study examining higher education STEM dropout, the authors found a lack of perceived relevance as the most vital SDT factor that emerged during a qualitative interview of first-year science students (Dyrberg & Holmegaard, 2019). Alignment between the course tasks and learners' values was the most frequently cited reason for finishing the course in a study (n = 643) across one humanities and one STEM MOOC on Coursera (Handoko et al., 2019). Notably, this study had mostly females in their voluntary response sample and used a mixture of quantitative scales and qualitatively coded open-ended responses to assess interest, value beliefs, and other self-regulation characteristics. These findings suggest that purposefully emphasizing the alignment between course content and the learners' values may help more people progress in online courses.

Within another MOOC on Coursera, authors explored how different aspects of intrinsic motivation and content engagement relate to learners' performance (de Barba et al., 2016). For this investigation, the authors sought a voluntary response sample of 862 students (26.8% female) in a macroeconomics MOOC offered by the University of Melbourne on the Coursera platform. To assess intrinsic motivation, the researchers adopted a previously developed scale containing questions on interest, mastery-approach, and value beliefs. They asked participants these questions in an online survey, including demographic and prior-knowledge questions. These researchers found value beliefs to be most positively correlated with the number of times learners accessed the course videos, suggesting this domain of intrinsic motivation most benefits students in their engagement with instructional material. Separately, mastery orientation, a different aspect of intrinsic motivation, was most correlated with learners' number of assessment attempts. Additionally, value beliefs showed the strongest correlation with final course performance across all intrinsic motivation areas analyzed, with no differences observed by gender. This study's findings suggest that learners' initial participation in instructional activities, such as watching the videos, may be more indicative of their likelihood of completing the course than continual assessment attempts since value beliefs and course video-watching correlated more strongly with final course performance. Hence, a learner's value beliefs may be a particularly fruitful intervention to obtain the largest statistical effect on their motivation and eventual completion.

More specifically, randomized experiments have exhibited the benefits of these value-based interventions in STEM courses. Peters and colleagues (2017) used a brief value-affirmation intervention in a statistics course with a sample of 290 undergraduate students (75% female). The treatment group wrote about the value most important to them from a list of six potential priorities, while the control group wrote about how the value in the list least important to them may be meaningful to others. All students completed this 10-minute writing activity twice during the course: once in class during week two and then online during week four. Each student retained the same random assignment to either the treatment or control groups. Students in the treatment group, who were making connections to their top values, demonstrated significant increases in numeracy, as shown through improved performance of their math skills. Drawing on past findings, these researchers suggested that students linking their own top values with the course, as done in the treatment group, helped increase their motivation to engage and sense they could succeed. With partial online delivery of the intervention, a majority-female sample, and a mathematics course setting, these researchers demonstrated how a value-focused activity might benefit women in STEM MOOCs, the target population for this intervention.

While these authors did not aim to support only the females in their statistics course, they recommended that others test the efficacy of a values-based intervention specifically to help close the STEM gender gap in higher education courses.

Other evidence highlights the utility of value prompting to help address the gender gap in STEM university course performance. In an introductory college physics course, researchers divided 399 students (29% female) into a double-blind, randomized experiment to test the effectiveness of a brief value affirmation activity (Miyake et al., 2010). Similar to the previously mentioned value prompts, this activity requested those in the treatment group to select their most important value from a list and write about its significance. Those in the control group were assigned to choose their least important value and write about why it may be meaningful to others. This 15-minute writing activity happened twice throughout the course, during weeks one and four, with consistent treatment and control groups. At the end of the course, men in the control group scored significantly higher than women in the control group. Comparatively, women in the value affirmation group had, on average, significantly higher final exam grades than their female peers in the control group; men did not show significantly different scores across the two experimental groups. In addition to this significant gender effect, the authors also found the value activity to have a stronger positive impact on women who endorsed the stereotype that men perform better than women in physics. These researchers found that a brief exercise to self-reflect on values, even when unrelated to the content, can help close the gender gap in STEM performance.

Others have also documented the quantitative benefits of value alignment to propel greater learning. Researchers at edX, a competitor platform to Coursera, recently published an extensive study wherein they implemented several different automated interventions across millions of learners (Kizilcec et al., 2020). While earlier studies had previously tested some of these messages and activities within smaller online learning contexts, this work was the first-of-its-kind research to test these automated messages at scale with thousands of students in hundreds of different courses. Across two and a half years and 247 MOOCs (38% in STEM subject areas) on the edX platform, these researchers tested automated learning nudges based on behavioral science. The 269,169 learners (37% female) were randomly assigned into one of five intervention groups at the start of their course: long-term planning, short-term planning, "value-relevance" to reflect on how their values aligned with this course, "intentions" to identify obstacles to achieving their goals and how they will overcome them, and "social accountability" to track their progress with others. The control group was subject to no intervention. The researchers tracked course progress and overall completion as the outcome measures of success for these learning nudges.

Although the other light-touch interventions, including the one-sentence planning prompts and social accountability statements, lost impact significance when expanded to hundreds of thousands of learners, the positive effect of the value relevance activity continued to raise completion rates for learners from developing countries (Kizilcec et al., 2020). This value-focused activity closed the global achievement gap in courses where it was present. This same positive effect was present in STEM and non-STEM courses where learners from developing countries had not previously been completing at the same rate as their peers. After controlling for learners' home country, formal education attainment, previous MOOC experience, and intention to complete, these researchers did not find any gender effects for their different in-course interventions. While not specifically about females, this large-scale MOOC study by Kizilcec and collaborators (2020) highlights the robustness of the value relevance activity's power, at least for particularly vulnerable learners in STEM MOOCs.

Building on the brain's motivational seeking system, researchers have explored the role of values-based interest in neural activity. Specifically, scientists recruited 15 university students (12 females) in Shanghai, China, to watch two-minute video clips from 15 different MOOCs, eight in STEM subject areas (Zhu et al., 2019). Using electroencephalography (EEG), these researchers recorded brain activation patterns while the participants watched the MOOC videos and then ranked these courses from their highest to lowest desire to learn more. Participants also responded to Likert-scale questions about their value connection to and interest in their top two and bottom two ranked courses. Based on these MOOC video rankings by learning desire, the researchers bucketed the videos into high, medium, and low motivational levels by the participants' average interest.

While not targeting specific brain areas, these researchers recorded frontal, parietal, and occipital regions, as well as picking up brain activity from subcortical regions, including the nucleus accumbens. Thus, these EEG recordings provided a general activation pattern of the brain's multifaceted seeking system while the participants watched the MOOC videos. The inter-subject correlation (ISC) of the EEG-recorded brain activity was higher, meaning more similar across participants, for the higher motivation videos, indicating how neural activity can become more similar when motivation is increased. Interestingly, each video's ISC neural similarity strength was predictive of individuals' reported course-learning desire. While not an intervention study, these researchers still demonstrated how stronger interest and value alignment between the student and MOOC content linked to more similar brain activation patterns across participants and collectively higher engagement.

Beyond the quantitative benefits, there are additional positives to implementing a value-focused intervention. First, creating impactful versions of this activity does not require extensive resources since it entails only a written prompt (Kizilcec, Saltarelli, et al., 2017; Peters et al., 2017; Walton et al., 2015). While some interventions require face-to-face interaction, a value-focused prompt has already been successfully implemented in online, asynchronous settings (Kizilcec et al., 2020; Peters et al., 2017). Given its successful impact at scale, relative ease to implement, and substantial effect seen for underrepresented groups, the value relevance activity may be beneficial when attempting to narrow the gender gap in STEM MOOC persistence on the Coursera platform.

As with all potential solutions, challenges and disadvantages exist for implementing a value relevance prompt in STEM MOOCs. For example, this intervention can sometimes lower completion rates for learners from developed countries if they find the prompt irrelevant and subsequently disengage from the MOOC (Kizilcec, Saltarelli, et al., 2017). However, in a larger, more recent sample across several MOOCs, these adverse effects were no longer observed for those in developed countries, providing more confidence about the efficacy of this option (Kizilcec et al., 2020). Additionally, while this type of reframing intervention can be useful, it fails to address the underlying social marginalization due to racism and sexism that is often present in higher education STEM courses (Walton et al., 2015). A value-based activity may provide meaningful benefits, but it will not solve all related inequities in online learning. While it may be helpful to reorient learners to focus on their values, it is equally essential to reorient instructors to meet the needs of diverse online learners and create an inclusive course community. In the meantime, a value-focused prompt provides a beneficial direction to pursue in STEM MOOCs, given its demonstrated success in supporting the most vulnerable learners.

Conclusion

Ultimately, all three interventions would likely assist different subpopulations of female learners in STEM MOOCs on Coursera. Thus, each one should be tested. Since these interventions necessitate a similar backend engineering infrastructure to surface the messages to learners, the researcher will start by coordinating with other teams to ensure this type of prompting within the courses is possible.

Proposed Intervention Design

The design of the Coursera-specific interventions will draw on successful experiments and learning experiences in similar contexts. For the value relevance variant, the researcher will adapt a previously tested activity in the MOOC setting (Kizilcec et al., 2020; Kizilcec, Saltarelli, et al., 2017) since this only requires a simple, written prompt presented to learners at the start of the course. The researcher will need to modify the other two interventions more substantially to align with this study's scope and resources. The self-efficacy instructional video and the personalized schedules with different chunking of the course materials are both time-intensive projects. Instead, this researcher will build more scalable versions of these interventions, suitable to the MOOC environment and Coursera learners' limited time. These new designs will draw on lighter-weight competency and autonomy interventions, such as the proven self-efficacy-boosting messages (Huang & Mayer, 2019) and written planning prompts (Yeomans & Reich, 2017). These three areas are vital to address and may help different female STEM learners to varying degrees.

During the implementation of these interventions, it will be crucial to monitor who, if anyone, each activity benefits and analyze the impact on different learner sub-groups. Specifically, retention at the week and course levels, as well as demonstrated learning gains determined by in-course performance, will be the primary indicators of success. Based on the results of previous studies, the researcher expects a self-efficacy-boosting intervention would most benefit younger women and those without previous subject matter knowledge (Chyung, 2007; Macphee et al., 2013). Alternatively, adding greater schedule flexibility may be most beneficial for female learners with full-time jobs, who are often juggling family and home responsibilities (Allione & Stein, 2016; Gütl et al., 2014; Perez, 2019). Lastly, a value-focused prompt may likely assist female learners from developing countries more than their peers in developed ones (Kizilcec, Saltarelli, et al., 2017). While it may not be possible to identify a single intervention that helps all female online learners, the researcher hopes to further the evidence of what works best for whom.

Future Directions

Using information collected in each Coursera user's profile, the researcher can investigate which of these interventions most benefit retention and performance for which subset of females, segmented by age, employment status, and home country. These observed patterns could inform which intervention the Coursera platform should surface to future female learners. In further iterations of this work, the Coursera platform could use an algorithm to choose from these intervention options depending on learners' profile information and previously observed results, consistent with other research-backed, machine-learning techniques for personalization (Urban, 2019). Leveraging automation to select the proper intervention for each learner maximizes the potential impact (Chandrasekaran et al., 2015; Sun et al., 2015; Urban, 2019).

In addition to algorithm-based tailoring, future work will include more involved versions of the current prompts being tested. Specifically, providing alternative schedule options based on individuals' time constraints, as suggested by the research literature, interests the product and design teams at Coursera. While not completable in the time period for this dissertation work, the researcher will also be involved with those further intervention projects. Beyond this particular study, the larger goal remains to improve overall support for females in STEM MOOCs through the use of intentional interventions at scale.

Chapter 4: Intervention Procedure and Program Evaluation Methodology

In STEM subjects, male learners complete MOOCs on Coursera at a rate more than 30% higher than their female peers, even after controlling for females' lower enrollment rates (Crues et al., 2018; Kizilcec & Cohen, 2017). The COVID-19 pandemic brought disproportionately more women to the platform than previous years and led to elevated female enrollment numbers, with 37% of STEM enrollments in 2021 from women compared to 31% in 2019 (Glassberg Sands et al., 2021). This increased share of female learners enrolling in scientific and technical courses makes in-course interventions to improve their retention even more critical. To address the pervasive and persistent gender gap in STEM retention, these interventions aim to counteract challenges ranging from national-level patterns of bias and inequality (Guiso et al., 2008; Kizilcec, Saltarelli, et al., 2017) to individual differences in confidence and connection to the material (Charleston et al., 2014; Handoko et al., 2019; Sax et al., 2017; Walton et al., 2015). This specific research plan offers light-touch changes to see how effectively brief, text-based messages can impact retention at scale.

The most critical factors from the recent needs assessment study included female learners' lower self-efficacy, immense time constraints, and challenges from gender-imbalanced societies. The researcher used regression models, hypothesis tests, and qualitative coding by theme to assess why learners drop out. Many high correlation values for behavioral relationships were found (Lochmiller & Lester, 2017; Thompson, 2002). National gender inequality, one factor tested in the recent needs assessment, showed a significant correlation with the average number of completed STEM MOOCs per female learner by country ($R^2 = 0.31$). In addition, female completion rate linked more strongly to the average course completion time ($R^2 = 0.25$) than male completion rate ($R^2 = 0.21$). This result indicates how schedules may be less flexible for females, and course length alone can explain one-quarter of females' variation in completion. Two of the most frequent themes from female learners' self-reported reasons for dropping out of a STEM MOOC before completing were "no time" (21%) and "not confident" (14%).

Subsequently, the proposed intervention includes four new versions of the course with prompts to (a) boost confidence, (b) improve planning, (c) align the content with individuals' values, and (d) a combination of these three interventions into an extra-strength version. While these intervention approaches have been shown to help female learners in varying contexts, each is also designed to help a specific group of learners needing greater support. The first intervention option may most assist younger students and those without a background in the topic (Chyung, 2007; Lambert, 2020; Macphee et al., 2013; Rabin et al., 2020). The second option could especially aid those with more family, home, and job responsibilities (Allione & Stein, 2016; Gütl et al., 2014; Perez, 2019). Finally, the third option is poised to most benefit those joining online courses from developing countries dealing with the challenges of more gender-inequal societies (Kizilcec, Davis, et al., 2017; Kizilcec et al., 2020). Combining the three prompt types for the fourth version may be most beneficial or may overwhelm learners with too many pop-up messages (Moulton, 2014), which is why that stronger treatment option will be tested separately. These four variants, tested alongside a fifth control version, provide a systematic experiment to assess the effectiveness of lightweight messages for empowering female learners to greater persistence and performance in STEM MOOCs.

Research Purpose and Questions

This study aims to apply insights from previous research to the Coursera context to support learners in their STEM learning journey. The goal was to increase female learners' persistence by counteracting the most salient negative findings unearthed in the earlier needs assessment. Leveraging the successes and learnings of others' studies, as explored in Chapter 3, this researcher aimed to implement a fully scalable online intervention with no human action needed after the initial implementation. Coursera team members had previously implemented retention-focused interventions, with some even leading to larger gains for female learners than their male peers (Hickey et al., 2018). However, no interventions explicitly targeting the empowerment of female learners had yet been tried. The goal of this study was to leverage other researchers' learnings and insights to design new, online course support to counteract females' challenges reaching course completion. In particular, this intervention focused on text-based, in-course messages to strengthen female learners' self-efficacy, add reflective planning, and highlight the relevance of individuals' values. With robust technological features and flexibility, the Coursera platform automatically surfaced these messages to learners in STEM MOOCs after the team finalized the design, wording, and timing.

This section documents the high-level plan for the intervention study. The treatment theory is presented first to highlight the theory behind the expected outcomes and how these causal mechanisms take shape in the Coursera context. Then, the research questions are offered to focus this experiment on the critical areas of investigation. The answers to these questions will uncover if the process and outcome metrics were met as expected and how future iterations of this intervention can be improved.

Theory of Treatment

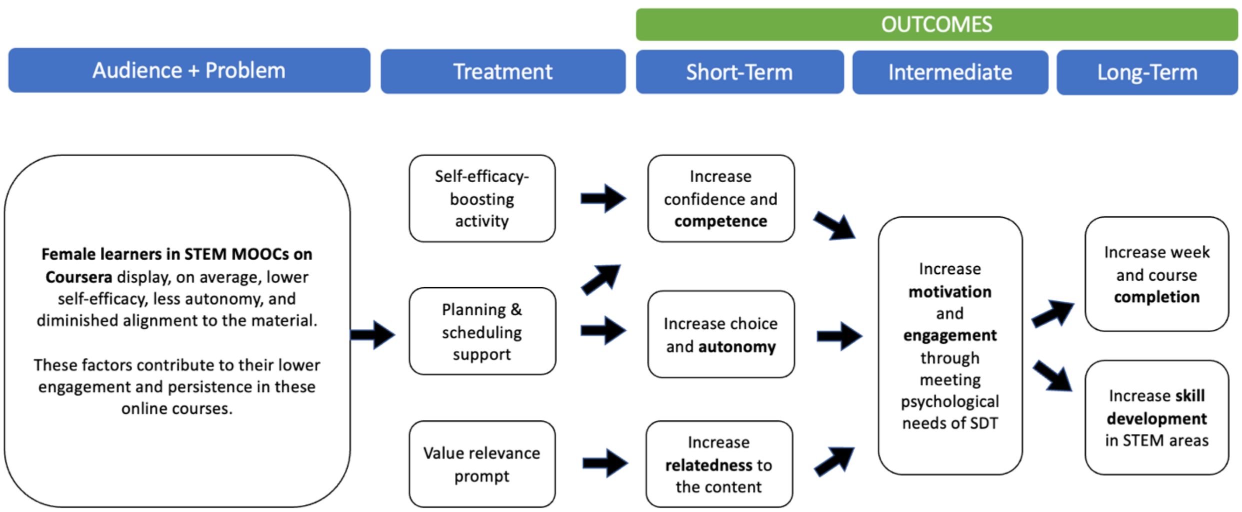
As Leviton and Lipsey (2007) describe, the theory of treatment provides the inputs, processes, and outputs for a given intervention, in this case aiming to influence female learners' engagement in STEM MOOCs. The goal of a treatment theory is to clarify how the intervention could enact positive change. The theory of treatment starts by outlining the problem in its current context and defining the intervention approaches before exploring the mechanisms involved to produce the desired outcomes.

For this treatment, the researcher provided text-based messages to learners directly in online courses. The first message appeared after they clicked on the first item (typically a video or reading) in a participating STEM MOOC and then periodically throughout the first few weeks of content. The design included three main areas of treatment, focusing on increasing learners' self-efficacy, time management, and value relevance. The researcher decided to keep the intervention as lightweight as possible, minimizing the intrusiveness of these additions to avoid interfering with the course material while still aiming to impact learners' engagement positively. While short, in-course messages surfaced to learners may seem too weak a treatment, other researchers have demonstrated success in the MOOC setting with similarly brief interventions, even as short as only one message in an entire course (Kizilcec, Saltarelli, et al., 2017; Yeomans & Reich, 2017).

Using the self-determination theory (SDT) framework, the researcher expected these prompts to increase motivation and further engagement with the course content by fulfilling the psychological needs of competency, autonomy, and relatedness (Deci & Ryan, 200). Female learners have consistently shown lower levels of competency, autonomy, and relatedness in STEM courses (Murphy et al., 2019; Simon et al., 2015; Stolk, Zastavker, et al., 2018). Female STEM learners have also displayed lower intrinsic motivation because of these SDT need deficiencies (OECD, 2015; Simon et al., 2015; Stolk, Zastavker, et al., 2018). This lower intrinsic motivation can cause females to drop out of higher education STEM courses at far higher rates than their male peers (León et al., 2015; Murphy et al., 2019; Simon et al., 2015). Encouragingly, interventions based in the SDT framework have successfully elevated women's persistence and performance in STEM courses (Dell et al., 2018; Huang & Mayer, 2019). Figure 11 presents these causal relationships as related to this study's overall design.

Figure 11

Theory of Treatment Model



Research Questions

This study included both process and outcome evaluation methods, with related research questions. Evaluating the implementation process ensures the intervention runs as desired, the right audience receives the correct dosage of services, and the intervention remains consistent across groups (Rossi et al., 2019). Process evaluation allows for data to be collected during the initial implementation to improve the delivery and eventual outcomes of the intervention. Process evaluation questions can also help researchers assess the causal mechanisms of their research design by examining early indicators of later outcome variables (Baranowski & Stables, 2000). The first two research questions focus on the successful deployment of this experiment by measuring process indicators.

Building on this process analysis, the researcher used the additional four research questions, RQ3 through RQ6, to measure the intervention's effects on the expected longer-term outcomes. These questions centered on assessing meaningful changes among the treatment and control groups in learners' persistence, skill development, reasons for dropping out for those not completing, and continued learning in subsequent MOOCs. These six questions measure concrete implementation and impact indicators, creating a holistic view of this intervention's success.

RQ1. To what extent did the intervention reach the target learner group?

RQ2. To what extent did learners find the prompt helpful?

RQ3. What differences in impact did each intervention have on week one and course completions?

RQ4. What differences in impact did each intervention have on course completers' performance and skill development?

RQ5. How did the intervention affect female learners' self-reported reasons for dropping out of the STEM MOOCs for those who did not complete?

RQ6. To what extent did the intervention spark learners to continue learning in other MOOCs?

Research Design

Building on the research purpose and questions, the researcher selected a methodological approach to align with these goals (Mertens, 2018; Onwuegbuzie & Leech, 2006). Any new intervention needs a thorough evaluation plan to assess the implementation and expected outcomes (Rossi et al., 2019). This section starts by providing the specifics of the intervention study and the mixed-methods approach used. Then, the strengths and limitations of this study design are explored alongside a rationale for the final research design choices. Third, the logic model of this study is presented, which has a theory-based framework of the experiment's activities, expected outputs, and desired outcomes (Cooksy et al., 2001). Fourth, the process evaluation components outline how the implementation of this intervention will be assessed, ensuring the right participants are reached, and the intervention is perceived as helpful. Finally, the outcome evaluation components act as an end appraisal of the intervention's success. This part of the chapter outlines the design and goals of the intervention study before the next section dives into the specific methods and measures that will be utilized.

Evaluation Study Design

The intervention research used an explanatory sequential mixed-methods approach, which emphasizes quantitative analysis first and follows with a shorter qualitative investigation to help elaborate on these results, with notation QUAN → qual (Teddlie & Tashakkori, 2003). This explanatory sequential design aligns with best practices for analyzing the data from a randomized control trial (RCT), assessing any statistically significant, quantitative findings before prioritizing qualitative analysis to enrich the understanding of identified underlying causes (Creswell & Plano Clark, 2011). This approach works best when the researcher has access to relevant numeric data and can apply quantitative data analysis skills, which are true for this study. While RCTs traditionally emphasize quantitative analysis as the primary method for assessing the intervention's impact, adding a qualitative lens strengthens the credibility of findings, uncovers unexpected consequences, and exposes nuances behind higher-level numeric trends (Bamberger et al., 2016).

As is typical for mixed-methods research, this evaluation plan is based on pragmatism, combining various measures and methods to reach a practical answer (Johnson et al., 2007). This pragmatic approach also aligns with the use branch of evaluation, which acknowledges how different methods are often needed to answer varied questions (Mertens, 2018). Proponents of the use branch recommend a quantitative-heavy analysis accompanying an RCT, followed by qualitative methods to uncover deeper insights, just as the researcher has proposed for this study. Advocates of the use branch further emphasize how the end goal of any study is continued learning, mirroring the backbone of improvement science (Christie et al., 2017). This researcher is committed not only to exploring how best to support female learners in this sample but also to sharing these findings with educators in similar contexts so that others can continue building on the insights and learnings from this intervention study.

Given the RCT design with thousands of learners in the sample, the researcher prioritized quantitative analysis to synthesize across these large groups while still assessing participants' sentiments. With one control and four treatment groups, this five-part RCT provides a robust research design not possible in most educational settings. While RCTs can be challenging to implement without contamination, bias, and pushback from other stakeholders, the Coursera online learning platform offers a rare opportunity to conduct a trial of this type. If achievable, randomized experiments are typically preferable to quasi-experimental designs, especially for the statistical analysis of outcome variables (Shadish et al., 2002). With group assignments chosen randomly by the platform as learners newly enroll in the selected MOOCs, the researcher will not see this selection process, and participants will remain unaware of their group assignment. This process eliminates researcher and participant bias in group selection (Rossi et al., 2019; Shadish et al., 2002). Also, since learners join from across the world instead of residing in the same building, as with traditional schooling settings, there is little opportunity for participants to hear about the other treatment activities or move between groups. This participant separation mitigates possible contamination between groups and improves the likelihood of delivering this RCT intervention successfully (Saunders et al., 2005; Shadish et al., 2002).

Strengths and Limitations of the Design

This quantitative-forward RCT study design with qualitative follow-up offers numerous benefits. When designed well, RCTs reduce most threats to internal validity by implementing a random assignment process at the start of the experiment (Rossi et al., 2019). This design allows each individual from the thousands of learners in the sample to have an equal chance of joining any of the experimental groups. Plus, the probabilistic process can be completely hidden from the participants. This randomization also isolates treatment effects as the only systematic differences across the groups, further ensuring strong internal validity for and causal interpretation of the final findings (Rossi et al., 2019; Shadish et al., 2002).

The double-blind RCT design also avoids potential issues affecting construct validity, including participant reactivity from knowledge of group assignment, experimenter's expectations interfering with the intervention delivery, and any compensatory treatment given to those in the control group (Shadish et al., 2002). The online design with automated deployment of the intervention messages removes human error and ensures greater consistency of implementation, a positive indicator of a successful RCT process (Baranowski & Stables, 2000). Finally, the outcome metrics will be operationalized in several ways for this study, including quantitative and qualitative indicators, to measure engagement more thoroughly across these STEM MOOCs. This diversification of indicators and methods improves the accuracy of this study's results (Johnson & Onwuegbuzie, 2004).

This RCT design plan also presents limitations. With quantitative analysis, the researcher can more easily test existing hypotheses than generate new theories relevant to the given population (Johnson & Onwuegbuzie, 2004). Given the ample evidence from similar online learning settings (Huang & Mayer, 2019; Kizilcec, Saltarelli, et al., 2017; Yeomans & Reich, 2017) and the magnitude of data to analyze, this quantitative approach aligns with the context, goals, and comparative research questions proposed (Onwuegbuzie & Leech, 2006). While this quantitative focus limits the type of possible interpretations, testing existing hypotheses suits the nature of this study and the researcher's goals.

With the volunteer nature of these online learners and their busy existing schedules, the researcher wanted the intervention to be as brief and lightweight as possible, which inherently provides a small expected effect size and potential threat to statistical conclusion validity (Shadish et al., 2002). Even with thousands of learners in the sample, it is unclear if this study can successfully balance these brief interventions with having a meaningful impact on persistence and progression. Furthermore, randomized groups do not ensure the intervention implementation proceeds according to the plan. Consequently, reach and exposure will need to be tracked carefully during the initial implementation to enable a valid test of this treatment, with the measures explored later in this chapter (Baranowski & Stables, 2000). Finally, an RCT does not ensure strong external validity or the ability to generalize beyond the setting of the current experiment (Shadish et al., 2002).

Fortunately, the researcher can mitigate many potential threats with the large sample available and the opportunity to conduct a double-blind RCT. Given the low Pearson's correlation of 0.1 found for similar experiments in previous MOOC studies (Kizilcec et al., 2020; Yeomans & Reich, 2017), a sample of at least 1,045 learners would be sufficient for a two-sample *t*-test, with the alpha set at 0.05 and a power of 0.8 (van den Berg, 2020). The researcher reduced the alpha level to 0.01 and increased the power to 0.9 to account for the multiple statistical tests planned and lessen the likelihood of Type I and Type II errors (Shadish et al., 2002). Using G*Power, this researcher obtained a 2,979 minimum for each of the five experimental groups, which should be achievable on the vast global platform of Coursera MOOCs. This large sample size and aligning with the a priori power analysis should mitigate any issues with the statistical conclusion validity (Shadish et al., 2002).

Beyond reaching the appropriate sample size, the researcher will take further actions to mitigate threats to validity. Unlike descriptive or correlational studies, the results of this research can be stated as causal impact relationships because of the experimental RCT design utilized and treatment isolation (Creswell & Plano Clark, 2011; Lochmiller & Lester, 2017; Rossi et al., 2019). Additionally, the researcher plans to extrapolate only to other fully open and online learning environments, keeping the setting consistent with the context of this sample, minimizing issues associated with the external validity (Shadish et al., 2002). Finally, a theory-based logic model, as explored in more depth in the following section, combined with this randomized assignment, creates the environment for authentic causal relationships if the final results are significant (Leviton & Lipsey, 2007). Even after accounting for the limitations, this QUAN → qual mixed-methods design builds on the strengths of the research context, maximizing the potential impact of the extensive data while still letting individual learners' voices influence the outcome results.

Logic Model

A logic model articulates the critical activities of an intervention, relating them to intended outputs and outcomes. In particular, a logic model displays a theory-based diagram of the program's design, focuses data collection on the key activities, and provides an integrative framework for combining findings across a multi-method evaluation (Cooksy et al., 2001). Since female learners display lower self-efficacy, greater time constraint barriers, and diminished value connections to content in Coursera STEM MOOCs, counteracting these factors are at the core of this intervention's logic model. Appendix C displays the full logic model for this intervention research study.

The intervention aims to combat challenges by providing pop-up prompts within the online course learning experience. A backend platform system on Coursera automatically surfaced messages to learners in particular STEM MOOCs at specific times. An A/B testing system was used, dividing each MOOC into multiple course versions and automatically assigning learners as they enrolled to create an approximately equal number of participants in each version (Urban & Greenblatt-Kolodny, 2017). This A/B testing setup is how Coursera and other online learning platforms run RCT-style experiments to assess the impact of content and platform changes on learners' behavior and performance in a course (Kizilcec & Brooks, 2017; Urban & Greenblatt-Kolodny, 2017; Yeomans & Reich, 2017). This intervention followed an A/B/C/D/E experimental design, signifying the four treatment variants and the single control group for each STEM MOOC in the sample.

The prompt variants were designed according to previous researchers' approaches for increasing self-efficacy, planning around busy schedules, and raising value connection to the material. Using the framework of SDT, as highlighted in the treatment theory, these prompts were expected to increase motivation and deepen engagement with the course content by fulfilling competency, autonomy, and relatedness psychological needs. Lower levels of competence, autonomy, and relatedness correspond to lower intrinsic motivation levels for females in STEM content (OECD, 2015; Simon et al., 2015; Stolk, Zastavker, et al., 2018). Plus, women's lower motivation directly correlates to their reduced likelihood to persist in STEM content (León et al., 2015; Murphy et al., 2019; Simon et al., 2015). However, interventions rooted in SDT have shown success in increasing the intrinsic motivation of women in STEM programs and their persistence (Dell et al., 2018). Thus, female learners' lower intrinsic motivation, resulting from a lack of competence, autonomy, and relatedness, is likely one cause of their lower persistence in STEM MOOCs.

Process Evaluation Components

This process evaluation is the first step of the broader explanatory sequential mixed-methods analysis plan. It provides formative indications about whether the STEM MOOC prompts were deployed and received as expected. This broader study's first two research questions explore two critical process evaluation components, reach and exposure, as synthesized by Baranowski and Stables (2000). Appendix D provides a table of these process evaluation questions and how each one was answered.

To explore how successfully this intervention was deployed in STEM MOOCs on the Coursera platform, the researcher focused on reaching the intended learner audience and exposing learners to an intervention they found helpful. As outlined on the left side of the logic model, the target audience and inputs need to be met, and process evaluation is how a researcher can ensure these crucial aspects of an intervention's deployment are satisfied (Rossi et al., 2019). For this study, the process component areas assessed the demographics and reception of the learners interacting with the different intervention designs by averaging across learners in each treatment group. This quantitative approach allowed for more generalizable findings and greater credibility to assess the causal relationships between treatment and outcome variables (Johnson

& Onwuegbuzie, 2004). In the Methods section later in this chapter, the researcher explains how these constructs were operationalized through metrics collected from the Coursera platform and directly from learners.

Outcome Evaluation Components

The researcher evaluated all outcomes for this intervention study by measuring any meaningful changes across the treatment and control groups. Specifically, the key expected outcomes centered on learners' course persistence, skill development, reasons for dropping out if they failed to complete, and later learning in subsequent MOOCs. Appendix E provides an overview of these outcome evaluation indicators.

These four primary outcomes of interest align directly with research questions three through six for this overall intervention study. Persistence can be measured through the week and course completion rates, which align with previous MOOC studies' indicators of success (Kizilcec, Saltarelli, et al., 2017; Yeomans & Reich, 2017). Skill development can be evaluated by a learner's "skill score," calculated by their previous assessment scores relative to their performance in the course, as shown in Coursera's recent Drivers of Quality report (Hickey et al., 2020). For individuals with insufficient previous learning on the Coursera platform, in-course assessment performance acts as a useful proxy for their demonstrated skill development. Finally, subsequent MOOC enrollments after joining this experiment can be used as an indicator of continued learning and commitment to further skill development. Each of these four outcome questions will be assessed by treatment and gender groupings to analyze impact in a more nuanced manner. While learners' home, work, and country ecosystems are external factors that cannot be directly altered, the treatment design focused on light-touch interventions to align the course with women's already demanding lives while increasing confidence. These four questions measure the intervention's impact on the primary outcomes summarized in this study's logic model.

Methods

With millions of total learners and a flexible platform, Coursera offered a dynamic environment for testing this multi-part intervention. The RCT provided a systematic approach for evaluating the utility of these lightweight intervention options through randomized learner assignment into treatment and control groups. The following section outlines the participants, measures, and procedures for this study.

Participants

The participants for this study were learners newly enrolled in specified STEM MOOCs on the Coursera platform while the experiment was open. After filtering to the top STEM MOOCs by active enrollment count and removing partners' courses that have opted out of experimentation, the researcher selected the top 150 courses as the sample for this RCT-style intervention. See Appendix F for the complete list of courses and their active enrollment counts. Learners decided of their own accord to enroll in these courses and could do so at any time. Subsequently, the exact size of the participant group depended on the enrollment patterns in these 150 STEM MOOCs on Coursera during the duration of this experiment.

The STEM MOOCs included in this study were all open-access online courses, meaning there were no admissions barriers for enrolling. Learners could audit the videos and readings for free but needed to pay the subscription fee to gain access to the entire course materials (between $29 and $79 per month) or apply for financial aid for this access. Upon successful completion, learners received a co-branded partner institution and Coursera course certificate but did not receive any university credit. Learners decided to enroll in the STEM MOOCs independently, and there were no monetary rewards or incentives to participate. Learners under 18 years of age were excluded from the experiment and removed from the sample before individuals were assigned an intervention group. The lightweight treatment was delivered automatically by the Coursera platform within the online course experience and could be easily disregarded if the learner did not want to engage with the prompted messages, highlighting the minimal risks posed by this type of intervention.

The final sample provided the size and scope desired for both statistical power and thorough analysis by subgroup. In total, the sample for this experiment consisted of 242,847 female and 365,949 male learners across the 150 STEM MOOCs included. This overarching group represents all learners who newly enrolled in one of the 150 MOOCs during the experiment's open period, from December 8, 2021, to March 20, 2022, and who had their gender identified by either their Coursera profile or a linked profile from another website. This massive group of individuals were used to track the process indicators around reach and exposure for this intervention study.

The demographics of this learner sample mirror Coursera's overall user trends. For individuals for whom age data was available, the majority of learners in this intervention sample were 26 to 45 years old, as is also seen across Coursera's full learner community. Figure 12 and Table 7 display the participants by age and gender for the individuals in the intervention sample for which both these demographic data points were known. Of the female learners in the total intervention sample with data available on their employment status, 42% were employed full time compared to 46% of the male learners. Table 8 displays the employment status of the learners in the sample by gender across all individuals for whom both variables were known. Representing Coursera's global learning community, the individuals in this experiment enrolled from more than 200 countries worldwide. Table 9 displays the number of female and male learners by their home country across the top 100 nations in the sample. This extensive and diverse sample enabled a productive analysis of the implementation at scale.

Figure 12

Age of Learners in the Intervention Sample by Gender

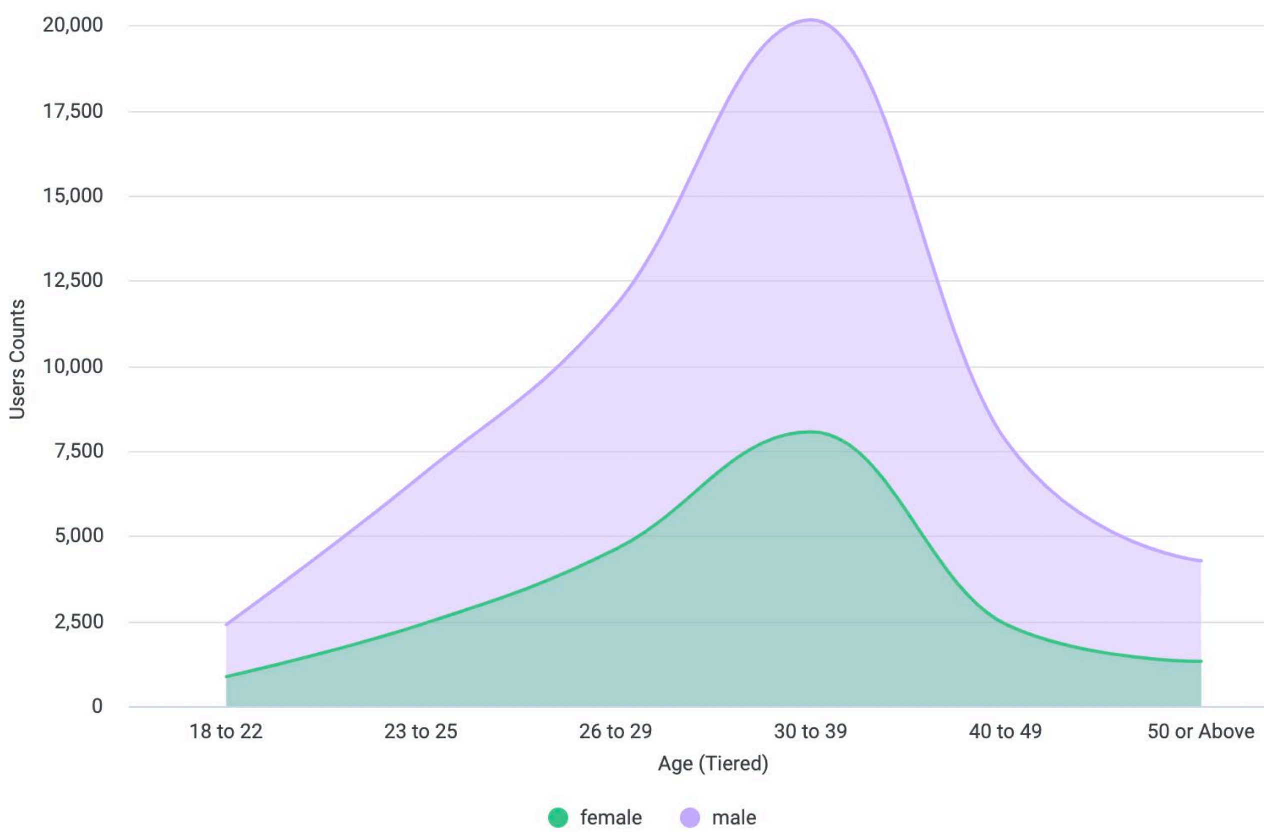


Table 7

Learners in the Intervention Sample by Age and Gender

Age Tier	Female Learners	Male Learners
18 to 22	866	2,394
23 to 25	2,399	6,772
26 to 29	4,618	11,788
30 to 39	8,062	20,165
40 to 49	2,399	7,786
50 or Above	1,316	4,270

Note. Age was not known for all learners in the intervention sample. This table presents the subset of the learner sample for whom age data was available.

Table 8

Learners in the Intervention Sample by Employment Status and Gender

Current Employment Status	Female Learners (%)	Male Learners (%)
Employed Full Time	29,795 (42%)	61,359 (46%)
Unemployed & Looking for Work	15,878 (22%)	30,974 (23%)
Unemployed & Not Looking for Work	13,922 (20%)	22,900 (17%)
Employed Part Time	3,786 (5%)	5,464 (4%)
Self-Employed Part Time	2,099 (3%)	3,418 (3%)
Homemaker	1,959 (3%)	864 (1%)
Self-Employed Full Time	1,589 (2%)	4,219 (3%)
Unable to Work	1,307 (2%)	2,717 (2%)
Retired	750 (1%)	1,434 (1%)

Note. Employment status was not known for all learners in the intervention sample. This table presents the subset of the learner sample for whom employment status was available.

Table 9

Learners in the Intervention Sample by Home Country and Gender

Home Country	Female Learners	Male Learners
United States of America	70,777	81,374
India	29,248	55,150
Canada	9,732	11,528
United Kingdom	7,692	11,145
Brazil	6,522	9,645
Germany	6,371	11,350
Mexico	5,029	6,656
Egypt	4,088	9,561
Turkey	4,052	6,356
Philippines	3,911	4,563
Spain	3,739	6,065
Australia	3,505	4,670
Italy	3,282	5,062
Russia	2,963	3,961
France	2,851	5,427
Netherland	2,631	4,430
Colombia	2,386	3,743
Nigeria	2,337	5,390
Pakistan	2,228	9,678
Singapore	2,196	3,314
Indonesia	2,148	2,437
Argentina	2,113	3,176
Poland	2,083	3,130
China	1,656	1,992
Ukraine	1,638	2,040
United Arab Emirates	1,563	2,864
Hong Kong	1,536	2,509
South Africa	1,499	2,153
Portugal	1,439	1,868
Israel	1,392	3,335
Romania	1,372	1,517
Peru	1,366	2,288
Kenya	1,313	2,311
Saudi Arabia	1,309	2,754
Switzerland	1,243	2,038

Home Country	Female Learners	Male Learners
Morocco	1,231	2,785
Greece	1,223	1,831
Serbia	1,188	1,247
Chile	1,153	1,821
Japan	1,136	2,367
Lebanon	1,046	1,299
Sweden	1,045	1,787
Malaysia	1,021	1,272
Ecuador	895	1,125
Taiwan	894	1,489
Ireland	873	1,333
Vietnam	834	1,515
New Zealand	688	716
Tunisia	671	1,226
Ghana	661	1,769
Austria	631	981
Czech Republic	621	845
Belgium	617	1,303
Denmark	603	1,118
South Korea	601	1,241
Bangladesh	550	2,331
Algeria	546	930
Kazakhstan	529	577
Jordan	508	872
Finland	505	810
Norway	495	889
Thailand	494	756
Croatia	487	551
Bulgaria	478	580
Hungary	458	726
Georgia	450	505
Uruguay	435	643
Armenia	429	367
Belarus	372	459
Lithuania	372	415
Venezuela	369	659
Azerbaijan	341	549
Costa Rica	338	615
Latvia	333	277

Home Country	Female Learners	Male Learners
Nepal	327	1,031
Qatar	314	503
Guatemala	310	436
Estonia	307	300
Sri Lanka	307	511
Dominican Republic	299	407
Ethiopia	273	1,074
Panama	260	332
Trinidad and Tobago	255	179
Slovakia	238	317
Albania	234	177
Bolivia	232	419
Oman	223	310
Cyprus	221	284
North Macedonia	208	213
Iraq	204	574
Honduras	201	246
Myanmar	196	228
El Salvador	186	271
Moldova	182	164
Paraguay	182	223
Kuwait	169	297
Jamacia	163	135
Palestine	161	319
Uganda	161	369
Slovenia	156	223

Note. This list is limited to the top 100 countries as ordered by female learners enrolled in the intervention study. A total of 216 countries had at least one female and one male learner in the full sample. This table displays only learners from the sample for whom home country information was known.

As outlined in the design plan, the outcome metric questions were intended to examine only those who had started the in-course learning activities. Specifically, the researcher narrowed the pool to only active learners, meaning individuals who had started at least one learning item in their enrolled course. This narrowing resulted in 131,804 active female and 192,653 active male learners. This active learner sample of 324,457 individuals was used to calculate first-week completion, course completion, and further MOOC enrollments. Final grade achieved was calculated by averaging across only those who completed the STEM MOOC they had enrolled in as part of this experiment.

Measures

Building on this study's research questions and logic model, the researcher identified specific metrics to assess the relevant constructs. For this intervention centered on better supporting females in STEM MOOCs on Coursera, the process evaluation assessed reaching the target audience and providing an intervention viewed as helpful by learners. The outcome evaluation examined the impact on learners' persistence, skill development, inactivity reasons, and continued learning. This section documents each measure's purpose, utility, and details, organized by process and outcome evaluation categories.

Process Evaluation Indicators

After exploring which process evaluation components would be most fruitful for this study, the researcher focused on the specific measures required to investigate these questions. Each of the following headers is a particular indicator to assess a construct of interest. Within each subsection, the researcher explains how the relevant data was collected and how the analysis plan to assess these measures throughout the implementation was conducted. To summarize the process evaluation protocol, the researcher created a detailed matrix outlining the questions, components, indicators, source, collection, and analysis plans. Appendix D shows this matrix and highlights the process section of the broader evaluation plan for this intervention study.

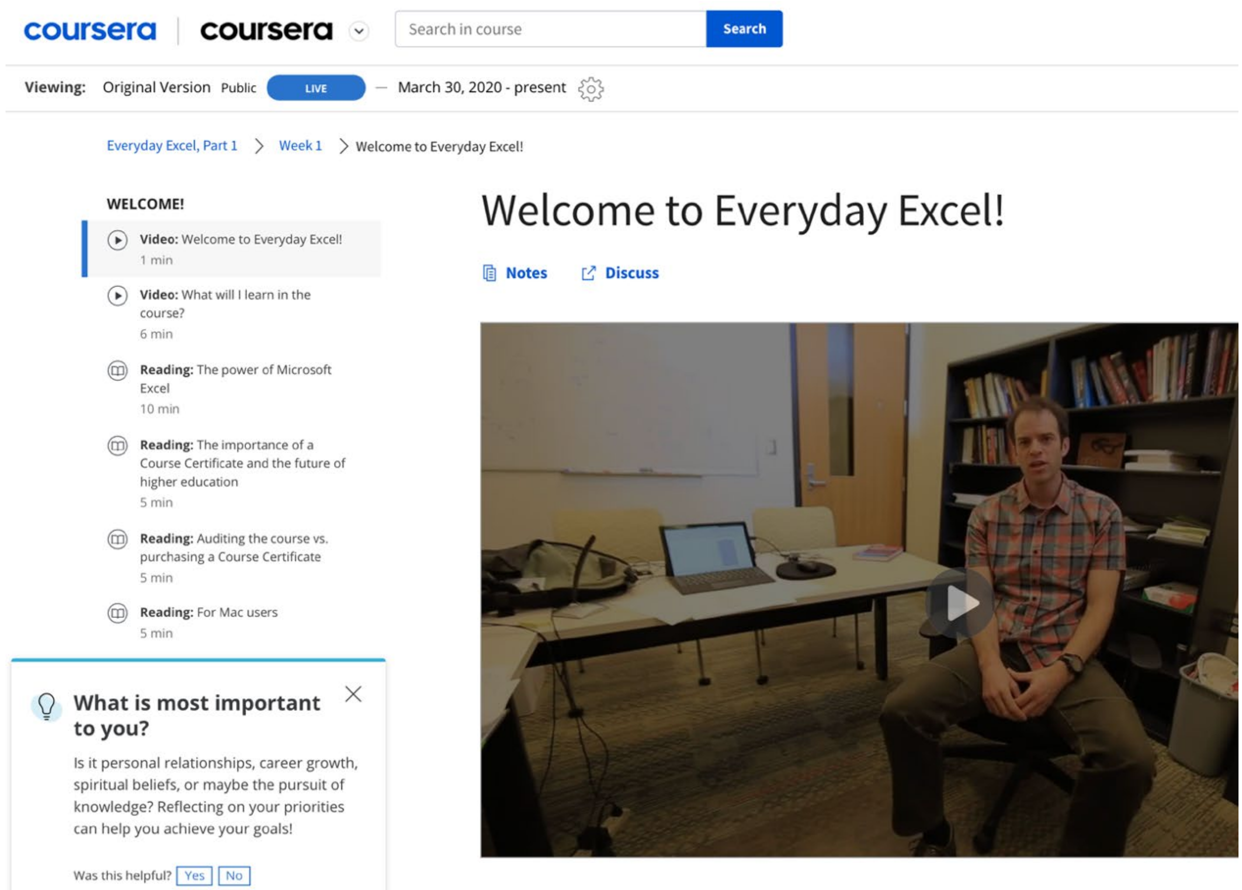
Learners Reached. The first key area of process evaluation was to ask if the intervention reached the appropriate audience (Rossi et al., 2019). This component was defined as the percentage of participants aligning with the intended target group (Baranowski & Stables, 2000). For this intervention, the different in-course messages were randomly displayed to certain subsets of newly enrolled learners in the selected STEM MOOCs. The number of female learners in each intervention group was determined randomly by the backend A/B testing system as learners decided to enroll in these participating STEM MOOCs on Coursera. Thus, the researcher could not limit this intervention to only female learners, the primary intended audience, and needed to ensure a meaningful percentage of learners in the sample were female. In addition, previous research suggests these interventions can help both male and female learners from developing countries as well as those working full time, suggesting benefits beyond the target group (Wang & Baker, 2015; Yeomans & Reich, 2017). The researcher calculated the percentage of females in each treatment group compared with the control to examine if the intervention reached sufficient female learners in STEM MOOCs. The learners' demographic information was collected and used for this calculation.

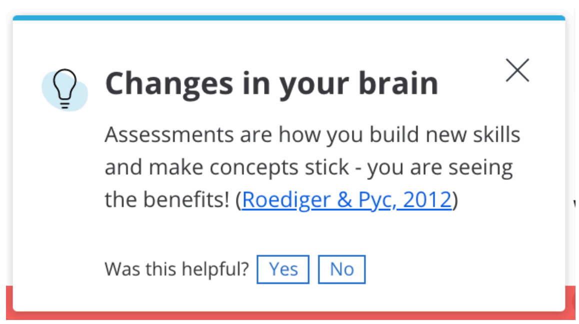
Perceived Helpfulness. The exposure component encompasses how a participant interacts with and responds to an intervention (Saunders et al., 2005). Baranowski and Stables (2000) define exposure as participants' component preference and observed utility of the intervention. To include participants' voices in this experience, the researcher explored learners' responses to the intervention and how enjoyable or valuable they found it (Baranowski & Stables, 2000; Saunders et al., 2005). In this study, individual learners provided a helpfulness rating after seeing each prompted activity, indicating their preference. Not all learners provided their answer to the helpfulness question, as the pop-up message could be easily dismissed or ignored if the individual desired.

Given the large participant group for this study, the researcher aimed to summarize this sentiment by examining learners' responses in aggregate. Specifically, learners were asked, "Was this helpful?" (see Figure 13) as part of the pop-up message. Individuals had the opportunity to respond directly within the STEM MOOC learning experience on Coursera, and the backend platform saved their answers. These responses were averaged across learners and the four or six total messages per treatment group to calculate an average helpfulness rating for each intervention variant. This method of collecting feedback ensured the implementation process worked as planned, acting as a user-level verification that the intervention was displayed as expected to real learners. Additionally, this direct learner feedback on the utility of the current in-course interventions helped measure success and iterate toward more useful solutions.

Figure 13

Intervention Message Examples in Live Course on Coursera





Note. These screenshots depict two examples of the intervention in-course messages live on the Coursera platform. The first displays the full-screen view of what a learner would see when this message pops up in the first item of the course. The second is an enlarged view of a message that contains a clickable hyperlink to the relevant research. Both include the helpfulness question.

Outcome Evaluation Indicators

After evaluating the process of implementing this intervention, the researcher concentrated on assessing the outcome variables of interest. The constructs in outcome-focused research questions needed to be operationalized into valid and reliable measures aligning with the study's learning goals (Moulton, 2014). These constructs included persistence, skill development, and continued learning beyond the course. The alignment between constructs and variables needs to occur before the experiment starts to ensure all stakeholders agree on the standards of quality and success (Rossi et al., 2019; Wholey et al., 2010). Appendix E highlights this study's overall outcome evaluation plan, documenting how these outcome metrics relate to the research questions and constructs of interest.

Week One Completion Rates. For all active learners in the course, the researcher calculated the number of learners who completed the first week of material, including all assessments. This "active learner" was the same as defined in the needs assessment, indicating a learner who has engaged with at least one item in the course. These learners must also be eligible to complete the course, meaning they have access to all graded assignments. Thus, this study was limited to paying and financial aid learners, all of whom have this eligibility. The average week one completion rate was calculated for the sample of learners in each treatment and control group as well as analyzed by gender.

Course Completion Rates. To assess longer-term persistence, the researcher used course completion to measure systematic engagement across several weeks. An individual learner "completes" the course by passing all graded assessments. Consistent with the week one completions, the average course completion rate was also calculated for each treatment and control learner group in the sample as well as by gender. These averages helped summarize information from the large number of learners in the study and assess wide-scale impact. Only active learners and those paying or on financial aid were included in this calculation.

Skill Development. The researcher measured skill development using Coursera's novel technique to determine learning gains. While researchers often estimate skill development at the country level and use broad survey measures (OECD, 2015), data scientists and learning experts at Coursera calculate skill development at the individual level and on the scale of days instead of months or years (Hickey et al., 2020). This algorithmic approach uses previous assessment performance to determine an individual learner's current skill level in a particular domain, such as programming or calculus. This skill score shows a learner's likelihood of passing an assessment in that domain, meaning a learner with a skill score of two is twice as likely to pass a given test than a learner with a skill score of one.

Coursera's novel approach to measuring skill development systematically for millions of learners required innovative approaches and several years of development. On the Coursera platform, a learner's skill score is calculated using the same algorithms often employed in chess competitions or sports tournaments (Reddick, 2019). Data scientists at Coursera can calculate a reliable estimate of both the learner's ability and the assessment's difficulty level by coding each assessment attempt as a single game or sporting match. This approach was made possible by the extensive data from the Coursera platform containing tens of millions of assessment attempts across thousands of courses. Furthermore, previous applications of this algorithmic process provide evidence of its validity (Reddick, 2019).

When taking courses, learners' skill scores can increase or decrease depending on their performance on the exams, assignments, and projects in that course. If a learner with a low skill score passes a challenging graded assessment, their skill score will increase significantly; if a learner with an already high skill score passes a difficult graded assessment in that same domain, their skill score will increase minimally or not at all (Hickey et al., 2020). As learners practice and complete more graded assessments, their skill score tends to increase, signifying their learning in this subject area (Reddick, 2019). For this outcome evaluation, the researcher calculated the average final grade achieved by course completers across treatment groups and genders. In-course performance acts as the lever by which skill scores update, so this average grade achieved could act as a proxy for skill development, since most learners on the Coursera platform do not yet have adequate learning data to provide an individual skill score in each domain (Hickey et al., 2020). With learners assigned randomly across treatment groups, the previous skill level of learners was expected to average out across groups.

Inactivity Survey. As used in the needs assessment study, this intervention study leveraged insights from Coursera's existing Inactivity Survey. If a learner has not been active in a course for three weeks, Coursera sends a brief survey by email to ask a few short questions on why they have not continued in the course. Any given learner receives this survey at most once, even if enrolled and inactive in several courses. These survey responses provide both multiple-choice answers and open-text answers, the latter of which was used as the qualitative data for this study. For this indicator, only female learners' responses from this survey during the time bounds of the intervention study were reviewed.

Subsequent Course Enrollments. The researcher examined the average number of enrollments in other MOOCs after joining this experiment to measure longer-term engagement and commitment to skill development. This metric was assessed one month after the experiment closed and included any course enrollments on the Coursera platform by a learner after their initial enrollment in the STEM MOOC in this experiment. This approach added a month between when learners could last join the experiment and the measurement of learners' subsequent course enrollments. This delay allowed the researcher to calculate the average number of courses the learners had engaged with on the Coursera platform after their initial engagement with the STEM MOOC in this experiment. As with previous calculations, the researcher filtered learners only to those who have both enrolled and actively engaged with at least one item in the experiment course—in other words, an active learner, as defined earlier. Furthermore, narrowing subsequent MOOC enrollments to look specifically at other STEM courses on Coursera aligns with this study's overall goal of furthering scientific and technical learning.

Procedure

This implementation focused on integrating research-based interventions within the online course learning experience. As a Silicon Valley education technology company, Coursera already has a strong data science team and has used machine learning to improve course recommendations, learner onboarding, and platform functionality (Hickey et al., 2018; Reddick, 2019; Urban, 2019). This intervention built on these innovations to surface time-relevant, in-course pop-up messages aligned to the three areas identified during the recent needs assessment study of low confidence, significant time constraints, and challenges from gender-imbalanced countries. The following section explores the specific design of each intervention group and the data collection and analysis across these groups.

Participant Recruitment

Learners were randomly assigned to the treatment or control groups as they enrolled in one of the selected STEM MOOCs on Coursera. The courses in this experiment included 150 STEM MOOCs currently live on the Coursera platform, as ordered by active learner enrollments after removing partners that did not wish for any experimentation in their courses. The time bounds of this intervention were from December 8, 2021, to March 20, 2022, meaning all new enrollments by adult learners during this period were automatically added to the experiment. The backend of the Coursera platform randomly assigned these learners to one of five groups: the four different treatments or the control group, which each contained approximately 20% of learners in the sample by the end of the experimental period.

Coursera's platform enabled this online, RCT-style experiment. The random assignment was completed using Coursera's A/B testing system, a standard online learning approach for testing new product features and content enhancements (Kizilcec & Brooks, 2017; Urban & Greenblatt-Kolodny, 2017). Thus, learners did not know which group they had joined and did not have an opportunity to switch groups. The researcher also did not have visibility into or control over which learners were assigned to the five different groups of this RCT. During this experimental period, all learners who enrolled in the participating STEM MOOCs on the Coursera platform automatically became a part of this study's sample.

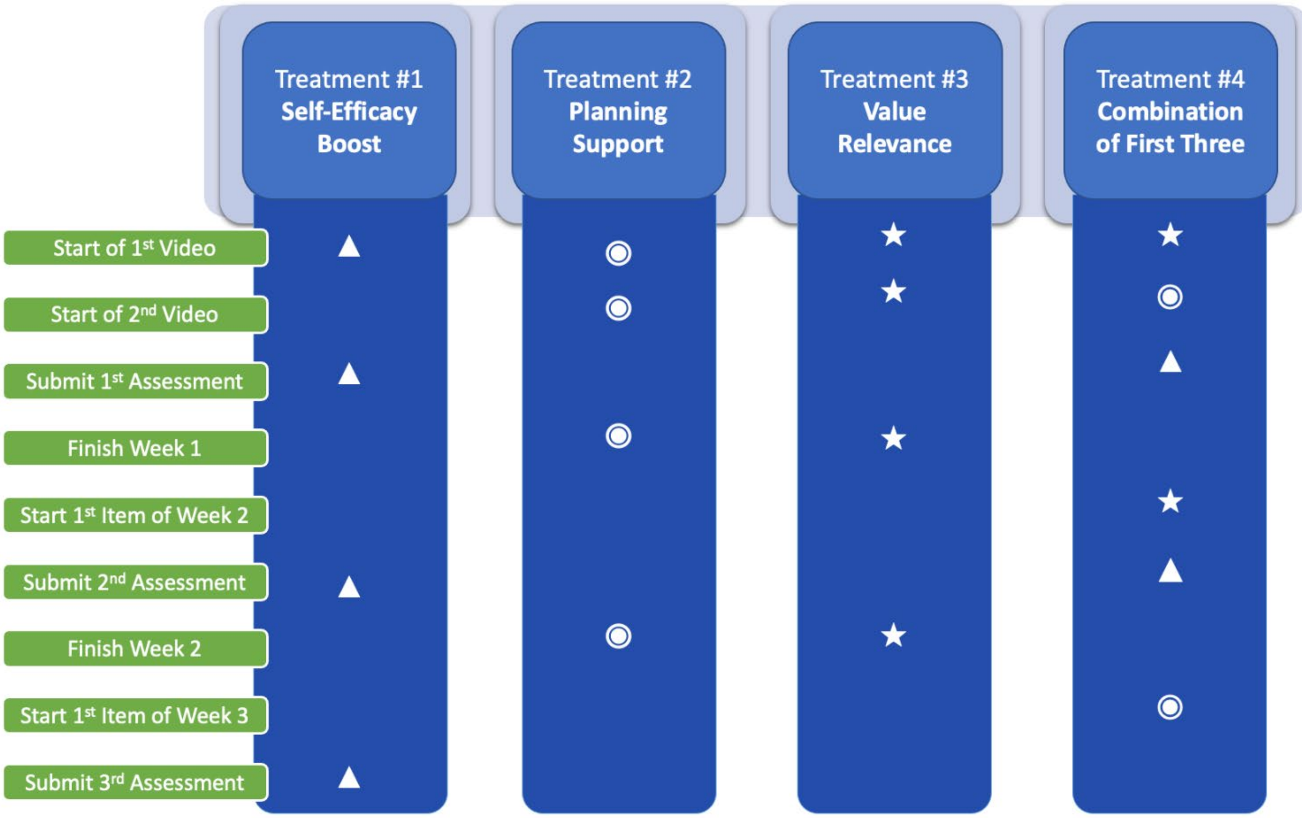
Intervention

This researcher divided newly enrolled learners into five intervention groups within each STEM MOOC. The first four groups (self-efficacy, planning, value connection, and a combination of the first three) aimed to assist the main areas identified in the recent needs assessment. The fifth acted as the control group with no new enhancements added. These intervention variants were designed to be as light-touch as possible to avoid interrupting each course's learning materials and activities. Across all five groups in the sample, the coursework looked the same for each given STEM MOOC. Thus, the videos, readings, quizzes, projects, and discussion forums were consistent across each STEM MOOC regardless of the treatment or control group in which a learner was placed.

After learners enrolled, they were shown prompted messages, each based on self-determination theory (SDT) literature, or nothing if they were in the control group. In each of the first three intervention designs, four text-based, in-course messages were surfaced directly to learners at similar milestones throughout the first three weeks of the online course in which they had enrolled (see Figure 13 for examples). In the fourth version, the combination of the previous three intervention approaches purposely offered an increased treatment strength with six total in-course messages. Each pop-up included a title, a text-based message, a clickable link to the research when appropriate, and a quick helpfulness question. These milestones were broad enough that all 150 STEM MOOCs included in the study had each of the following trigger moments for the planned messages. However, in a small subset of courses, these milestones were in a different order than appears in the diagram below—for example, when there were two graded assessments in the first week of content. Figure 14 summarizes the type and timing of these in-course messages across the different treatment groups.

Figure 14

Timing of In-Course Messages Across Intervention Treatment Groups



Note. Each icon indicates a pop-up message was shown to the learner at that milestone of the course. The triangles represent self-efficacy messages, the circles correspond to planning support, and the stars show value relevance prompts. The control group is not included in this summary since no messages were shown in that variant of each course.

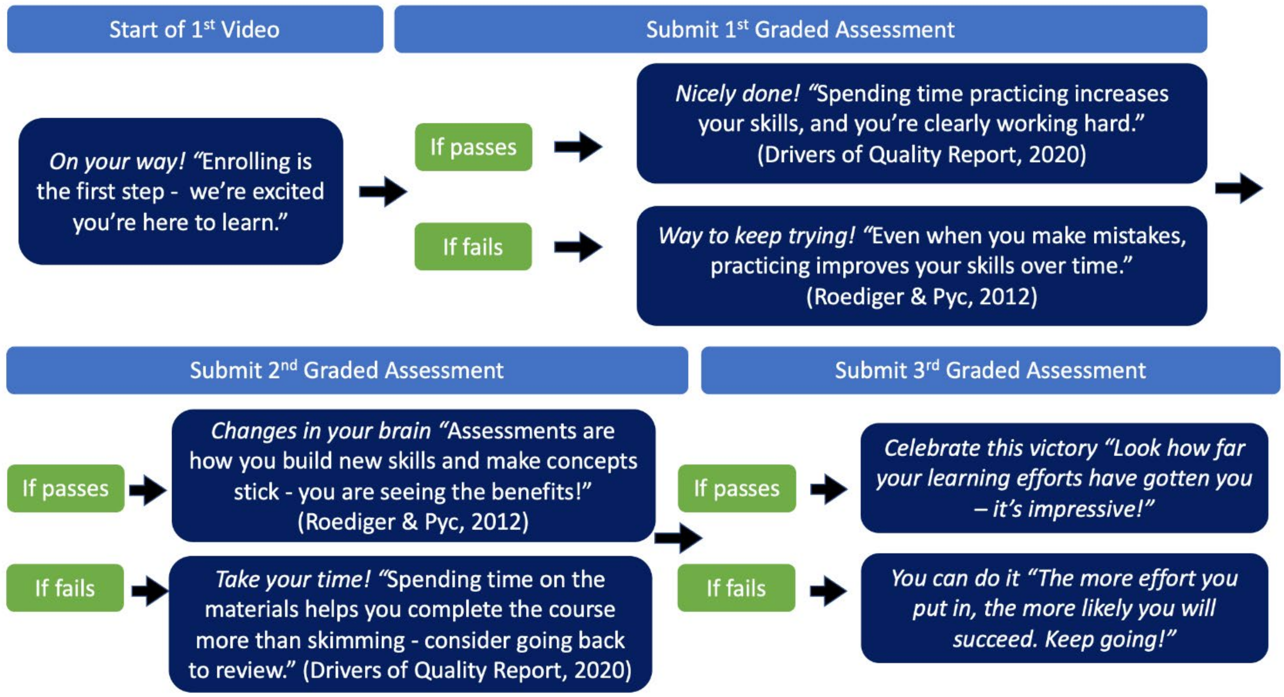
Learners, as they desired, engaged with each prompt that appeared and indicated the helpfulness of these pop-up messages. The overall design assumed a substantial number of learners continued to enroll in these STEM MOOCs on Coursera and that at least a meaningful subset would engage with the pop-up prompts. The following sections explore the differences unique to each of the five groups, connecting to the literature review of solutions in the previous chapter.

Self-Efficacy Treatment. Learners in this treatment group received new messaging after entering the course and submitting the first three graded assessments (quizzes, programming assignments, or peer-review assignments) of the STEM MOOC in which they enrolled. In addition to the existing feedback supplied by the Coursera platform on their grade received, learners in the self-efficacy variant were shown messages rooted in all they have already accomplished or how slightly more effort and review could help them succeed. These messages differed depending on if the learner passed or failed each given assessment.

This treatment applied learnings from Huang and Mayer's (2019) experiment to test the utility of self-efficacy-boosting messages in online statistics courses. Consistent with these researchers' findings, this intervention's messaging emphasized students' already demonstrated efforts and accomplishments while also offering encouragement in the form of praise to increase their self-confidence and persistence (Huang & Mayer, 2019). When learners enrolled, they were congratulated on taking the first step to increase their learning. Then, after each of the first three graded assessments, students received automated messages depending on their performance. Figure 15 summarizes the progression through the MOOC with corresponding trigger points and in-course statements for this treatment group.

Figure 15

Intervention Points and Messages for the Self-Efficacy Boosting Treatment



Note. The text in italics within the visualization above maps to the title when the message is shown to learners. The citations in parentheses were also hyperlinked in the final messages shown on the Coursera platform. These same specifics apply to Figures 16, 17, and 18 as well.

Planning Treatment. Learners in the planning treatment group also received in-course pop-up messages throughout the first weeks of their MOOC experience. The initial prompt requested that individuals reflect on where and when they will make time to progress in this content, just as other researchers have found helpful for MOOC learners (Yeomans & Reich, 2017). As learners finished items in the course and completed the first few weeks, this treatment provided messages emphasizing the importance of planning and making learning a habit in their lives.

The researcher aimed to combine insights from other settings and previous Coursera trials into the proposed design. For example, highlighting data on how certain practices can increase their likelihood of success has been beneficial in other in-course interventions previously tested on Coursera (Hickey et al., 2018; Urban, 2019). This intervention variant also congratulated learners on their successful planning as they progressed through the course material. This praising message leveraged other researchers' insights on how adding planning support can increase learners' feelings of confidence and competence (Lung-Guang, 2019; Sambe et al., 2017). Figure 16 outlines when and what messages were surfaced to learners in this planning treatment group of each course.

Figure 16

Intervention Points and Messages for the Planning Treatment

Value Relevance Treatment. In the third variant of this intervention, learners received in-course messages prompting reflection on their values and goals. Just as previous researchers have tested in other course settings, the researcher started by surfacing specific value areas and asking learners to reflect on which is most important to them (Kizilcec, Saltarelli, et al., 2017; Miyake et al., 2010; Peters et al., 2017). As learners continued to progress in the MOOC, they received encouragement and reminders related to their values and goals. Reminding learners of their values during the course journey has proven helpful in STEM content, especially for females (Peters et al., 2017) and those from less developed countries (Kizilcec, Saltarelli, et al., 2017). Figure 17 shows these value relevance messages and when they appeared for learners throughout the STEM MOOC learning journey.

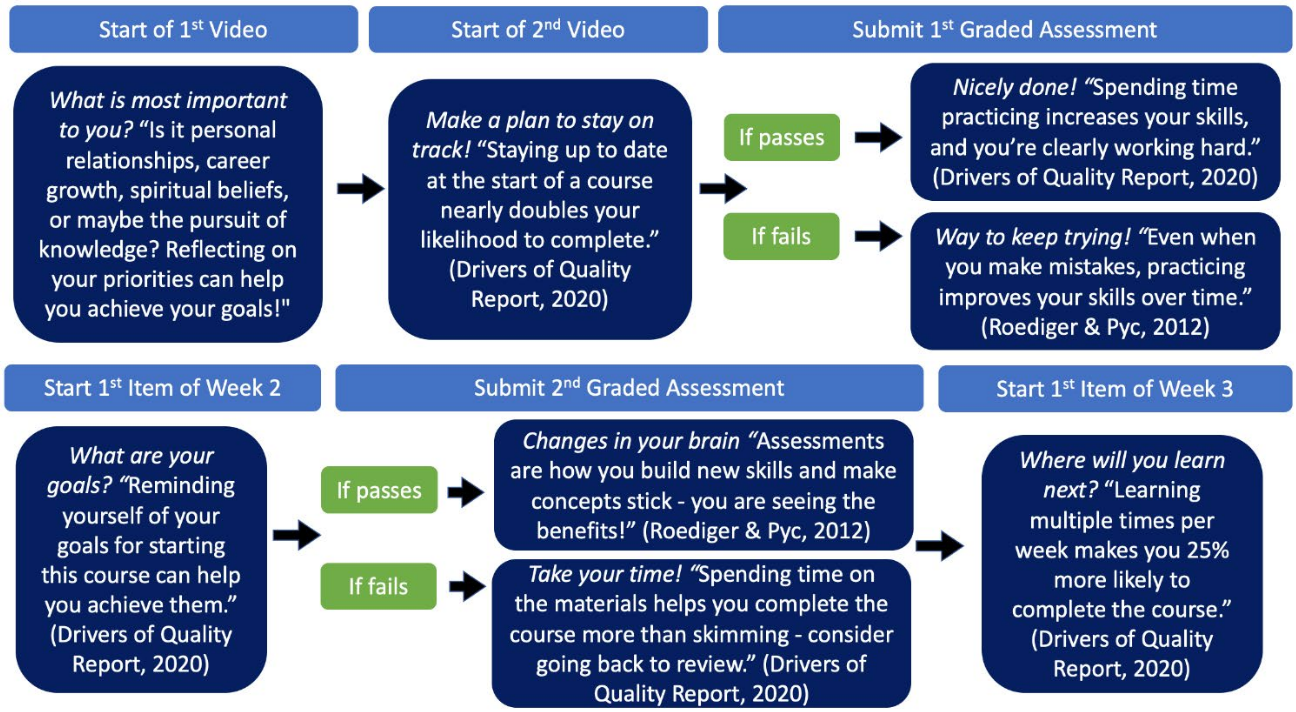
Figure 17

Intervention Points and Messages for the Value Relevance Treatment

Combination Treatment. The fourth treatment group combined prompts from the first three intervention groups to test both a stronger overall treatment (i.e., more messages) and if these interventions could be beneficial when presented together. For this fourth version, the course had six in-course messages with two prompts taken directly from each of the first three variant conditions. Cognitive scientists warn educational researchers about overwhelming students with too much to process (Moulton, 2014), and some investigators have even found additional items to process linked with lower self-efficacy in STEM undergraduate courses (Feldon et al., 2018), which would be counterproductive to the goals of this study. However, just-in-time nudges can be useful to learners, particularly in online settings (Hickey et al., 2018; Yeomans & Reich, 2017). Thus, this combination group allowed the testing of a stronger treatment of messages in order to assess if the benefits outweigh the negatives when surfacing these prompts together in a single course experience. Figure 18 presents this combination of messages for the final intervention treatment.

Figure 18

Intervention Points and Messages for the Combination Treatment



Control. Learners in the control group saw the same course experience as learners before this experiment was implemented. These learners in the control sample did not see any of the prompted messages designed for this intervention. Additionally, all other in-course messages from other Coursera experiments were turned off during this experiment to isolate these SDT messages as the only systematic change between intervention groups. Each STEM MOOC included the same coursework across all treatment and control groups as well as the typical functionality within the Coursera learning experience. As is recommended for RCT designs, the control group provided an identical experience besides the intervention variants described above for the four treatment groups, enabling a more successful causal impact analysis (Rossi et al., 2019; Shadish et al., 2002). Thus, learners in the control sample saw the same STEM MOOC experience on Coursera as before this experiment began, without any of the above intervention messages shown.

Data Collection

The explanatory sequential mixed-methods design guided the type of data collected and the following analysis plan. First, the researcher utilized quantitative indicators to summarize the effectiveness of this large-scale intervention. The process-focused, numeric indicators centered on which learners were reached and their perceived helpfulness of the intervention. Second, to ensure implementation fidelity, the researcher collected descriptive statistics monthly after the experiment began to monitor if the intended audience was reached and if learners in each treatment group reported the intervention as helpful.

All data were collected through the Coursera online learning platform. Learners' RCT participant assignment, behavior in the courses, and survey responses were saved automatically in the backend of the Coursera system. The researcher had access to these actions and answers only in aggregate without any personally identifiable information (PII). This data collection plan aligns with the Johns Hopkins University Institutional Review Board protocols and falls within Coursera's terms concerning educational research (see Appendix G).

Data Analysis

Process Evaluation Data Analysis. The researcher analyzed each process evaluation metric one month after implementation and monthly thereafter as well as at the end of the study. Given the large expected sample size and RCT design, the percentage of females in each group was not expected to differ significantly. Descriptive statistics summarized the reach and exposure process evaluation components within each intervention variant. To test the learners reached, a two-way analysis of variance (ANOVA) statistical test was used to check for any association between the intervention groups and gender after the final learner sample had been assembled. A two-way ANOVA test was also used for the perceived helpfulness indicator to reveal any differences by treatment group or gender. These process indicators were monitored monthly to determine if the intervention prompts required any adjustment and test if the A/B testing system and messages worked as projected.

Outcome Evaluation Data Analysis. To investigate the outcomes of this intervention, the researcher started by assessing the quantitative impact of the different intervention variants on learners' persistence in these STEM MOOCs. These results were compared with those of female learners in the control group using a two-way ANOVA difference of means statistical test. Further analysis by age, employment level, and home country were also conducted to assess the impact on learner subgroups. Second, the average achieved course grades across treatment groups were analyzed, given this in-course performance is how skill scores are updated on the Coursera platform, using a systematic skills framework (Hickey et al., 2020; Reddick, 2019). Third, ANOVA statistical tests determined any significant differences among learners' persistence and skill development in each intervention group by gender, aligning with RCT quantitative impact evaluation norms (Creswell & Plano Clark, 2011; Rossi et al., 2019). Finally, ANOVA tests were used to analyze the average subsequent STEM course enrollments by treatment group and gender.

To complement the quantitative analysis, the researcher qualitatively coded female learners' self-reported reasons for dropping out. The complete list of themes from the needs assessment qualitative analysis was used as the a priori list of codes for this study to enable more precise comparisons with the earlier samples (Hsieh & Shannon, 2005; Lochmiller & Lester, 2017). This list included the themes of no time, content problems, lack of confidence, and missing prerequisite knowledge. Qualitative analysis of these written responses and "quantizing" these verbal data into percentages by theme clarified quantitative trends from the earlier research questions (Sandelowski, 2000). These outcome measures were assessed one month after learners could no longer join the cohort experiencing the RCT-style intervention. When the experiment closed, no new learners could enroll in the treatment versions of the MOOCs, but those already enrolled could continue progressing and retain access to the experiment messages. This additional month between closing the experiment and conducting the outcome analysis allowed sufficient time for learners to progress through the full course material or report their reason for inactivity if they decided to drop out before completing.

Appendix E presents how the outcomes of this study will be evaluated, including the questions, measures, data collection, and analysis details. Throughout both the quantitative and qualitative portions of this outcome evaluation, the researcher could compare individuals in the intervention and control groups to learners in the previous needs assessment sample from Chapter 2. Overall, this numeric-first approach aligned with the explanatory sequential mixed-methods design, the extensive data available, and the research questions posed (Creswell & Plano Clark, 2011; Onwuegbuzie & Leech, 2006).

Conclusion

This implementation study tested the utility of adding brief, in-course messages to STEM MOOCs to improve female learners' persistence and performance. The research questions outlined the process and outcome components to assess, ensuring that this intervention's implementation and results aligned with expectations. The study used an explanatory sequential mixed-methods design, emphasizing the quantitative power of an RCT setup with follow-up qualitative analysis to explore learners' personal narratives. The study design included five experimental groups: the self-efficacy boost, planning support, value relevance connection, combination of the first three, and control. These different variations of the experiment tested the utility of augmenting the three primary psychological needs of the SDT framework with a lightweight, fully automated intervention to increase females' motivation. The process components focused on reaching the desired target group and exposing learners to prompts they found helpful. The main outcome components centered on learner persistence, skill development, inactivity reasons, and subsequent enrollments in the STEM MOOCs.

This RCT-style experiment took place from December 8, 2021, to March 20, 2022, in 150 of the highest enrollment STEM MOOCs on Coursera. Through a robust A/B testing system, the Coursera platform randomly assigned newly enrolled learners into the five different versions of each MOOC. The researcher collected the data, including learners' behavior and responses, directly from the Coursera platform. This quantitative-focused study utilized statistical tests, such as two-way ANOVA, to assess any significant differences among the treatment and control groups. Qualitative coding by theme was used for the open-text answers from the learner inactivity survey. Insights from these various data sources and methods were then combined in a final analysis to assess the utility of these interventions to support female learners in STEM MOOCs. The overall goal of this study was to produce results meaningful for the learners in this experiment and for others desiring to increase retention, skill development, and further learning in their own educational settings. Chapter 5 presents the complete results of this experiment and the insights from this study for future iterations.

Chapter 5: Results and Discussion

The purpose of this chapter is to reflect on the implementation and impact of the intervention study. As described in the previous chapter, this study aimed to empower female learners in STEM MOOCs on the Coursera platform through brief, in-course messages designed to boost their self-efficacy, planning focus, and value connection. Through learners' random assignments to the control or treatment groups, this experiment provided the opportunity to draw causal conclusions on any observed outcome differences. The researcher reports results from the two process indicator questions and four outcome questions of interest. After the research questions are discussed, the remainder of this chapter covers the conclusions, insights, and implications of this study. Future directions are also considered to further this research and continue empowering female learners in online courses worldwide.

Focus of Findings

From the initial group assignment and helpfulness reported via persistence and performance, the research questions for this experiment focused on the accuracy and efficacy of the intervention design. This section explores the results of each question, examining both the process and outcome indicators for this study. The following research questions guided the data collection and analysis for this experimental research study:

RQ1. To what extent did the intervention reach the target learner group?

RQ2. To what extent did learners find the prompt helpful?

RQ3. What differences in impact did each intervention have on week one and course completions?

RQ4. What differences in impact did each intervention have on course completers' performance and skill development?

RQ5. How did the intervention affect female learners' self-reported reasons for dropping out of the STEM MOOCs for those who did not complete them?

RQ6. To what extent did the intervention spark learners to continue learning in other STEM MOOCs?

Results

This section explores the findings related to each research question. Given the hundreds of thousands of learners in the sample, summarizing statistics were used throughout this analysis. Additionally, *t*-tests were used to determine statistical significance between a pair of variables, and the results of these tests are reported in this section. The error bars on the graphs in this chapter show a 99% confidence interval for the true value of each metric. Thus, these error bars represent visually the same information as two-sample independent *t*-tests conducted at the 0.01 alpha level. To determine statistical significance among the tables of variables, the researcher utilized two-way analysis of variance (ANOVA) tests with gender and intervention group as the two independent variables. The key findings from these tests are presented in the chapter text, and the complete ANOVA test results are available in the Appendices.

Process Evaluation

Throughout the implementation of this intervention study, the researcher tracked process indicators to ensure a successful deployment of these novel in-course messages. As described in previous studies, process evaluation provides evidence regarding how a study is implemented and construed, offering early feedback on a project's design (Saunders et al., 2005). Process evaluation indicators can also aid in causal conclusions during the outcome evaluation stage by revealing insights about mediating variables (Baranowski & Stables, 2000). Given that the study sample included thousands of learners in each treatment group, the researcher assessed reach and exposure indicators using quantitative summarizing metrics. This section explores the specific answers to the process research questions posed.

RQ1: Reach

Before testing whether the intervention had the desired results, the researcher examined whether the intervention reached the desired target audience. The reach question for this study focused on assessing the percentage of female learners in each treatment and control group, ensuring that there were an adequate number of females and no significant differences in this percentage among the groups. Table 10 summarizes the quantitative indicators for the number of female and male learners reached by each experimental intervention condition throughout the implementation.

Table 10

Learners in the Experiment by Intervention Group and Gender Over Time

	Control	Self-Efficacy Boost	Planning Support	Value Connection	Combined
ONE MONTH					
Female Learners	16,435	16,203	16,334	16,314	16,377
Male Learners	27,078	26,621	26,603	27,056	26,933
Percentage Female	37.8%	37.8%	38.0%	37.6%	37.8%
TWO MONTHS					
Female Learners	32,081	31,816	31,931	31,757	31,986
Male Learners	49,382	49,299	48,908	49,299	49,145
Percentage Female	39.4%	39.2%	39.5%	39.2%	39.4%
THREE MONTHS					
Female Learners	42,853	42,704	42,640	42,601	42,676
Male Learners	64,735	64,601	64,419	64,803	64,439
Percentage Female	39.8%	39.8%	39.8%	39.7%	39.8%
FINAL RESULTS					
Female Learners	48,762	48,578	48,589	48,420	48,498
Male Learners	73,683	73,375	73,284	73,922	73,159
Percentage Female	39.8%	39.8%	39.9%	39.6%	39.9%

Note. Courses, n = 150

As shown in Table 10, the overall percentage of females in each intervention are similar over time and between groups. This ratio results in a final distribution of females in the 39.6% to 39.9% range across intervention groups. The backend randomization process occurred as designed to create an RCT experiment with approximately equal-sized groups. Since this backend system was randomizing at the learner level and blind to gender during group assignment, the researcher wanted to ensure approximate consistency in the female-to-male balance across the intervention sub-groups. This aspect of the implementation relied on the law of large numbers, meaning that, given the large number of learners in the experiment, the percentage of females in each treatment should closely approximate the more stable population percentage of females enrolling in the selected 150 STEM MOOCs over time (Salkind, 2010).

A two-way ANOVA test did not detect any meaningful differences in the gender balance of these groups ($p = 0.093$, see Table 11), indicating insufficient evidence to suggest any meaningful differences in these groups' gender ratio or size. There was, as expected, a statistically significant difference in the number of males versus females in each treatment group ($p < 0.0001$, see Table 11). However, the overall sample portion that was female (nearly 40%) was a higher percentage than in the needs assessment presented in Chapter 2, where females comprised less than 30% of STEM enrollments. This larger proportion in the intervention study shows how female learners were especially interested in the 150 top-performing STEM MOOCs included in the experiment, accounting for a more substantial share of the learner cohort than in the larger sample of 2,300 STEM MOOCs explored earlier.

Table 11

ANOVA Test Results for Gender of Learners by Intervention Group

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Rows	4557114867.60	1	4557114867.60	377898.423	0.000	21.198
Columns	208633.60	4	52158.40	4.325	0.093	15.977
Error	48236.40	4	12059.10			
Total	4557371737.60	9				

RQ2: Exposure

After determining that the intervention reached the desired target audience, the researcher assessed learners' responses to the different treatments. Each new in-course message was designed to include a "yes or no" question of helpfulness within the same pop-up screen to investigate whether they found the message useful. This quick question gave participants an immediate and straightforward opportunity to indicate their perceived usefulness of each message they encountered. With hundreds of thousands of learners in the sample, summarizing sentiment across individuals seemed a more fair and effective method for understanding the overall exposure of this intervention while retaining the purpose of this process indicator (Baranowski & Stables, 2000).

The percentage of respondents who found the message helpful was calculated for each message, and this calculation was restricted to only those learners who responded to the helpfulness question. This percentage was then calculated for each treatment group, summing across all messages in that intervention. Each treatment group demonstrated a helpfulness average across messages of 86–92%, which was higher than predicted. As hoped, the overall average helpfulness indicated by female learners was similar to or higher than the average of the overall learner sample. This finding is also reflected in the specific helpfulness rating for each message within the different treatments (see Appendix H).

As a process indicator, it was useful to examine the helpfulness rating throughout the implementation of this intervention study. Table 12 displays the helpfulness ratings of all learners and females in each treatment group over time. For a learner who enrolled within the first month of the intervention and responded to the helpfulness question on the first item prompt, her "yes" or "no" was included in the first section of Table 12 under "One Month." For that same learner, if she returned to the course during the second month of the intervention and responded to the helpfulness question on a later prompt, her new response was recorded under the "Two Months" header. These different sections of Table 12 are cumulative, with the "Final Results" section including all responses given to the helpfulness question through the implementation of this intervention.

Table 12

Learners' Reported Message Helpfulness by Treatment and Gender, Over Time

	Self-Efficacy Boost	Planning Support	Value Connection	Combined
ONE MONTH				
% of Total Report Helpful	90.8%	86.3%	88.4%	86.6%
learners, n = 64,277				
% of Females Report Helpful	91.2%	86.8%	89.8%	86.4%
learners, n = 15,817				
TWO MONTHS				
% of Total Report Helpful	91.0%	86.5%	88.7%	86.7%
learners, n = 95,778				
% of Females Report Helpful	91.4%	86.7%	89.8%	86.3%
learners, n = 19,962				
THREE MONTHS				
% of Total Report Helpful	91.2%	86.4%	88.6%	86.8%
learners, n = 135,807				
% of Females Report Helpful	91.8%	86.6%	89.4%	86.3%
learners, n = 28,569				
FINAL RESULTS				
% Total Report Helpful	91.2%	85.6%	88.5%	86.9%
learners, n = 160,999				
% of Females Report Helpful	91.7%	86.7%	89.4%	86.6%
learners, n = 33,680				

Note. Courses, n = 150

Across the millions of previous messages shown with this same pop-up system on the Coursera platform, an average helpfulness rating of 70–75% had been observed when the messages were randomly shown to different users (Hickey & Urban, 2019). This average helpfulness rating across messages tends to increase to approximately 80% after personalization is used to select the message thought to be most helpful for that learner, given what the Coursera platform has already encoded about the learner's demographics and learning patterns. Impressively, the messages in this study had significantly higher helpfulness ratings than the average of previously tested in-course pop-ups ($t - test, p < 0.0001$). Even with the randomized deployment of these messages, nine out of ten responses indicated that the self-efficacy and value relevance messages were helpful. With personalization from machine-learning recommendations, the researcher would expect this number to increase further.

A two-way ANOVA test was used to assess any meaningful differences in perceived helpfulness across treatment and gender groupings. The analysis revealed a significant difference between treatment group messages ($p < 0.0001$, see Table 13) resulting from the self-efficacy and value relevance treatments having consistently higher helpfulness ratings than the planning and combined groups. On the other hand, the difference between the two columns was not significant ($p = 0.146$, see Table 13), meaning there is insufficient evidence to suggest any systematic differences between males' and females' perceived helpfulness of the messages across treatment groups. Learners, on average, responded positively to the helpfulness indicator question, with the self-efficacy and value relevance treatments receiving especially strong utility ratings.

Table 13

ANOVA Test Results for Helpfulness Rating by Gender and Intervention Group

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Rows	0.051	22	0.002	14.571	0.000	2.785
Columns	0.000	1	0.000	2.272	0.146	7.945
Error	0.003	22	0.000			
Total	0.054	45				

It is also critical to examine the proportion of learners choosing to answer the helpfulness question. In addition to the high helpfulness ratings, this intervention demonstrated a response rate of 17.7%, meaning that more than one in six learners responded to the question that appeared in the experiment's messages across all treatment groups. This response rate was 17.4% for female learners, demonstrating similarly high engagement. Previous messages on the Coursera platform received an average response rate to the real-time helpfulness question of 5–6% (Hickey & Urban, 2019). The dramatically higher engagement observed for the in-course messages in this experiment represents a positive indicator of learners' response and this study's broader implementation.

In addition, learners who responded to the helpfulness question showed no difference in completion rate, with the female learners who indicated "yes" or "no" to the helpfulness question having an average completion rate across the four active treatment groups of 16.8% as compared with the overall completion rate for females across these four groups of 17.1%. This non-significant difference (t = -1.73, p = 0.083) demonstrates that the females who responded to the helpfulness question did not differ meaningfully from the overall population of female learners and can be used as a useful indicator of these learners' broader sentiments.

Outcome Evaluation

With 324,457 active learners in the experiment, and the majority responding positively to the process indicators, this study established a strong basis for assessing the outcome questions. Building on the evidence that the intervention had reached the intended learner audience and appeared helpful to that audience, the outcome research questions aimed to investigate the longer-term impact on progression and performance. Given the RCT design of this intervention study, the outcome results can be stated as causal findings, not merely correlational claims. The following section examines these outcome results and what can be learned from this online learning experiment.

RQ3: Persistence

On the path toward skill development and career impact, learners first need to progress through the course content, successfully completing assignments and, ultimately, the entire course. The researcher focused on first-week and course completion rates to assess progression across gender and treatment groups. Table 14 summarizes the rates for completing the first week of learning material, and Figure 19 provides the graphical version of these findings.

Table 14

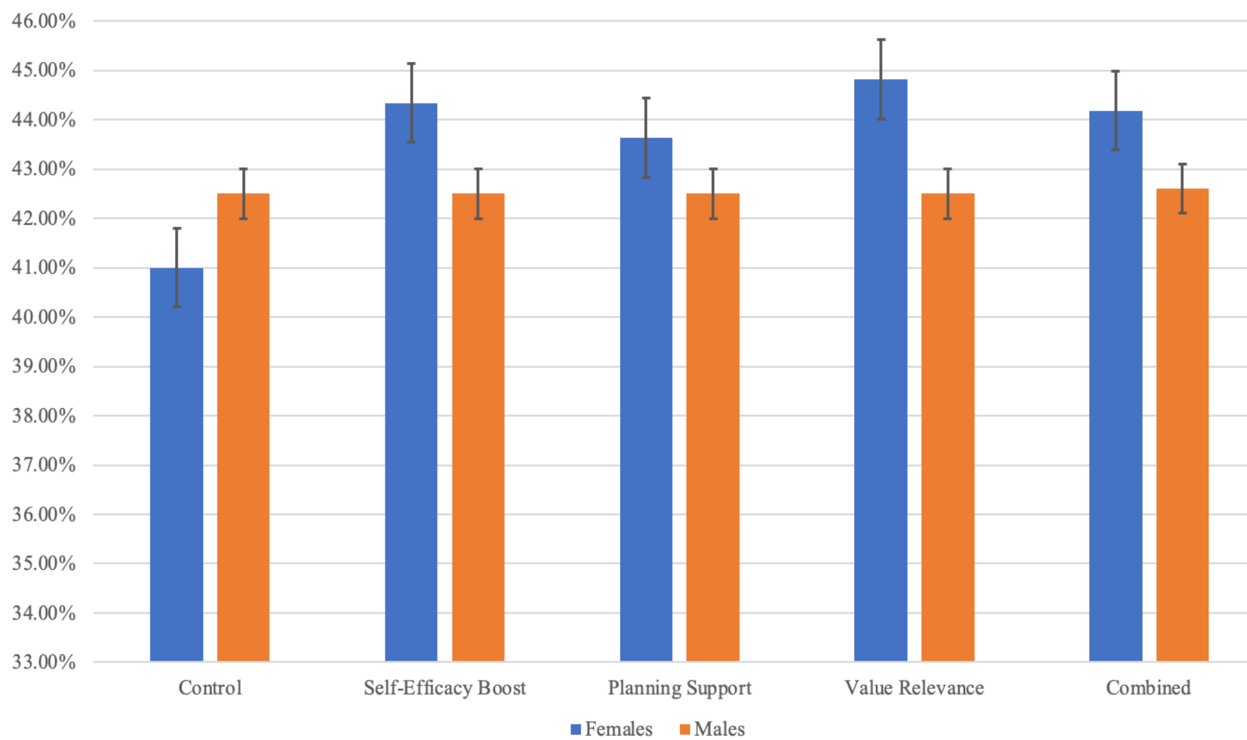
Learners' First-Week Completion Rates by Intervention Group and Gender

	Control	Self-Efficacy Boost	Planning Support	Value Connection	Combined
Female Active Learners *n = 131,804*	41.00%	44.34%	43.63%	44.82%	44.18%
Male Active Learners *n = 192,653*	42.51%	42.52%	42.54%	42.52%	42.53%

Note. Courses, n = 150

Figure 19

Graph of Learners' First-Week Completion Rates by Intervention Group and Gender



The first-week completion rates of males and females in the control group were substantially closer than in the needs assessment study and more similar than expected. This control finding is likely the result of stronger pedagogical design: the STEM courses with the highest enrollment also tend to have the highest ratings and best teaching strategies with clear, useful assessments. This experiment was limited to only 150 of the top STEM MOOCs on the Coursera platform instead of the 2,300 STEM MOOCs analyzed for the needs assessments. The female and male persistence rates were more similar in this narrowed sample, underscoring the importance of high-quality course design in reducing the retention gender gap. However, females in the control group still demonstrated a significantly lower first-week completion rate than their control group male peers ($t - test, p = 0.0085$).

The impact of the different treatments varied noticeably by gender. Notably, all four treatment groups demonstrated significantly higher persistence than the control when examining only female learners ($t - test, p < 0.001$ for each pair). Additionally, the female learners showed significantly higher first-week completion rates than male peers in the self-efficacy-boost, value relevance connection, and combined intervention groups ($t - test, p < 0.001$ for each pair by gender). While an increase from 41% to 44% of learners finishing the first week of the content may not appear to be a large absolute gain, this increase represents an additional 500 total female learners who completed this first week of materials and assessments in the self-efficacy, value relevance, and combined intervention groups than would have if all these females were in the control group.

These significant differences can also be observed visually. The error bars in Figure 19 show the intervals for the true value of each subgroup at the 99% confidence level. The intervals do not overlap between the treatment and control groups. This separation demonstrates the statistically significant differences between the values at the 0.01 alpha level. The two-way ANOVA test revealed a significant difference in how gender affected learners' first-week completion rates in the experiment ($p = 0.0097$, see Table 15). While these initial outcome results suggest potential benefits to female learners, their further progression, performance, and future learning must be evaluated to assess broader impact.

Table 15

ANOVA Test Results for First-Week Completion by Gender and Intervention Group

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Rows	0.001	4	0.000	1.026	0.4812	3.6975
Columns	0.001	1	0.001	8.948	0.0097	8.8616
Error	0.001	4	0.000			
Total	0.004	9				

The treatment effects also varied in learners' course completion rates. The male learners showed impressively similar course completion rates across all groups in the experiment, mirroring the same consistency observed in their first-week completion rates. In contrast, the designed interventions demonstrated divergent effects on female learners' likelihood of finishing the course. Table 16 and Figure 20 display these varying effects by intervention and gender.

Table 16

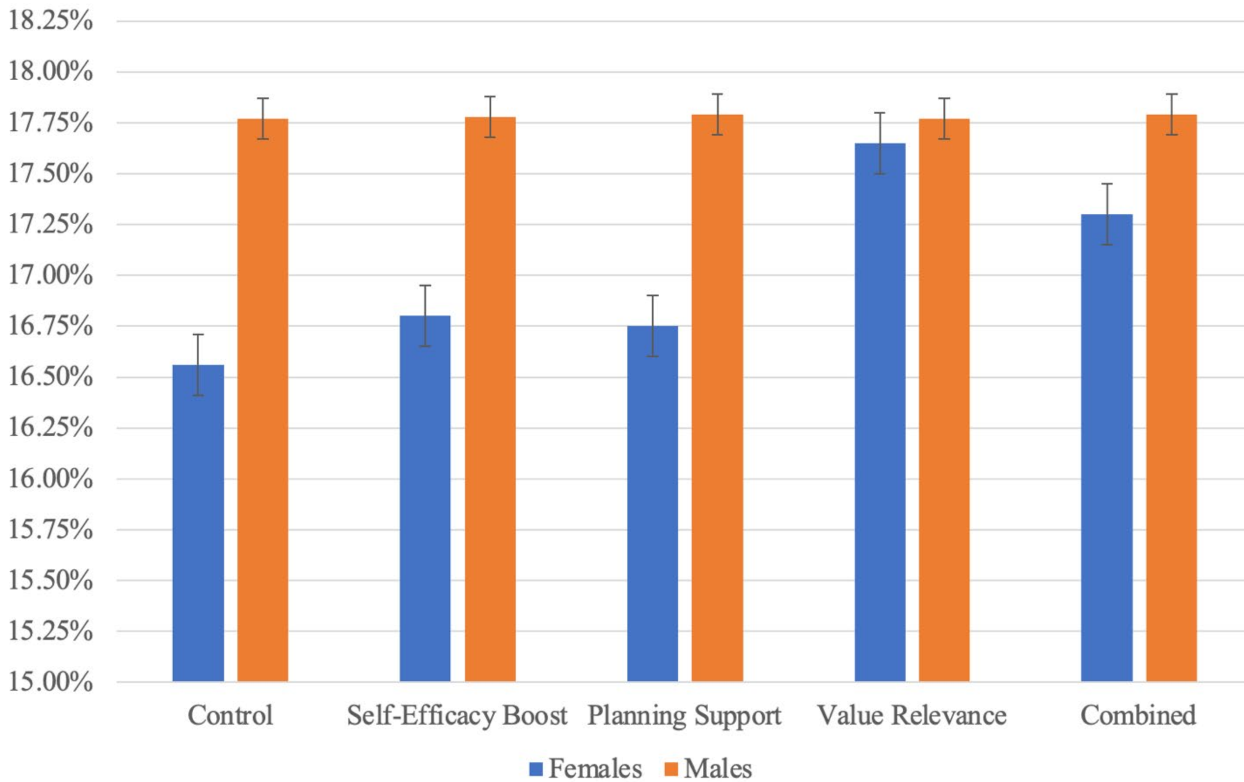
Learners' Course Completion Rate by Intervention Group and Gender

	Control	Self-Efficacy Boost	Planning Support	Value Connection	Combined
Female Active Learners *n = 131,804*	16.56%	16.80%	16.75%	17.65%	17.30%
Male Active Learners *n = 192,653*	17.77%	17.78%	17.79%	17.77%	17.79%

Note. Courses, n = 150

Figure 20

Graph of Learners' Course Completion Rate by Intervention Group and Gender



As seen in the graph, there is a clear interaction between gender and completion rate when examining by treatment group (ANOVA, $p = 0.0054$, see Table 17), with the error bars representing confidence intervals at the 99% level. Although the impact of the self-efficacy boost and planning support waned over the duration of the course, the value relevance emphasis and combined intervention groups showed sustained benefits. Notably, the female learners in the value relevance group demonstrated a completion rate indistinguishable from the male learners in the sample ($t - test, p = 0.076$). Thus, the value relevance intervention successfully closed the gender gap in learners' STEM MOOC completion rates.

Table 17

ANOVA Test Results for Course Completion by Gender and Intervention Group

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Rows	0.000	4	0.000	0.298	0.866	15.977
Columns	0.001	1	0.001	29.992	0.005	21.198
Error	0.000	4	0.000			
Total	0.001	9				

Furthermore, both the value relevance emphasis and combined intervention groups increased female learners' course completion rates significantly above the female learners' control group completion rate ($t - test, p < 0.001$ for each pair). While the increase in female course completion rate from the control to value relevance groups may not appear meaningful, this difference resulted in 7% more females completing the course. Moving all the females in this experimental sample from the control group to the value relevance condition would result in approximately 1,400 additional female learners completing their STEM courses.

Given the extant literature, the researcher explored these progression metrics by learners' demographic groups, including age, home country, and employment status. The researcher's hypothesis, in part based on previous studies, was that the self-efficacy-boost treatment would be especially beneficial for female learners in the youngest age tier. For females aged 18 to 24 years, the control treatment had a first-week completion rate of 34.50% compared to 39.86% for the self-efficacy boost ($t - test, p < 0.001, n = 315$). This increase, resulting from the self-efficacy intervention, brought first-week completion rates for the traditionally under-progressing group of young adult females up to nearly the average seen across all active female learners.

This initial impact was retained and became more exaggerated when examining course completion rates. Young active females in the self-efficacy-boost group demonstrated a course completion rate of 13.21% compared to only 8.97% in the control group ($t - test, p < 0.0001, n = 315$). While still lower than the overall completion rate for active female learners, this finding demonstrates the importance of investigating the impact by varying demographic groups. Although the self-efficacy-boost treatment did not meaningfully alter the overall female learner MOOC completion rate in this experiment, this intervention treatment did demonstrate a significant, positive impact on younger females. In fact, the self-efficacy-boost treatment increased the number of 18- to 24-year-old female learners who completed their STEM course by 50%.

Female learners join courses on Coursera from vastly different home, family, and work circumstances, which often affect their ability to persist in the material. Home country is a useful macro-indicator of their broader situation (Guiso et al., 2008; Kizilcec, Saltarelli, et al., 2017). With this intervention sample, the researcher mapped countries according to their 2018 United Nations Gender Inequality Index (UN GII) score, the same country score used during the earlier needs assessment study. After narrowing to countries with at least 35 unique active female learners enrolled in the RCT, this researcher conducted linear regressions to analyze the impact of the value relevance treatment group on completion rate by county's UN GII score. This model found a negative correlation between the course completion rate in the control group and gender inequality level, where a lower number equates to greater gender equality ($\beta = -0.04, R^2 = 0.02$) across the 82 countries. For the value relevance treatment group, the plotting of active female learners' course completion rate by country's UN GII resulted in a nearly flat trend line ($\beta = -0.007, R^2 = 0.0006; n = 82\ countries$). While not a significant correlation ($p =$

0.202 for control and $p = 0.831$ for value relevance), this flattening of the correlation to almost zero suggests the value relevance intervention potentially helped counterbalance the effect of a nation's gender inequality on female learners' likelihood to complete the MOOC. Table 18 summarizes these test statistics, and Figures 21 and 22 show the corresponding graphs.

Table 18

National Gender Inequality's Effect on Female Course Completion by Intervention Group

CONTROL

	df	*SS*	*MS*	*F*
Regression	1	0.004	0.004	1.65
Residual	80	0.186	0.002	
Total	81	0.190		

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	0.140	0.010	14.135	0.000
X Variable 1	-0.043	0.034	-1.285	0.202

VALUE RELEVANCE

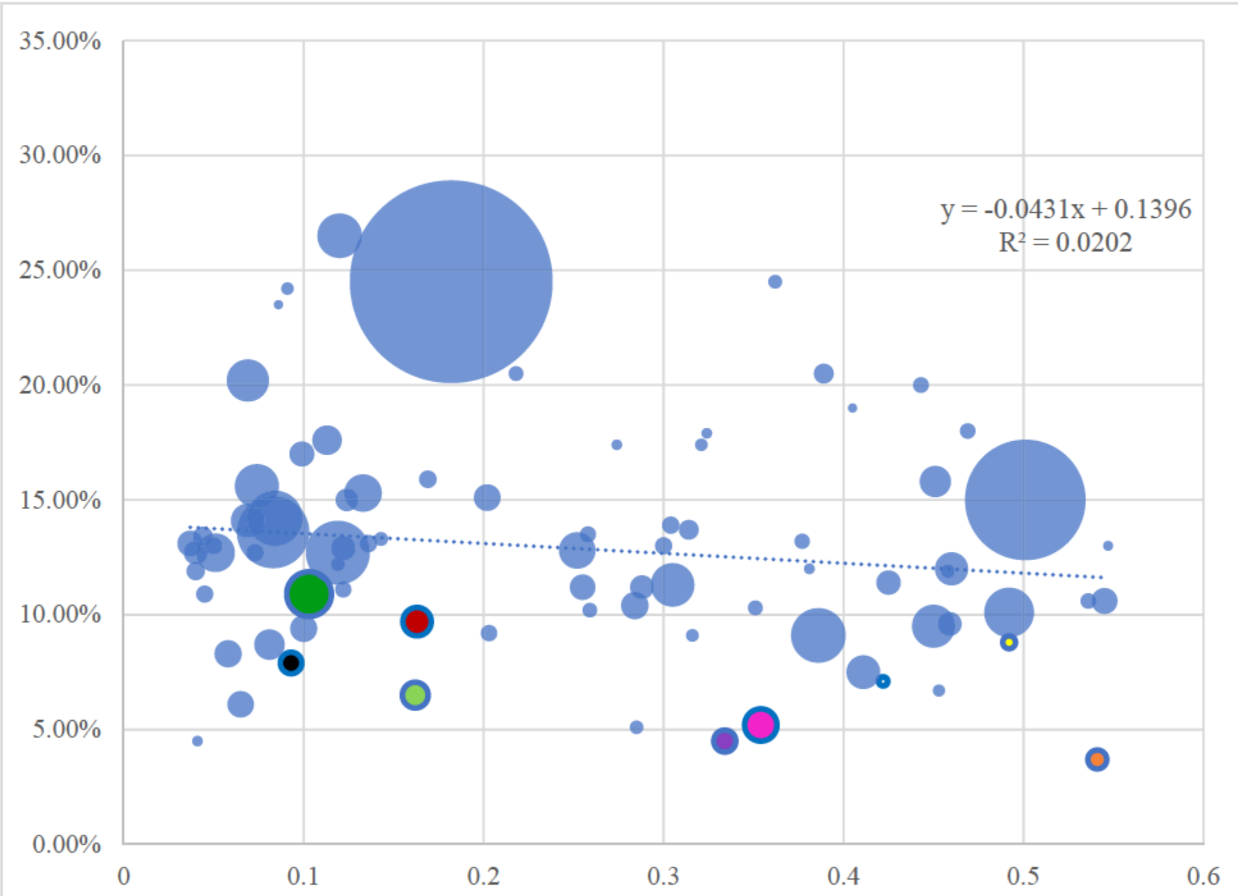
	df	*SS*	*MS*	*F*
Regression	1	0.000	0.000	0.046
Residual	80	0.209	0.003	
Total	81	0.209		

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	0.134	0.010	12.833	0.000
X Variable 1	-0.008	0.035	-0.2141	0.831

Note. Countries, n = 82; courses, n = 150

Figure 21

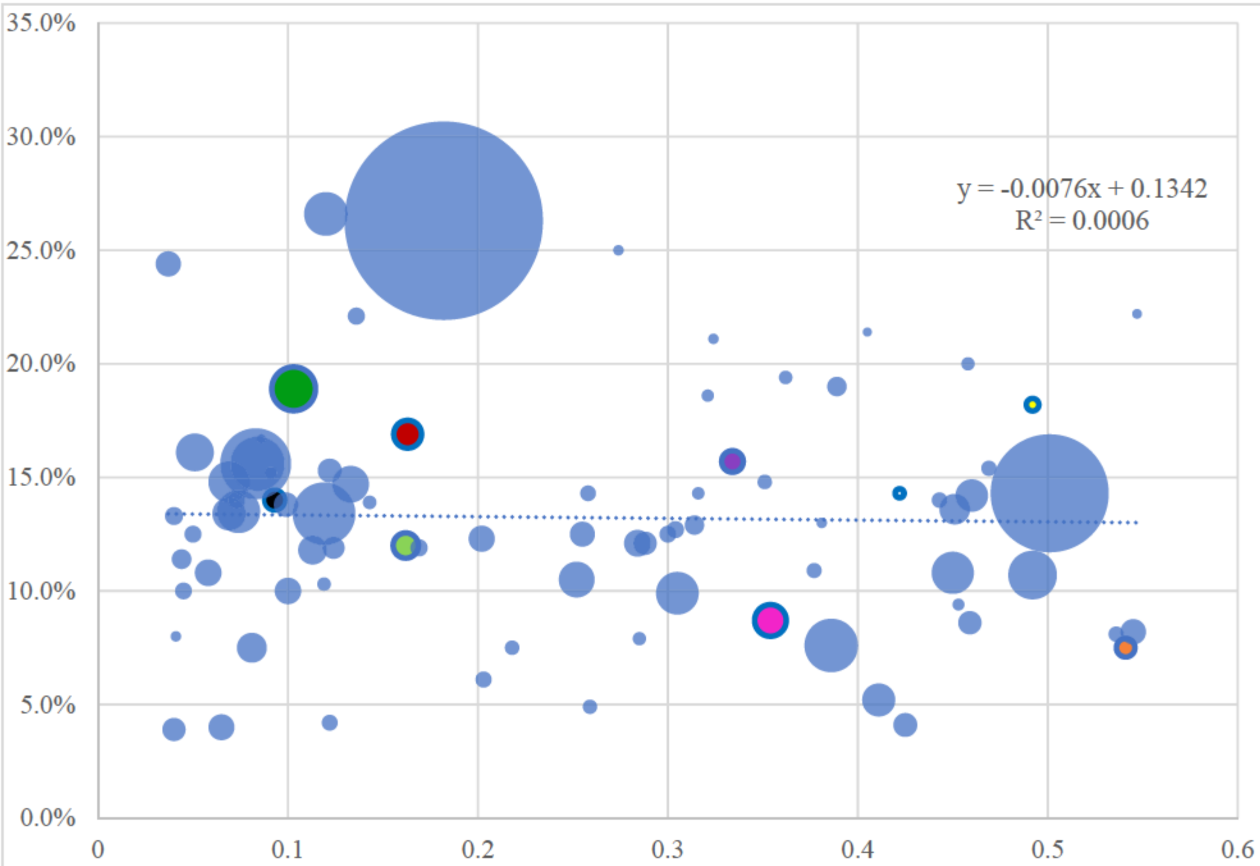
Female Learners' Control Course Completion Rate by Country's Gender Inequality



Note. Each bubble represents the active female learners from a single country. The vertical axis is the percentage of these learners who completed the STEM MOOC they enrolled in as part of the control group of this intervention experiment, and the horizontal axis is the country's UN GII score, with higher scores indicating greater inequality. The bubble's diameter represents the relative number of active female learners from that country enrolled in the RCT. The pink bubble represents the active female learners in Argentina. Active female learners from Australia are shown in dark green, China in red, Ghana in orange, Guatemala in yellow, Ireland in black, Malaysia in purple, Saudi Arabia in light green, and Slovenia in white.

Figure 22

Female Learners' Value Relevance Course Completion Rate by Country's Gender Inequality



Note. Each bubble represents the active female learners from a single country. The vertical axis is the percentage of these learners who completed the STEM MOOC they enrolled in as part of the value relevance group of this intervention experiment, and the horizontal axis is the country's UN GII score, with higher scores indicating greater inequality. The bubble's diameter represents the relative number of active female learners from that country enrolled in the RCT. The pink bubble represents the active female learners in Argentina. Active female learners from Australia are shown in dark green, China in red, Ghana in orange, Guatemala in yellow, Ireland in black, Malaysia in purple, Saudi Arabia in light green, and Slovenia in white.

With the unfortunate ubiquity of gender inequality across the globe, it is useful to explore the effect of the value relevance treatment on active female learners' completion rate across different countries. While the UN GII scores summarize high-level factors of gender inequality, even the wealthiest countries still have a widely documented lack of female equity in the workplace and an imbalance of empowerment opportunities (OECD, 2019). As the UN team emphasizes, not a single country gained a perfect equality score, even on the crude metrics they examined for their UN GII rankings. These scores are derived from female representation in each nation's education, labor market, and government alongside financial access and federal rights granted to women. The overall trend of the UN GII scores emphasizes the abundant gender inequality across all corners of the globe. Thus, this researcher investigated the impact of the value relevance treatment across Africa, Europe, Asia Pacific, and the Americas to gain insight into how this treatment may help counteract gender inequality as it appears in different regions of the world.

While the overall trend may not appear meaningful, the active female learner completion rate in several countries doubled, or nearly did so, in the value relevance group relative to the control group. For example, the course completion rate for learners from Ghana, one of the largest female enrollment countries in sub-Saharan Africa, was 3.7% in the control group and 7.5% in the value relevance treatment ($t - test, p < 0.001, n = 149$). Moving to the Middle East, the female completion rate in Saudi Arabia jumped from 6.5% in the control to 12.5% in the value relevance treatment ($t - test, p < 0.001, n = 280$). Looking north to Europe, the completion rate for active female learners in Ireland went from 7.9% in the control to 14.0% in the value relevance group ($t - test, p < 0.001, n = 194$). Slovenia displayed the same trend, with a control completion rate for active female learners of 7.1% and the value relevance treatment raising it to 14.3% ($t - test, p < 0.001, n = 35$). Furthermore, Guatemala, the country in Central and South America with the highest UN GII score in this sample, demonstrated a completion rate of 8.8% for active female learners in the control group and 18.2% in the value relevance group ($t - test, p < 0.001, n = 67$). Female learners in Argentina displayed a similar trend, jumping from 12% completion rate in the control to 22.2% in the control ($t - test, p < 0.001, n = 442$).

Countries in the Asia Pacific region replicated these effects of the value relevance treatment. For example, active female learners in China completed at a rate of 9.7% in the control compared to 16.9% in the value relevance treatment ($t - test, p < 0.001, n = 340$). In Malaysia, female learners demonstrated the largest effects from the value relevance treatment across all nations in the sample, raising the control completion rate of 4.5% to 15.7%, demonstrating a nearly 250% increase in the number of females completing their STEM MOOC ($t - test, p < 0.001, n = 204$). In Australia as well, females saw a more than 70% boost in completion rate from the value relevance intervention, moving from 10.9% to 18.9% ($t - test, p < 0.001, n = 850$). Across this list of diverse countries, the benefits from the value relevance treatment represent a 67% increase or greater in the number of females completing the course compared with the control group. Table 19 displays the effects of the value relevance treatment on active female completion rate across these selected countries. Although the samples from individual countries range in size, these gains provide evidence for how emphasizing individuals' values and greater goals can boost female course completion rates across a wide variety of countries likely by counteracting the pervasive presence of gender inequality.

Table 19

Value Relevance Effect on Active Female Course Completion for Select Countries

Theme	Control Completion Rate	Value Relevance Completion Rate	Increase
Argentina *n = 442*	5.2%	8.7%	67%
Australia *n = 850*	10.9%	18.9%	73%
China *n = 340*	9.7%	16.9%	74%
Ghana *n = 149*	3.70%	7.5%	103%
Guatemala *n = 69*	8.80%	18.2%	107%
Ireland *n = 194*	7.9%	14.0%	77%
Malaysia *n = 204*	4.5%	15.7%	249%
Saudi Arabia *n = 280*	6.5%	12.5%	85%
Slovenia *n = 35*	7.1%	14.3%	101%

Note. Courses, n = 150. Countries are shown in alphabetical order. The sample noted for each country is the total active female enrollments in the intervention sample from that nation across all RCT groups. Control and value relevance completion rate were calculated by counting the number of active females who completed the STEM MOOC in the experiment from that treatment group. The increase column represents the percentage gain seen for the value relevance active female course completion rate compared to the rate for the control group of active females in the same country.

It is also beneficial to examine the impact of these interventions on female learners working full time. Control group females who reported full-time employment during the experiment implementation demonstrated a higher baseline completion rate (17.82%) than all control group female learners combined. This higher control completion rate for females working full time is likely the result of two factors: career alignment with the coursework, which provides external motivation, and companies paying for their employees to take courses through the Coursera for Business offering. The only intervention that significantly impacted their average course progress was value relevance, which resulted in 18.62% of active female learners with full-time jobs completing the course ($t - test, p < 0.001, n = 12{,}282$). This finding mirrors the same overall positive effect seen for females across the experiment, even though this subset of females started at a higher baseline completion rate.

RQ4: Skill Development

Coursera courses are designed for learners to progress and gain meaningful benefits from the material. Examining course completers' performance on in-course assessments while controlling for previous learning enables a fair and systematic evaluation of skill development. For this sample, the vast majority of learners (greater than 80%) had not yet attempted sufficient previous assessments on the Coursera platform in the domain of the MOOC in which they enrolled for the experiment, as seen in the anonymized stored data linked to each learner. These previous assessment attempts are needed to establish a baseline "skill score" to assess the impact of the new MOOC enrollment on their overall skill development (Hickey et al., 2020; Reddick, 2019).

Without an accurate baseline measure, this researcher decided to focus purely on learners' demonstrated learning gains within each MOOC to yield the most reliable results. Plus, with a sample this large, the average of learners' previous skills and knowledge should be relatively uniformly distributed across the five intervention groups. Given that in-course performance drives the increase of learners' individual "skill score" on the Coursera platform (Reddick, 2019), the researcher analyzed the course completers' final course grades across gender and treatment groups. Table 20 and Figure 23 summarize the average performance achieved by experimental group and gender.

Table 20

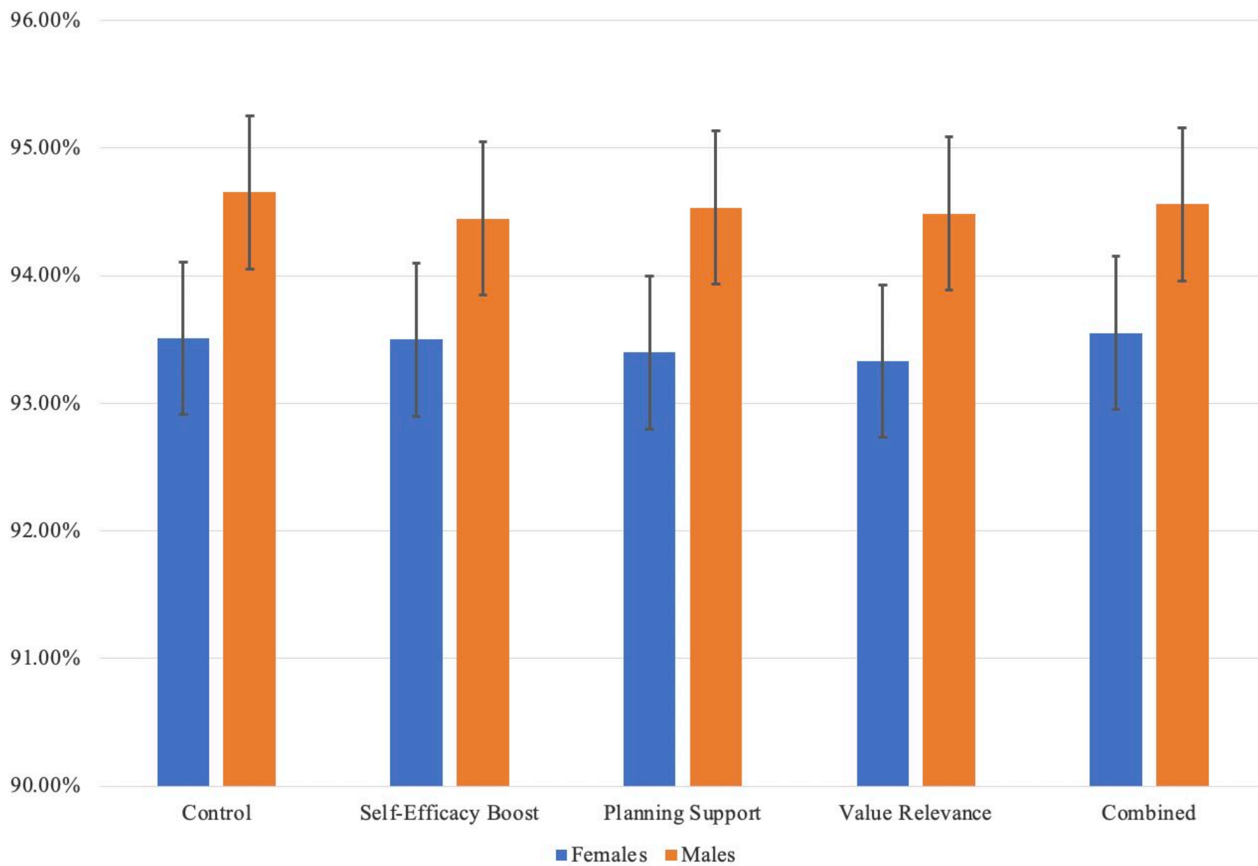
Course Completers' Average Grade Achieved by Intervention Group and Gender

	Control	Self-Efficacy Boost	Planning Support	Value Connection	Combined
Female Course Completers *n = 22,540*	93.51%	93.55%	93.40%	93.33%	93.55%
Male Course Completers *n = 30,248*	94.65%	94.45%	94.54%	94.49%	94.56%

Note. Courses, n = 150

Figure 23

Course Completers' Average Grade Achieved by Intervention Group and Gender



As shown in Figure 23, the average STEM MOOC grade achieved by course completers was consistent across treatment groups. While the males achieved slightly higher (percentage) grades in all groups, there is insufficient evidence to suggest a significant difference between the females' average final course grade of 93.5% and the males' average of 94.5% at the 0.01 alpha level ($ANOVA, p = 0.037$, see Table 21). More crucially, this result also likely lacks practical significance, as meaningful differences in learners' later career outcomes are rarely traceable to a one-percentage-point difference in final course grade. Additionally, this metric appears to have a strong ceiling effect: with many students earning final course grades at or close to the maximum

100% possible, it becomes difficult to distinguish differences across groups by gender or treatment (Salkind, 2010). While not significant when examining average grades among the experimental groups, this performance indicator is revisited below in RQ6, which considers individuals' learning gains beyond their MOOC enrollment in this experiment.

Table 21

ANOVA Test Results for Final Course Grade by Gender and Intervention Group

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Rows	0.000	4	0.000	0.833	0.568	15.977
Columns	0.000	1	0.000	9.525	0.037	21.198
Error	0.000	4	0.000			
Total	0.000	9				

RQ5: Dropoff Survey

For the learners who did not complete their STEM MOOC in the experiment, it is useful to examine their self-reported reasons for stopping the course before finishing. Given the short intervention timeline and the delay in sending the Coursera automatic Inactivity Survey (three weeks after the most recent activity in the course), only 41 responses were collected from female learners across the four treatment groups and three responses from the control group. Thus, the researcher compared the treatment responses in aggregate to the needs assessment findings discussed in Chapter 2, with the latter acting as the pseudo-control. While the inclusion of a larger sample of MOOCs in the needs assessment may result in different reasons for female learners dropping out, it is reasonable to expect many similarities since this experiment represents a smaller but largely overlapping group of courses from the original needs assessment sample. Table 22 summarizes by theme the new responses collected by the Inactivity Survey from female learners in the four treatment groups of the experiment.

Table 22

Females' Reasons for Stopping Before Course Completion in Treatment Groups

Theme	Prevalence
Content Itself	39%
Technical Difficulties	22%
"Not interested in assignments"	10%
"Not confident"	7%
"No time"	5%
Prerequisite Knowledge	5%
"Don't want to pay"	2%
Other	10%

Note. Learners, n = 41

For females in one of the four treatment groups, most respondents cited a misalignment with the content or frustration with the technical assessments. Specifically, these female learners expressed disappointment regarding the topics and materials covered. Many learners expressed similar sentiments across the Content Itself theme, including learners explaining that the course was "not the one I need," "not what I was looking for," "not what I was interested in," and "not what I expected." While these narratives offer critical feedback on the enrollment flow and how more clarity may be needed when selecting courses, these insights are not as useful when considering how best to support learners who want to gain the skills in their current course. The Technical Difficulties theme consisted of learners lamenting their inability to open, access, or install needed software to complete one of the assignments.

In contrast, the themes related to the motivational areas of the intervention's treatment groups showed promising decreases. While only a small sample, it is encouraging that lower proportions of the female non-completers referenced their time, confidence, or prerequisite knowledge. Under the "no time" theme, learners emphasized not being able to "afford to spend my time watching… right now" and how they "question if it [is] worth my time." These insights capture how female learners often must calculate where and when to spend their limited free time while juggling work, home, and family responsibilities (Perez, 2019).

The confidence and prerequisite responses highlighted two distinctive sets of challenges. In the "not confident" bucket, females emphasized how they "need to understand more before [I] go for the assignments" and how it is "too hard for me to understand." These quotations suggest certain learners were dropping out even before attempting the assessments in a course. However, the Prerequisite Knowledge theme had females pinpointing precisely what skills or tools they needed to learn before they could be successful in the content. Learners in this category noted, "PyTorch," "TensorFlow," and "Matlab" as specific frameworks and software they needed greater knowledge of before returning to the course. While the Prerequisite Knowledge learners had actionable next steps—attempting to learn the basics of PyTorch functions or Matlab programming, for example—the learners in the "not confident" group expressed a more abstract feeling of not being able to understand or continue.

Many learners in the Prerequisite Knowledge theme from the needs assessment study expressed a lack of both skills and confidence. For example, in the original sample, females noted, "I don't understand how to complete the assignment," or, "I do not have the resources needed." These learners also noted that they needed more specific knowledge, such as "computer skills" and "language abilities." Identifying their specific skill gaps pushes learners toward a growth mindset of improvement, while focusing on current failures often immobilizes them, believing improvement is impossible (Hickey et al., 2018). While these are only preliminary results because of the smaller sample size, these self-reported reasons show an optimistic decrease in time and confidence limitations causing female dropouts.

RQ6: Continued Learning

The researcher was curious to assess whether these intervention treatments would spark continued learning beyond retention in the current MOOC. To examine this phenomenon, the researcher investigated new active enrollments in other STEM MOOCs on the Coursera platform after the learner became active in their experiment course. Unlike traditional educational settings in which students can only start new courses at specific times of the year, learners on Coursera can enroll in new MOOCs and start making progress at any time. To standardize this metric of new enrollments, the researcher divided new active STEM enrollments by the number of active learners per treatment group to obtain the average number of new STEM enrollments per active learner. These results are summarized in Table 23 and Figure 24.

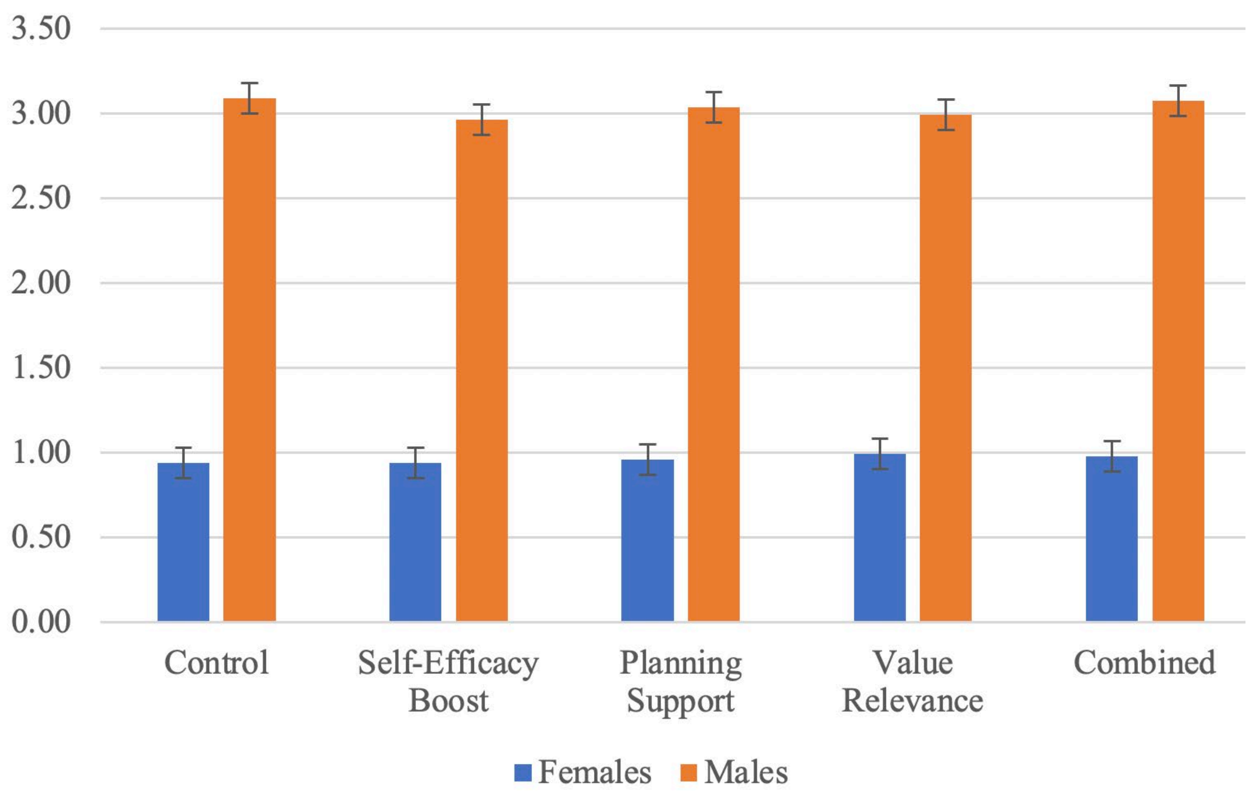
Table 23

Learners' Average New STEM MOOC Enrollments Per Active Learner

	Control	Self-Efficacy Boost	Planning Support	Value Connection	Combined
Female Active Learners *n = 131,804*	0.937	0.936	0.956	0.991	0.978
Male Active Learners *n = 192,653*	3.09	2.96	3.04	2.99	3.08

Figure 24

New STEM MOOC Enrollments per Active Learner by Intervention Group and Gender



The intervention may influence females' continued learning by pursuing other STEM MOOCs on the Coursera platform, but there is insufficient evidence to suggest a significant causal relationship. The value relevance and combined treatment groups both showed a slight increase in active STEM enrollments for females who were active in the experiment course, with female learners, on average, increasing from nine to 10 new enrollments for every 10 learners. However, males still enrolled in other MOOCs beyond the experiment at much higher rates than their female peers, at an average of 30 new STEM enrollments for every 10 learners ($p < 0.0001$, see Table 24); no significant differences were seen across treatment groups for the female or male learners ($p = 0.52$, see Table 24). While the males' averaged three new enrollments per active learner, female learners averaged under one, displaying the same gender enrollment gap observed in the needs assessment.

Table 24

ANOVA Test Results for New STEM Enrollments by Gender and Intervention Group

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Rows	66.450	4	16.613	0.958	0.516	6.388
Columns	107205.316	1	107205.316	6183.079	0.000	7.709
Error	69.354	4	17.339			
Total	107341.120	9				

Notably, the combination treatment significantly impacted female learners' grades in these future STEM enrollments. While more than 22,540 females completed the STEM MOOC in the experiment, 15,394 of the active females in the RCT went on to complete a future STEM MOOC outside the confines of this experiment. The combination treatment proved beneficial for these female course completers beyond the experiment. Combining self-efficacy, planning, and value relevance prompts led to female learners earning significantly higher end-of-course grades in the STEM courses they completed during the experiment than females in the control group ($t - test, p < 0.001, n = 6{,}246$). While the grades of females and males remained indistinguishable across all groups, the females from the combination treatment significantly outperformed all four other groups of females in their final grades in future STEM enrollments they completed. Appendix I presents a graph of the average course completers' final grades in future STEM enrollments by gender and intervention group.

Discussion

These research-based, light-touch interventions significantly increased female persistence and completion in the most popular STEM MOOCs on Coursera. All four treatments raised females' first-week completion rate above that of females in the control group. In addition, the self-efficacy-boost, value relevance emphasis, and combination treatment groups demonstrated significantly higher first-week completion rates than the males in the first week. This result was more conclusive than expected and showed how potent even brief interventions could be in asynchronous online courses. Minor encouragements designed to increase female learners' intrinsic motivation, drawing on their competence, autonomy, and relatedness, yielded promising results.

While several of the treatments had only minor effects on overall course completions, it is important to remember that these prompts were only offered during the first three weeks of the courses. Most of the STEM MOOCs included in the experiment lasted four to eight weeks, suggesting the potential benefits of extending these self-efficacy, planning, and value connection prompts into the later weeks of the course. Given all four treatments had a meaningful, positive impact on first-week completion, including these prompts in later weeks may result in continued elevated retention rates for the females.

The impact of these prompts on course completion rate was also encouraging. While the 4% gender gap witnessed during the needs assessment would be challenging to close with only a few text-based prompts, the baseline difference in the completion rate for these top 150 STEM MOOCs on the platform was 1.2%. The most popular courses on the Coursera platform tend to have the most thoughtful designs and highest quality pedagogical approaches, which previous authors have found can provide the most assistance to female learners (Hickey et al., 2018; Stolk, Jacobs, et al., 2018; Vennix et al., 2018). With this reduced baseline gender gap and the impact of these novel prompts, the value relevance group resulted in male and female course completion rates that were indistinguishable. This finding demonstrates that females' completion rate significantly increased when their values were highlighted and linked to their learning journey, successfully closing the gender gap in STEM MOOC completion.

While other authors have observed similar benefits derived from value relevance prompting for learners from developing countries, the effects have been negative for those from more developed nations (Kizilcec et al., 2017). However, in this intervention, the researcher integrated general value relevance prompting with course content alignment to those goals. This integrated design significantly impacted the completion rate in a diverse sample of 131,804 active female learners. The correlation between the effect of this value relevance intervention and each country's UN GII ranking was nearly zero, indicating this treatment had a particularly strong effect on female learners in nations with greater inequality, such as those in Ghana and Guatemala. Overall, this research provided a beneficial experience across a diverse group of females in the value relevance treatment group.

For younger females, the self-efficacy intervention was especially helpful. Younger female students have dramatically lower completion rates than their older and male peers (Allione & Stein, 2016; Rabin et al., 2020) and also show the largest increases when provided a self-efficacy-enhancing intervention (Chyung, 2007). Thus, the researcher investigated whether the self-efficacy boost led to any positive effects for younger females. Despite not resulting in significant increases in overall female course completion, this intervention did show large, significant benefits for the females aged 18 to 24. Specific subgroups of females are most likely to drop out before course completion, such as younger learners and those joining from developing countries, so it is encouraging to observe the self-efficacy boost increasing the number of young female course completers by 50%.

Finally, females with full-time jobs demonstrated a higher baseline completion rate, with many of these learners gaining skills that may benefit their current job and company. Notably, the planning support treatment completion rate did not increase their persistence, but the value relevance emphasis significantly increased an already elevated completion rate (from the control group) to an impressive 18.6%. While greater schedule flexibility is likely still needed for these women juggling work, home, and courses (Allione & Stein, 2016; Watson et al., 2018), it is promising to see how career alignment and value connection may work synergistically to raise the completion rate significantly for full-time working females.

Overall, the self-efficacy-boost and value relevance treatments displayed the highest helpfulness ratings from females, as seen during the process evaluation. In particular, the “This learning is for you” message, as part of the value relevance treatment condition, was the most highly rated message across the entire experiment for all learners and females, with 94.2% of the latter group indicating it as helpful. Ultimately, this message was part of the treatment that had the largest benefits for female course completion rate. This connection between process and outcome metrics demonstrates how learners perceived utility may at least partially predict their later progression patterns. By exploring early process indicators, researchers can start to preview downstream effects of an intervention (Baranowski & Stables, 2000).

Average final course grades remained stable across all five intervention groups and were indistinguishable across genders. With only a one-percentage-point difference between males and females, it was impossible to discern any difference in skill development gained from their in-course learning. Given that this analysis section was limited to course completers, all the learners in each MOOC had successfully finished the same exams and projects. With this narrowed sample and the ability to examine only one course per unique learner in the experiment, there was insufficient differentiation to result in large skill development variations. Even so, the researcher was encouraged to see both females and males, on average, scoring within the A letter-grade range when completing these STEM MOOCs. With only a 1% gender gap in average final grade achieved, this disparity represents neither a statistical nor a practical difference. Furthermore, despite no skill development differences at the individual course completer level by intervention group, more females were completing the course, resulting in greater skill development at the population level across active female learners in the successful treatments.

The written reasons reported by females in the experiment who dropped out of the course before completion differed from those in the needs assessment. From the treatment groups, a larger proportion of females who did not finish the STEM MOOC referred to a mismatch with the material, a technical difficulty, or no interest in completing the assessments. These female learners in the experiment rarely expressed challenges with their time or confidence as barriers. While only a small sample ($n = 41$) provided reasons for why they dropped out, the researcher was encouraged by the rare mentions of a busy schedule and lacking self-efficacy across the four treatment groups (i.e., all except the control). Despite more clearly needing to be done, especially to manage females' frequent time constraints (Perez, 2019), these females' reported reasons for their inactivity before completion demonstrates a shift from the needs assessment. Instead of focusing on their lack of time and perceived lack of ability, these females identified the content and specific prerequisite skills that aligned with their goals.

Despite small increases in female learners' future enrollments, male learners enrolled in significantly more MOOCs beyond the study across all intervention groups. The value relevance emphasis resulted in the largest increase in active STEM enrollments after female learners' first experience in the experiment, although there is insufficient evidence to suggest a meaningful increase or any causal impact. Since this experiment focused on retention and persistence within their current MOOC and not future enrollments, any directionally positive effects in female learners' new STEM enrollments underscore the potential of lasting benefits for even light-touch interventions in online learning environments.

Returning to final course grades earned, female learners in the combination treatment group displayed significantly higher average final course grades in their future STEM MOOC enrollments. While not a primary research outcome metric of this experiment, it is valuable to consider the potential downstream impacts of these intervention treatments. The combination of boosting female learners' self-efficacy, supporting their planning, and connecting the course to their values may lead learners to enroll in more appropriate courses or increase their motivation to achieve in their later course enrollments. If learners see the ease and success of learning in a MOOC, they often quickly enroll in additional MOOCs, building on that momentum (Hickey & Urban, 2019).

Strengths and Limitations

Given the large sample size of more than 300,000 active learners and the overall RCT design, this study lacked the common threats to validity seen in education research. When executed successfully, RCTs mitigate most threats to internal validity by removing the human selection process through a double-blind, randomized process and isolating treatment effects as the only experience changes across groups (Hardiman et al., 2019; Shadish et al., 2002). The double-blind RCT design also avoided the potential issues of participant reactivity from knowledge of group assignment, experimenter's expectations, and any compensatory treatment provided to the control group (Shadish et al., 2002). Given the repeated statistical tests planned, the researcher corrected the alpha to a 0.01 level to increase statistical validity. Finally, the causal relationships, enabled by the RCT design and large sample for stronger statistical power, boosted the external validity of this study.

However, this experiment still had limitations. The sample size for the qualitative analysis was small ($n = 41$), especially compared to the overall hundreds of thousands of learners enrolled in the study. This smaller qualitative sample challenges the representativeness of these data for the broader group of female learners and presents issues for triangulation across the research questions (Creswell & Miller, 2000). Operationalizing the retention and skill development metrics in various ways to include both qualitative and quantitative indicators would have strengthened the construct validity of this study (Shadish et al., 2002). Furthermore, any significant findings are only applicable to MOOCs. While there may be lessons that could transfer to smaller online courses or in-person, lecture-based classes, the results of this experiment are from large-scale, online, asynchronous learning experiences. Maintaining a setting consistent with the context of this experiment is the best method for ensuring external validity (Shadish et al., 2002). Knowing these limitations can help future practitioners and researchers apply the learnings from this study in their own contexts.

Implications for Practice

Even brief, text-based interventions can significantly increase females' retention in MOOCs, which has enormous implications for how instructors orient their lessons and online platforms supplement the coursework. Focusing on the levers of intrinsic motivation, especially connecting individuals' values and long-term goals to the learning experience, raises persistence and can erase the gender gap in course completion (Lee et al., 2020; Peters et al., 2017). Instructors or other content providers could apply these insights into their learning activities with more salient value queues, whether in videos, written material, or at the start of assessments. More explicitly connecting the learning to the students' values may likely benefit female learners, especially those facing challenges from gender inequality.

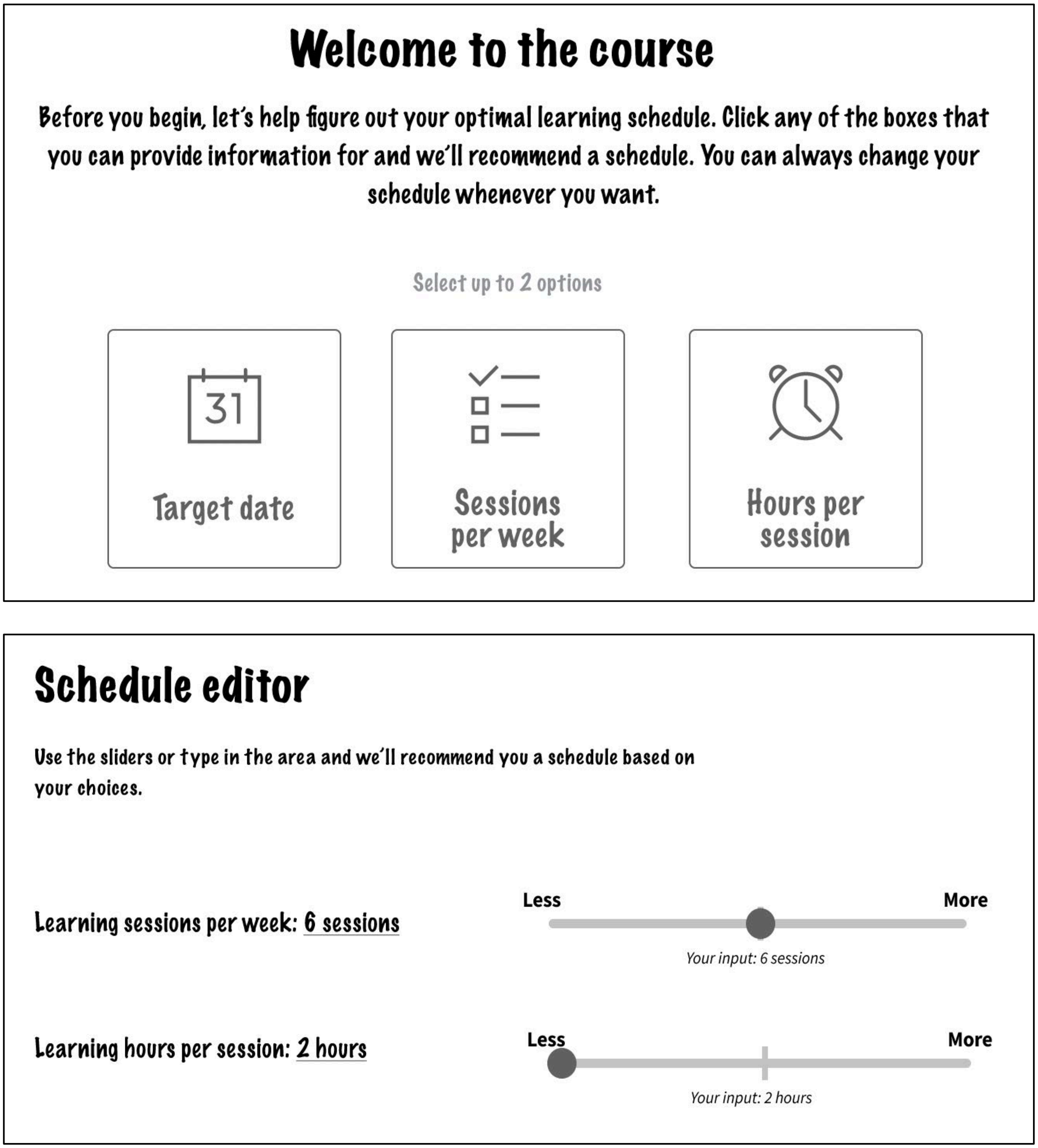
The combination treatment resulting in higher grades for females in future course enrollments also has important implications. This blend of self-efficacy boost, planning support, and value emphasis led females, on average, to stronger future STEM performance. This longer-term outcome suggests that attempting to elevate intrinsic motivation can have lasting benefits for female learners, which is consistent with earlier studies (Loizzo et al., 2017; Stolk, Zastavker, et al., 2018). Highlighting their previous successes and goals can also help female learners identify the appropriate courses for their desires and abilities, which may contribute to higher performance (Grella & Meinel, 2016). Instructors and teaching staff should explore how to integrate these moments of competence, autonomy, and relatedness into their future learning experiences (Stolk, Zastavker, et al., 2018).

Even the treatment groups that did not result in raising course completion rates demonstrate valuable takeaways. For example, the planning support prompt stemmed from previous literature on the difficulty of females' busy schedules and time demands. The use of minimal prompts in this study resulted in female retention gains during the first week of courses, suggesting further schedule and time support may be useful. Specifically, Chapter 3 presents the potential benefits of providing personalized schedules (Bonk et al., 2018; Gütl et al., 2014). Given the substantial product and engineering resources needed, this type of change was not feasible for this intervention study. However, the researcher introduced the idea of this larger update to product leaders at Coursera, who have since designed and pilot-tested personalized deadlines.

An elongated enrollment flow was designed to incorporate these new personalized schedules for MOOCs on the Coursera platform. Upon enrolling, learners are asked to select up to two desired parameters from the three options of the target completion date, learning sessions per week, and learning time per session. Then, learners can use on-screen sliders to select times that align with their personal schedules. Figure 25 shows a prototype of this enrollment flow. After learners input their desired schedule, the platform will provide a customized schedule aligning with their time demands. While still in its early phases, an initial personalized schedule pilot was recently completed and showed an increased active learner retention rate (Coursera, 2022a). Aligning new product features with intrinsic motivation research makes the courses more achievable and helps females fit learning into their busy lives (Stolk, Jacobs, et al., 2018; Watson et al., 2018).

Figure 25

Prototype of Personalized Schedule Enrollment Flow for Learners on Coursera



Furthermore, the findings from this intervention study underscore the need for personalization. Females aged 18 to 24 showed large benefits from the self-efficacy-boosting intervention, despite the effects of this treatment declining to negligible for the overall sample of female learners. Females with full-time jobs benefited from the value relevance treatment, even though the baseline for this group was a significantly higher completion rate than the entire female sample. Using machine-learning algorithms to match learners with the prompts that will work best for them will increase the overall utility of these treatments. Instead of randomly assigning different learners to the various treatment groups, an algorithm could identify demographic and learning patterns regarding which individuals respond best to which prompts (Hickey & Urban, 2019; Yu et al., 2017).

Even if machine-learning matching is not possible in their contexts, other instructors or platforms can also personalize their teaching by aligning their course materials with the demographics of their specific students. For example, those teaching younger students may want to emphasize self-efficacy enhancements throughout their lessons and projects (Chyung, 2007). This RCT was a first step toward assessing the utility of these treatments and will inform future experimentation. Collecting learners' reported helpfulness and longer-term outcome retention data has provided insights for later personalization trials.

Future Directions for Research

The next area for experimentation should be tailoring these prompts through personalization by matching learners with the messages most likely to help them succeed. Scaling up one-size-fits-all interventions can eliminate most of the gains observed, just as others have seen in the MOOC setting when intentional treatment assignment is not included (Kizilcec et al., 2020). Given the promising impact observed in this experiment, the researcher and broader

Coursera team are well-positioned to start testing these interventions with personalization, which should increase their total impact (Hickey & Urban, 2019; Yu et al., 2017). For example, a female learner younger than 25 years old would likely most benefit from the self-efficacy prompts. Therefore, instead of randomly placing learners into different intervention groups, their demographic information and previous learning patterns could determine their treatment group assignment. This personalization can also be done at the message level, making it more granular and nuanced than the treatment level observed in this experiment.

Given the concentration of these interventions within the first few weeks of the course and the corresponding impact on first-week completion, the researcher also suggests testing additional messages in the later weeks. Extending the intervention into the remainder of the course may extend the strikingly positive impacts observed for week-one completion rates. Additionally, a more robust version of the combination treatment may boost female course completers' final grades, as seen in the experiment for their future STEM enrollments. With how positively younger females responded to the self-efficacy treatment, there is reason to hypothesize that more of these message types for this demographic group may be beneficial as well. Future research should examine these possibilities.

Finally, there is an opportunity for instructors teaching MOOCs to incorporate and test the effectiveness of self-efficacy boosts, planning support, and value relevance in their lessons. For any researchers looking to replicate this study, it would be useful to test in STEM MOOCs with greater initial disparity in the gender completion rate to see if the value relevance would produce similar benefits. While a 1.2% increase closed the gender gap in these 150 most-popular STEM MOOCs, that increase would not be sufficient to erase the gender disparity in completion seen across the wider population of STEM online courses. Further studies could examine this impact on different types of courses.

When presenting a new concept in a video, instructors could reflect on how challenging these skills used to be for them or how much practice was needed to become facile with the material, which others have used to increase students' self-efficacy (Chyung, 2007). Different schedules could be provided to help learners plan their time and continue to progress through the content (Watson et al., 2018). Additionally, learners could write about and share their values as part of an assessment within the course (Miyake et al., 2010; Peters et al., 2017). To assess the true impact of these updates, the course teams could utilize the same A/B testing setup on the Coursera platform as in the current intervention study (Urban & Greenblatt-Kolodny, 2017). While these more extensive changes were not possible during this study, it would be beneficial for a university or company designing their own MOOCs to test the utility of these motivation-based interventions within their pedagogical plan.

Conclusion

Addressing female learners' intrinsic motivation through light-touch interventions in online courses on Coursera eliminated the gender gap in initial retention and course completion. Short, in-course prompts focused on increasing females' self-efficacy, planning, and value relevance resulted in nearly identical STEM MOOC first-week completion rates for females and males. In the learner group receiving value relevance emphasis prompts, this increase was retained throughout the course, thereby closing the gender gap in STEM MOOC completion. Employing the self-determination theory of intrinsic motivation (Deci & Ryan, 2000), this researcher designed and implemented an RCT experiment to document the causal impact of brief, novel messages to increase female learners' motivation to progress. While females have traditionally not completed STEM MOOCs at the same rate as their male peers (Grella & Meinel, 2016; Healy, 2017), applying this intervention offers new insights regarding how online course platforms and providers can better support underrepresented groups and erase this persistent gender gap.

The results of this intervention study illustrate the utility and power of connecting online learning experiences with individuals' personal values to further their retention. More than 170 additional females receiving the value relevance treatment completed their STEM MOOC in this treatment group than in the control condition. With learners worldwide using course completion certificates earned on Coursera to gain promotions and new job opportunities (Hadavand et al., 2018; Hickey et al., 2020), each female with new course completions would have the opportunity for greater career advancement. Evidence from data science MOOC completers suggests a greater than $8,000 salary increase and a 30% greater likelihood of job mobility resulting from these courses (Hadavand et al., 2018). Crucially, these impressive benefits are only seen for those who complete the MOOCs (Hadavand et al., 2018; Hickey et al., 2020). With the return on investment at least one order of magnitude larger than the educational costs to the learner, completing MOOCs has clear economic value for individuals.

Plus, increasing online course persistence impacts not only the learner but also their broader society. Recent research highlights how learning in MOOCs can significantly increase retaining employment and decrease the likelihood of unemployed periods, benefiting both individual and company productivity (Castaño-Muñoz & Rodrigues, 2021). Hiring managers have learned to equate the presence of MOOC certificates on a resumé with two additional years of work experience (Rivas et al., 2020). This update to companies' evaluation of MOOC completions suggests how low-cost, large-scale online learning can help companies and governments prepare workers for more advanced roles. In addition, recent randomized experiments suggest similar MOOC learning outcomes to in-person or hybrid teaching, with the MOOC version delivered at much lower costs (Chirikov et al., 2020).

Finally, diversity in STEM is a global issue. Economists argue that, when females remain disproportionately unrepresented in STEM fields, countries fail to meet their capacity for innovation, potential for entrepreneurship, or competitiveness on the world stage (Beede et al., 2011; Dilli & Westerhuis, 2018). STEM MOOCs are an affordable, effective method for helping diverse learners enter and remain in high-demand careers (Castaño-Muñoz & Rodrigues, 2021; Hadavand et al., 2018). An increase in females completing STEM MOOCs has cascading benefits for these learners, plus their families, employers, and nations. Looking ahead, educators and researchers need to ask hard questions around equity, inclusion, and belonging to realize the full potential of MOOCs for uplifting females and underserved populations worldwide.

References

Aguilar, S. J. (2018). Learning analytics: At the nexus of big data, digital innovation, and social justice in education. *TechTrends*, *62*, 37–45. https://doi.org/10.1007/s11528-017-0226-9

Alario-Hoyos, C., Estévez-Ayres, I., Pérez Sanagustín, M., Leony, D., & Delgado Kloos, C. (2015). MyLearningMentor: A mobile app to support learners participating in MOOCs. *Journal of Universal Computer Science*, *21*(5), 735–753. https://doi.org/10.3217/jucs-021-05-0735

Alario-Hoyos, C., Estévez-Ayres, I., Pérez-Sanagustín, M., Kloos, C. D., & Fernández-Panadero, C. (2017). Understanding learners' motivation and learning strategies in MOOCs. *International Review of Research in Open and Distance Learning*, *18*(3), 119–137. https://doi.org/10.19173/irrodl.v18i3.2996

Allione, G., & Stein, R. M. (2016). Mass attrition: An analysis of drop out from principles of microeconomics MOOC. *Journal of Economic Education*, *47*(2), 174–186. https://doi.org/10.1080/00220485.2016.1146096

Bamberger, M., Tarsilla, M., & Hesse-Biber, S. (2016). Why so many "rigorous" evaluations fail to identify unintended consequences of development programs: How mixed methods can contribute. *Evaluation and Program Planning*, *55*, 155–162. https://doi.org/10.1016/j.evalprogplan.2016.01.001

Bandura, A. (1977a). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, *84*(2), 191–215. https://doi.org/10.1037/0033-295X.84.2.191

Bandura, A. (1977b). *Social learning theory*. Prentice Hall.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Prentice Hall.

Baranowski, T., & Stables, G. (2000). Process evaluations of the 5-a-day projects. *Health Education &Behavior*, *27*(2), 157–166. https://doi.org/10.1177/109019810002700202

Bayeck, R. Y. (2016). Exploratory study of MOOC learners' demographics and motivation: The case of students involved in groups. *Open Praxis*, *8*(3), 223–233. https://doi.org/10.5944/openpraxis.8.3.282

Beede, D., Julian, T., & Langdon, D. (2011). Women in STEM: A gender gap to innovation. *Economics and Statistics Administration Issue Brief*, *4*(11). https://doi.org/10.2139/ssrn.1964782

Bernacki, M. L., Vosicka, L., & Utz, J. C. (2019). Can a brief, digital skill training intervention help undergraduates "learn to learn" and improve their STEM achievement? *Journal of Educational Psychology*, *112*(4), 765–781. https://doi.org/10.1037/edu0000405

Blackburn, H. (2017). The status of women in STEM in higher education: A review of the literature 2007–2017. *Science and Technology Libraries*, *36*(3), 235–273. https://doi.org/10.1080/0194262X.2017.1371658

Bonk, C. J., Zhu, M., Kim, M., Xu, S., Sabir, N., & Sari, A. R. (2018). Pushing toward a more personalized MOOC: Exploring instructor selected activities, resources, and technologies for MOOC design and implementation. *International Review of Research in Open and Distance Learning*, *19*(4), 92–115. https://doi.org/10.19173/irrodl.v19i4.3439

Bouarab, C., Thompson, B., & Polter, A. M. (2019). VTA GABA neurons at the interface of stress and reward. *Frontiers in Neural Circuits*, *13*, 1–12. https://doi.org/10.3389/fncir.2019.00078

Bronfenbrenner, U. (1979). *The ecology of human development: Experiments by nature and design*. Harvard University Press.

Bronfenbrenner, U. (1986). Ecological models of human development. *Oxford Review of Education*, *12*(1), 77–93. https://doi.org/10.1080/0305498860120107

Brooks, C., Gardner, J., & Chen, K. (2018). How gender cues in educational video impact participation and retention. *Proceedings of International Conference of the Learning Sciences, ICLS*, *3*, 1835–1842. https://doi.org/10.22318/cscl2018.1835

Buffington, C., Cerf, B., Jones, C., & Weinberg, B. A. (2016). STEM training and early career outcomes of female and male graduate students: Evidence from UMETRICS data linked to the 2010 census. *American Economic Review*, *106*(5), 333–338. https://doi.org/10.1257/aer.p20161124

Carlone, H. B., & Johnson, A. (2007). Understanding the science experiences of successful women of color: Science identity as an analytic lens. *Journal of Research in Science Teaching*, *44*(8), 1086–1109. https://doi.org/10.1002/tea

Carrell, S. E., West, J. E., & Page, M. E. (2013). Sex and science: How professor gender perpetuates the gender gap. *National Bureau of Economic Research*, *53*(9), 1689–1699. https://doi.org/10.1017/CBO9781107415324.004

Castaño-Muñoz, J., & Rodrigues, M. (2021). Open to MOOCs? Evidence of their impact on labour market outcomes. *Computers and Education*, *173*, 104289. https://doi.org/10.1016/J.COMPEDU.2021.104289

Chandrasekaran, M. K., Ragupathi, K., Kan, M. Y., & Tan, B. C. Y. (2015). Towards feasible instructor intervention in MOOC discussion forums. *2015 International Conference on Information Systems: Exploring the Information Frontier*, 1–9. https://www.comp.nus.edu.sg/~kanmy/papers/icis2015.pdf

Charles, M., & Bradley, K. (2009). Indulging our gendered selves? Sex segregation by field of study in 44 countries. *American Journal of Sociology*, *114*(4), 924–976. https://doi.org/10.1086/595942

Charleston, L. J., Lang, N. M., Adserias, R. P., & Jackson, J. F. L. (2014). Intersectionality and STEM: The role of race and gender in the academic pursuits of African American women in STEM. *Journal of Progressive Policy & Practice*, *2*(3), 273–293. https://caarpweb.org/wp-content/uploads/2014/12/Charleston-Adserias-Lang-Jackson-2014.pdf

Cheryan, S., Ziegler, S. A., Montoya, A. K., & Jiang, L. (2017). Why are some STEM fields more gender balanced than others? *Psychological Bulletin*, *143*(1), 1–35. https://doi.org/10.1037/bul0000052

Chirikov, I., Semenova, T., Maloshonok, N., Bettinger, E., & Kizilcec, R. F. (2020). Online education platforms scale college STEM instruction with equivalent learning outcomes at lower cost. *Science Advances*, *6*(15), 1–11. https://doi.org/10.1126/sciadv.aay5324

Christie, C. A., Lemire, S., & Inkelas, M. (2017). Understanding the similarities and distinctions between improvement science and evaluation. In C. A. Christie, M. Inkelas, & S. Lemire (Eds.), *Improvement Science in Evaluation: Methods and Uses. New Directions for Evaluation* (Issue 153, pp. 11–21).

Chyung, S. Y. (2007). Age and gender differences in online behavior, self-efficacy, and academic performance. *The Quarterly Review of Distance Education*, *8*(3), 213–222. https://www.infoagepub.com/qrde-issue.html

Cooksy, L. J., Gill, P., & Kelly, P. A. (2001). The program logic model as an integrative framework for a multimethod evaluation. *Evaluation and Program Planning*, *24*(2), 119–128. https://doi.org/10.1016/S0149-7189(01)00003-9

Coursera. (2022a). *Product update: Personalized schedules*. Unpublished internal company document.

Coursera. (2022b, April 18). *10 things that have always been true about Coursera learners*. Coursera Blog. https://blog.coursera.org/10-things-that-have-always-been-true-about-coursera-learners/

Creswell, J. W., & Miller, D. L. (2000). Determining validity in qualitative inquiry. *Theory into Practice*, *39*(3), 124–130. https://doi.org/10.1021/cen-v025n050.p3743

Creswell, J. W., & Plano Clark, V. L. (2011). *Designing and conducting mixed methods research*. SAGE Publications Inc. https://doi.org/10.1007/978-1-4419-5659-0_519

Creswell, J. W., & Plano Clark, V. L. (2018). *Analyzing and interpreting data in mixed methods research* (3rd ed.). SAGE Publications Inc. https://doi.org/10.1007/978-1-4419-5659-0_519

Crues, R. W., Henricks, G. M., Perry, M., Bhat, S., Anderson, C. J., Shaik, N., & Angrave, L. (2018). How do gender, learning goals, and forum participation predict persistence in a computer science MOOC? *ACM Transactions on Computing Education*, *18*(4), 1–14. https://doi.org/10.1145/3152892

Dasgupta, N., Scircle, M. M. M., & Hunsinger, M. (2015). Female peers in small work groups enhance women's motivation, verbal participation, and career aspirations in engineering. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(16), 4988–4993. https://doi.org/10.1073/pnas.1422822112

de Barba, P. G., Kennedy, G. E., & Ainley, M. D. (2016). The role of students' motivation and participation in predicting performance in a MOOC. *Journal of Computer Assisted Learning*, *32*(3), 218–231. https://doi.org/10.1111/jcal.12130

Deci, E. L., & Ryan, R. M. (2000). The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry*, *11*(4), 227–268. https://doi.org/10.1360/982004-431

Dell, E. M., Verhoeven, Y., Christman, J. W., & Garrick, R. D. (2018). Using self-determination theory to build communities of support to aid in the retention of women in engineering. *European Journal of Engineering Education*, *43*(3), 344–359. https://doi.org/10.1080/03043797.2017.1410522

Di Domenico, S. I., & Ryan, R. M. (2017). The emerging neuroscience of intrinsic motivation: A new frontier in self-determination research. *Frontiers in Human Neuroscience*, *11*(145), 1–14. https://doi.org/10.3389/fnhum.2017.00145

Diederen, K. M. M. J., Spencer, T., Vestergaard, M. D. D., Fletcher, P. C. C., & Schultz, W. (2016). Adaptive prediction error coding in the human midbrain and striatum facilitates behavioral adaptation and learning efficiency. *Neuron*, *90*(5), 1127–1138. https://doi.org/10.1016/j.neuron.2016.04.019

Dilli, S., & Westerhuis, G. (2018). How institutions and gender differences in education shape entrepreneurial activity: A cross-national perspective. *Small Business Economics*, *51*(2), 371–392. https://doi.org/10.1007/s11187-018-0004-x

Dubuc, B. (2020). *The brain from top to bottom*. McGill University. https://thebrain.mcgill.ca/

Dyrberg, N. R., & Holmegaard, H. T. (2019). Motivational patterns in STEM education: A self-determination perspective on first year courses. *Research in Science and Technological Education*, *37*(1), 90–109. https://doi.org/10.1080/02635143.2017.1421529

Elliott, V. (2018). Thinking about the coding process in qualitative data analysis. *Qualitative Report*, *23*(11), 2850–2861. https://doi.org/10.46743/2160-3715/2018.3560

Eriksson, T., Adawi, T., & Stöhr, C. (2017). "Time is the bottleneck": A qualitative study exploring why learners drop out of MOOCs. *Journal of Computing in Higher Education*, *29*(1), 133–146. https://doi.org/10.1007/s12528-016-9127-8

Ertl, B., Luttenberger, S., & Paechter, M. (2017). The impact of gender stereotypes on the self-concept of female students in STEM subjects with an under-representation of females. *Frontiers in Psychology*, *8*, 1–11. https://doi.org/10.3389/fpsyg.2017.00703

Feldon, D. F., Franco, J., Chao, J., Peugh, J., & Maahs-Fladung, C. (2018). Self-efficacy change associated with a cognitive load-based intervention in an undergraduate biology course. *Learning and Instruction*, *56*, 64–72. https://doi.org/10.1016/j.learninstruc.2018.04.007

García-Martín, J., & García-Sánchez, J.-N. (2020). The effectiveness of four instructional approaches used in a MOOC promoting personal skills for success in life. *Revista de Psicodidáctica (English Ed.)*, *25*(1), 36–44. https://doi.org/10.1016/j.psicoe.2019.08.001

Gayles, J. G., & Ampaw, F. (2016). To stay or leave: Factors that impact undergraduate women's persistence in science majors. *NASPA Journal About Women in Higher Education*, *9*(2), 133–151. https://doi.org/10.1080/19407882.2016.1213642

Glassberg Sands, E., Reddick, R., & Karsten, E. (2021). *Women and skills report: Addressing gender gaps through online learning*. https://about.coursera.org/press/wp-content/uploads/2021/09/Coursera-Women-and-Skills-Report-2021.pdf

Good, C., Rattan, A., & Dweck, C. S. (2012). Why do women opt out? Sense of belonging and women's representation in mathematics. *Journal of Personality and Social Psychology*, *102*(4), 700–717. https://doi.org/10.1037/a0026659

Grella, C., & Meinel, C. (2016). MOOCs as a promoter of gender diversity in STEM? *The International Scientific Conference ELearning and Software for Education*, *2*(1), 516–521. https://doi.org/10.12753/2066-026X-16-164

Guba, E. (1981). Criteria for assessing the trustworthiness of naturalistic inquiries. *Educational Communication and Technology: A Journal of Theory, Research, and Development*, *29*(2), 75–91. https://doi.org/10.1007/BF02766777

Guiso, L., Monte, F., Sapienza, P., & Zingales, L. (2008). Culture, gender, and math. *Science*, *320*(5880), 1164–1165. https://doi.org/10.1126/science.1154094

Gütl, C., Rizzardini, R. H., Chang, V., & Morales, M. (2014). Attrition in MOOC: Lessons learned from drop-out students. In L. Uden, J. Sinclair, Y.-H. Tao, & D. Liberona (Eds.), *Learning technology for education in cloud. MOOC and big data* (pp. 37–48). Springer International Publishing. https://doi.org/10.1007/978-3-319-10671-7_4

Hadavand, A., Gooding, I., & Leek, J. (2018). *Ca MOOC programs improve student employment prospects?* https://ssrn.com/abstract=3260695

Handoko, E., Gronseth, S. L., Mcneil, S. G., Bonk, C. J., & Robin, B. R. (2019). Goal setting and MOOC completion: A study on the role of self-regulated learning in student performance in massive open online courses. *International Review of Research in Open and Distributed Learning*, *20*(3), 39–58. https://doi.org/10.19173/irrodl.v20i4.4270

Hardiman, M. M., JohnBull, R. M., & Carran, D. (2019). The effects of arts-integrated instruction on students' memory for science content: Results from a randomized control trial study. *Journal of Chemical Information and Modeling*, *53*(9), 1689–1699. https://doi.org/10.1017/CBO9781107415324.004

Healy, P. A. (2017). Georgetown's first six MOOCs: Completion, intention, and gender achievement gaps. *Undergraduate Economic Review*, *14*, 1–37. http://digitalcommons.iwu.edu/

Hickey, A., Bakthavachalam, V., Urban, A., & Kolodny, T. (2018). *A growth mindset can reduce the gender gap in STEM*. Coursera Blog. https://blog.coursera.org/a-growth-mindset-can-reduce-the-gender-gap-in-stem/

Hickey, A., & Urban, A. (2019). *The machine learning advantage*. Unpublished internal company document.

Hickey, A., Urban, A., & Karsten, E. (2020). *Drivers of quality in online learning*. https://about.coursera.org/press/wp-content/uploads/2020/10/Coursera_DriversOfQuality_Book_MCR-1126-V4-lr.pdf

Hiser, J., & Koenigs, M. (2018). The multifaceted role of ventromedial prefrontal cortex in emotion, decision-making, social cognition, and psychopathology. *Biological Psychiatry*, *83*(8), 638–647. https://doi.org/10.1016/j.biopsych.2017.10.030

Hsieh, H. F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, *15*(9), 1277–1288. https://doi.org/10.1177/1049732305276687

Huang, X., & Mayer, R. E. (2019). Adding self-efficacy features to an online statistics lesson. *Journal of Educational Computing Research*, *57*(4), 1003–1037. https://doi.org/10.1177/0735633118771085

Hughes, C. C., Schilt, K., Gorman, B. K., & Bratter, J. L. (2017). Framing the faculty gender gap: A view from STEM doctoral students. *Gender, Work and Organization*, *24*(4), 398–416. https://doi.org/10.1111/gwao.12174

Ihsen, S., Jeanrenaud, Y., Vries, P. De, & Hennis, T. (2015). Gender and diversity in engineering MOOCs: A first appraisal. *43rd Annual Conference of the European Society for Engineering Education*, 1–9. https://doi.org/10.13140/RG.2.1.1345.2886

Jansen, R. S., van Leeuwen, A., Janssen, J., Conijn, R., & Kester, L. (2020). Supporting learners' self-regulated learning in massive open online courses. *Computers and Education*, *146*(103771), 1–17. https://doi.org/10.1016/j.compedu.2019.103771

Jiang, S., Schenke, K., Eccles, J. S., Xu, D., & Warschauer, M. (2016). Females' enrollment and completion in science, technology, engineering, and mathematics massive open online courses. *ArXiv Preprint*. https://doi.org/arXiv:1608.05131

Jiang, S., Schenke, K., Eccles, J. S., Xu, D., & Warschauer, M. (2018). Cross-national comparison of gender differences in the enrollment in and completion of science, technology, engineering, and mathematics massive open online courses. *PLoS ONE*, *13*(9), 1–16. https://doi.org/https://doi.org/10.1371/journal.pone.0202463

Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher*, *33*(7), 14–26. https://doi.org/10.3102/0013189X033007014

Johnson, R. B., Onwuegbuzie, A. J., & Turner, L. A. (2007). Towards a definition of mixed methods research. *The Cambridge Handbook of Sociology*, *1*(2), 112–133. https://doi.org/10.1017/9781316418376.015

Jones, B. D., Ruff, C., & Paretti, M. C. (2013). The impact of engineering identification and stereotypes on undergraduate women's achievement and persistence in engineering. *Social Psychology of Education*, *16*(3), 471–493. https://doi.org/10.1007/s11218-013-9222-x

Kay, J., Reimann, P., Diebold, E., & Kummerfeld, B. (2013). MOOCs: So many learners, so much potential... *IEEE Intelligent Systems*, *28*(3), 70–77. https://doi.org/10.1109/MIS.2013.66

Kizilcec, R. F., & Brooks, C. (2017). Diverse big data and randomized field experiments in MOOCs. In C. Lang, G. Siemens, A. Wise, & D. Gašević (Eds.), *Handbook of learning analytics* (pp. 211–222). Society for Learning Analytics Research. https://doi.org/10.18608/hla17.018

Kizilcec, R. F., & Cohen, G. L. (2017). Eight-minute self-regulation intervention raises educational attainment at scale in individualist but not collectivist cultures. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(17), 4348–4353. https://doi.org/10.1073/pnas.1611898114

Kizilcec, R. F., Davis, G. M., & Cohen, G. L. (2017). Towards equal opportunities in MOOCs. *Proceedings of the Fourth (2017) ACM Conference on Learning @ Scale - L@S '17*, 121–130. https://doi.org/10.1145/3051457.3051460

Kizilcec, R. F., & Halawa, S. (2015). Attrition and achievement gaps in online learning. *Proceedings of the Second ACM Conference on Learning @ Scale*, 57–66. https://doi.org/10.1145/2724660.2724680

Kizilcec, R. F., Pérez-Sanagustín, M., & Maldonado, J. J. (2017). Self-regulated learning strategies predict learner behavior and goal attainment in massive open online courses. *Computers and Education*, *104*, 18–33. https://doi.org/10.1016/j.compedu.2016.10.001

Kizilcec, R. F., Reich, J., Yeomans, M., Dann, C., Brunskill, E., Lopez, G., Turkay, S., Williams, J. J., & Tingley, D. (2020). Scaling up behavioral science interventions in online education. *Proceedings of the National Academies of Sciences*, *117*(26), 14900–14905. https://doi.org/10.1073/pnas.1921417117

Kizilcec, R. F., & Saltarelli, A. J. (2019). Psychologically inclusive design: Cues impact women's participation in STEM education. *CHI Conference on Human Factors in Computing Systems*, 1–10. https://doi.org/10.1145/3290605.3300704

Kizilcec, R. F., Saltarelli, A. J., Reich, J., & Cohen, G. L. (2017). Closing global achievement gaps in MOOCs. *Science*, *355*(6322), 16–18. https://doi.org/10.1126/science.aag2063

Knapp, H. (2018). *Intermediate statistics using SPSS*. SAGE Publications Inc. https://doi.org/10.4135/9781071802625

Lambert, S. R. (2020). Do MOOCs contribute to student equity and social inclusion? A systematic review 2014–18. *Computers and Education*, *145*, 1–17. https://doi.org/10.1016/j.compedu.2019.103693

Lee, D., Watson, S. L., & Watson, W. R. (2020). The relationships between self-efficacy, task value, and self-regulated learning strategies in massive open online courses. *International Review of Research in Open and Distance Learning*, *21*(1), 1–22. https://doi.org/10.19173/irrodl.v20i5.4564

León, J., Núñez, J. L., & Liew, J. (2015). Self-determination and STEM education: Effects of autonomy, motivation, and self-regulated learning on high school math achievement. *Learning and Individual Differences*, *43*, 156–163. https://doi.org/10.1016/j.lindif.2015.08.017

Leviton, L. C., & Lipsey, M. W. (2007). A big chapter about small theories: Theory as method: Small theories of treatments. In *New Directions for Evaluations* (Issue 114, pp. 27–62). https://doi.org/10.1002/ev.224

Liu, W. C., Wang, C. K. J., Kee, Y. H., Koh, C., Lim, B. S. C., & Chua, L. (2014). College students' motivation and learning strategies profiles and academic achievement: A self-determination theory approach. *Educational Psychology*, *34*(3), 338–353. https://doi.org/10.1080/01443410.2013.785067

Lochmiller, C. R., & Lester, J. N. (2017). *An introduction to educational research: Connecting methods to practice*. SAGE Publications Inc.

Lodge, J. M., & Corrin, L. (2017). What data and analytics can and do say about effective learning. *Science of Learning*, *2*(1), 1–2. https://doi.org/10.1038/s41539-017-0006-5

Loizzo, J., & Ertmer, P. A. (2016). MOOCocracy: The learning culture of massive open online courses. *Educational Technology Research and Development*, *64*(6), 1013–1032. https://doi.org/10.1007/s11423-016-9444-7

Loizzo, J., Ertmer, P. A., Watson, W. R., & Watson, S. L. (2017). Adult MOOC learners as self-directed: Perceptions of motivation, success, and completion. *Online Learning Journal*, *21*(2). https://doi.org/10.24059/olj.v21i2.889

Lung-Guang, N. (2019). Decision-making determinants of students participating in MOOCs: Merging the theory of planned behavior and self-regulated learning model. *Computers and Education*, *134*, 50–62. https://doi.org/10.1016/j.compedu.2019.02.004

Macphee, D., Farro, S., & Canetto, S. S. (2013). Academic self-efficacy and performance of underrepresented STEM majors: Gender, ethnic, and social class patterns. *Analyses of Social Issues and Public Policy*, *13*(1), 347–369. https://doi.org/10.1111/asap.12033

Mannella, F., Gurney, K., & Baldassarre, G. (2013). The nucleus accumbens as a nexus between values and goals in goal-directed behavior: A review and a new hypothesis. *Frontiers in Behavioral Neuroscience*, *7*(135), 1–29. https://doi.org/10.3389/fnbeh.2013.00135

Martin, N. I., Kelly, N., & Terry, P. C. (2018). A framework for self-determination in massive open online courses: Design for autonomy, competence, and relatedness. *Australasian Journal of Educational Technology*, *34*(2), 35–55. https://doi.org/10.14742/ajet.3722

Master, A., & Meltzoff, A. N. (2016). Building bridges between psychological science and education: Cultural stereotypes, STEM, and equity. *Prospects*, *46*(2), 215–234. https://doi.org/10.1007/s11125-017-9391-z

Mayer, R. E. (2019). Thirty years of research on online learning. *Applied Cognitive Psychology*, *33*(2), 152–159. https://doi.org/10.1002/acp.3482

Mertens, D. M. (2018). *Mixed methods designs in evaluation*. SAGE Publications Inc.

Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative data analysis: A methods sourcebook* (3rd ed.). SAGE Publications Inc. https://doi.org/https://doi.org/10.4135/9781849209939.n11

Miller, D. I., Eagly, A. H., & Linn, M. C. (2015). Women's representation in science predicts national gender-science stereotypes: Evidence from 66 nations. *Journal of Educational Psychology*, *107*(3), 631–644. https://doi.org/10.1037/edu0000005

Miyake, A., Kost-Smith, L. E., Finkelstein, N. D., Pollock, S. J., Cohen, G. L., & Ito, T. A. (2010). Reducing the gender achievement gap in college science: A classroom study of values affirmation. *Science*, *330*(6008), 1234–1237. https://doi.org/10.1126/science.1195996

Moulton, S. T. (2014). *Applying psychological science to higher education: Key findings and open questions*. https://hilt.harvard.edu/

Murayama, K., Matsumoto, M., Izuma, K., Sugiura, A., Ryan, R. M., Deci, E. L., & Matsumoto, K. (2015). How self-determined choice facilitates performance: A key role of the ventromedial prefrontal cortex. *Cerebral Cortex*, *25*(5), 1241–1251. https://doi.org/10.1093/cercor/bht317

Murphy, S., MacDonald, A., Wang, C. A., & Danaia, L. (2019). Towards an understanding of STEM engagement: A review of the literature on motivation and academic emotions. *Canadian Journal of Science, Mathematics and Technology Education*, *19*(3), 304–320. https://doi.org/10.1007/s42330-019-00054-w

National Center for Education Statistics. (2019). Status and trends in the education of racial and ethnic groups 2018. *Institute of Education Sciences*, 182. https://doi.org/10.1037/e571522010-001

National Science Foundation. (2019). *2019 women, minorities, and persons with disabilities report goes live*. https://www.nsf.gov/news/

O'Dea, R. E., Lagisz, M., Jennions, M. D., & Nakagawa, S. (2018). Gender differences in individual variation in academic grades fail to fit expected patterns for STEM. *Nature Communications*, *9*(1). https://doi.org/10.1038/s41467-018-06292-0

OECD. (2015). *The ABC of gender equality in education: Aptitude, behaviour, confidence*. https://doi.org/10.1787/9789264229945-en

OECD. (2019). *Trends shaping education 2019*. https://doi.org/10.1787/trends_edu-2019-en

Onah, D., Sinclair, J., & Boyatt, R. (2014). Dropout rates of massive open online courses: Behavioural patterns. *Proceedings of the 6th International Conference on Education and New Learning Technologies*, 1–10. https://doi.org/10.13140/RG.2.1.2402.0009

Onwuegbuzie, A. J., & Leech, N. (2006). Linking research questions to mixed methods data analysis procedures. *The Qualitative Report*, *11*(3), 474–498. https://doi.org/10.46743/2160-3715/2006.1663

Parson, L. (2018). An institutional ethnography of higher education: The experiences of undergraduate women majoring in math and physics. *Journal of Ethnographic & Qualitative Research*, *13*, 18–33. https://eric.ed.gov/?id=EJ1257110

Peechapol, C., Na-Songkhla, J., Sujiva, S., & Luangsodsai, A. (2018). An exploration of factors influencing self-efficacy in online learning: A systematic review. *International Journal of Emerging Technologies in Learning*, *13*(9), 64–86. https://doi.org/10.3991/ijet.v13i09.8351

Perez, C. C. (2019). *Invisible women: Data bias in a world designed for men*. Abrams.

Peters, E., Shoots-Reinhard, B., Tompkins, M. K., Schley, D., Meilleur, L., Sinayev, A., Tusler, M., Wagner, L., & Crocker, J. (2017). Improving numeracy through values affirmation enhances decision and STEM outcomes. *PLoS ONE*, *12*(7), 1–19. https://doi.org/10.1371/journal.pone.0180674

Pursel, B. K., Zhang, L., Jablokow, K. W., Choi, G. W., & Velegol, D. (2016). Understanding MOOC students: Motivations and behaviours indicative of MOOC completion. *Journal of Computer Assisted Learning*, *32*(3), 202–217. https://doi.org/10.1111/jcal.12131

Qiu, J., Tang, J., Liu, T. X., Gong, J., Zhang, C., Zhang, Q., & Xue, Y. (2016). Modeling and predicting learning behavior in MOOCs. *Proceedings of the Ninth ACM International Conference*, 93–102. https://doi.org/10.1145/2835776.2835842

Rabin, E., Henderikx, M., Kalman, Y. M., & Kalz, M. (2020). What are the barriers to learners' satisfaction in MOOCs and what predicts them? The role of age, intention, self-regulation, self-efficacy and motivation. *Australasian Journal of Educational Technology*, *36*(3), 119–131. https://doi.org/10.14742/AJET.5919

Reddick, R. (2019). Using a glicko-based algorithm to measure in-course learning. *EDM 2019 - Proceedings of the 12th International Conference on Educational Data Mining*, 754–759. https://files.eric.ed.gov/fulltext/ED599237.pdf

Reeves, T. D., & Chiang, J. L. (2019). Effects of an asynchronous online data literacy intervention on pre-service and in-service educators' beliefs, self-efficacy, and practices. *Computers and Education*, *136*, 13–33. https://doi.org/10.1016/j.compedu.2019.03.004

Renzel, D., & Klamma, R. (2013). From micro to macro: Analyzing activity in the ROLE sandbox. *ACM International Conference Proceeding Series*, 250–254. https://doi.org/10.1145/2460296.2460347

Riegle-Crumb, C., & King, B. (2010). Questioning a white male advantage in STEM: Examining disparities in college major by gender and race/ethnicity. *Educational Researcher*, *39*(9), 656–664. https://doi.org/10.3102/0013189X10391657

Riegle-Crumb, C., King, B., Grodsky, E., & Muller, C. (2012). The more things change, the more they stay the same? Prior achievement fails to explain gender inequality in entry into STEM college majors over time. *American Educational Research Journal*, *49*(6), 1048–1073. https://doi.org/10.3102/0002831211435229

Riegle-Crumb, C., & Moore, C. (2014). The gender gap in high school physics: Considering the context of local communities. *Social Science Quarterly*, *95*(1), 253–268. https://doi.org/10.1111/ssqu.12022

Rivas, M. J., Baker, R. B., & Evans, B. J. (2020). Do MOOCs make you more marketable? An experimental analysis of the value of MOOCs relative to traditional credentials and experience: *Https://Doi.Org/10.1177/2332858420973577*, *6*(4). https://doi.org/10.1177/2332858420973577

Roediger, H. L., & Pyc, M. A. (2012). Inexpensive techniques to improve education: Applying cognitive psychology to enhance educational practice. *Journal of Applied Research in Memory and Cognition*, *1*(4), 242–248. https://doi.org/10.1016/j.jarmac.2012.09.002

Rossi, P. H., Lipsey, M. W., & Henry, G. T. (2019). *Evaluation: A systematic approach* (8th ed.). SAGE Publications Inc.

Russell, L. (2017). Can learning communities boost success of women and minorities in STEM? Evidence from the Massachusetts Institute of Technology. *Economics of Education Review*, *61*, 98–111. https://doi.org/10.1016/j.econedurev.2017.10.008

Salkind, N. (2010). *Encyclopedia of research design* (Vols. 1–0). SAGE Publications Inc. https://doi.org/10.4135/9781412961288

Sambe, G., Bouchet, F., & Labat, J.-M. (2017). Towards a conceptual framework to scaffold self-regulation in a MOOC. *Sixième Colloque National Sur La Recherche En Informatique et Ses Applications*, 245–256. https://doi.org/10.1007/978-3-319-72965-7_23

Sandelowski, M. (2000). Focus on research methods: Combining qualitative and quantitative sampling, data collection, and analysis techniques in mixed-method studies. *Research in Nursing & Health*, *23*(3), 246–255. https://doi.org/10.1002/1098-240x(200006)23:3<246::aid-nur9>3.0.co;2-h

Saunders, R. P., Evans, M. H., & Joshi, P. (2005). Developing a process-evaluation plan for assessing health promotion program implementation: A how-to guide. *Health Promotion Practice*, *6*(2), 134–147. https://doi.org/10.1177/1524839904273387

Sax, L. J., Lehman, K. J., Jacobs, J. A., Kanny, M. A., Lim, G., Monje-Paulson, L., & Zimmerman, H. B. (2017). Anatomy of an enduring gender gap: The evolution of women's participation in computer science. *Journal of Higher Education*, *88*(2), 258–293. https://doi.org/10.1080/00221546.2016.1257306

Schunk, D. H., Pintrich, P. R., & Meece, J. L. (2020). *Motivation in education: Theory, research, and applications* (4th ed.). Pearson Merrill Prentice Hall.

Shadish, W., Cook, T., & Campbell, D. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.

Shenton, A. K. (2004). Strategies for ensuring trustworthiness in qualitative research projects. *Education for Information*, *22*, 63–75. https://doi.org/10.3233/EFI-2004-22201

Shin, J. E. L., Levy, S. R., & London, B. (2016). Effects of role model exposure on STEM and non-STEM student engagement. *Journal of Applied Social Psychology*, *46*(7), 410–427. https://doi.org/10.1111/jasp.12371

Simon, R. A., Aulls, M. W., Dedic, H., & Hall, N. C. (2015). Exploring student persistence in STEM programs: A motivational model. *Canadian Journal of Education*, *38*(1), 1–27. http://www.jstor.org/stable/canajeducrevucan.38.1.09

Small, M. L. (2011). How to conduct a mixed methods study: Recent trends in a rapidly growing literature. *Annual Review of Sociology*, *37*, 57–86. https://doi.org/10.1146/annurev.soc.012809.102657

Smeding, A. (2012). Women in science, technology, engineering, and mathematics (STEM): An investigation of their implicit gender stereotypes and stereotypes' connectedness to math performance. *Sex Roles*, *67*(11–12), 617–629. https://doi.org/10.1007/s11199-012-0209-4

Smith, E. (2008). Pitfalls and promises: The use of secondary data analysis in educational research. *British Journal of Educational Studies*, *56*(3), 323–339. https://doi.org/10.1111/j.1467-8527.2008.00405.x

Smyth, F. L., & Nosek, B. A. (2015). On the gender-science stereotypes held by scientists: Explicit accord with gender-ratios, implicit accord with scientific identity. *Frontiers in Psychology*, *6*(415), 1–19. https://doi.org/10.3389/fpsyg.2015.00415

Stolk, J. D., Jacobs, J., Girard, C., & Pudvan, L. (2018). Learners' needs satisfaction, classroom climate, and situational motivations: Evaluating self-determination theory in an engineering context. *Proceedings of the Frontiers in Education Conference, FIE*, 1–5. https://doi.org/10.1109/FIE.2018.8658880

Stolk, J. D., Zastavker, Y. v., & Gross, M. D. (2018). Gender, motivation, and pedagogy in the STEM classroom: A quantitative characterization. *ASEE Annual Conference and Exposition, Conference Proceedings*. https://www.asee.org/public/conferences/106/papers/22784/view

Sujatha, R., & Kavitha, D. (2018). Learner retention in MOOC environment: Analyzing the role of motivation, self-efficacy and perceived effectiveness. *International Journal of Education and Development Using Information and Communication Technology*, *14*(2), 62–74. https://www.learntechlib.org/p/184685/

Sun, G., Cui, T., Guo, W., Beydoun, G., Xu, D., & Shen, J. (2015). Micro learning adaptation in MOOC: A software as a service and a personalized learner model. *Faculty of Engineering and Information Sciences*, *9412*, 174–184. https://doi.org/10.1007/978-3-319-25515-6_16

Teddlie, C., & Tashakkori, A. (2003). Major issues and controversies in the use of mixed methods in the social and behavioral sciences. In *Handbook of mixed methods in social and behavioral research* (Vol. 129, pp. 1926–1936). SAGE Publications Inc.

Thompson, B. (2002). “Statistical,” “practical,” and “clinical”: How many kinds of significance do counselors need to consider? *Journal of Counseling and Development*, *80*, 64–71. https://doi.org/10.1002/j.1556-6678.2002.tb00167.x

Urban, A. (2019). *Nail it, then scale it: Doubling down on data to achieve platform success*. Coursera Blog. https://blog.coursera.org/doubling-down-on-data-to-achieve-platform-success/

Urban, A., & Greenblatt-Kolodny, T. (2017). *Why practice matters: Using data to transform the learning experience*. Coursera Blog. https://blog.coursera.org/using-data-transform-learning-experience/

van den Berg, R. G. (2020). *Effect size: A quick guide*. SPSS Tutorials. https://www.spss-tutorials.com/effect-size/

Vennix, J., den Brok, P., & Taconis, R. (2018). Do outreach activities in secondary STEM education motivate students and improve their attitudes towards STEM? *International Journal of Science Education*, *40*(11), 1263–1283. https://doi.org/10.1080/09500693.2018.1473659

Walton, G. M., Logel, C., Peach, J. M., Spencer, S. J., & Zanna, M. P. (2015). Two brief interventions to mitigate a “chilly climate” transform women’s experience, relationships, and achievement in engineering. *Journal of Educational Psychology*, *107*(2), 468–485. https://doi.org/10.1037/a0037461

Wang, M., & Degol, J. L. (2017). Gender gap in science, technology, engineering, and mathematics (STEM): Current knowledge, implications for practice, policy, and future directions. *Physiology Review*, *29*(1), 119–140. https://doi.org/10.1007/s10648-015-9355-x

Wang, Y., & Baker, R. (2015). Content or platform: Why do students complete MOOCs? *Journal of Online Learning and Teaching*, *11*(1), 17–30. https://jolt.merlot.org/vol11no1/Wang_0315.pdf

Wang, Y., & Baker, R. (2018). Grit and intention: Why do learners complete MOOCs? *International Review of Research in Open and Distributed Learning*, *19*(3), 20–42. https://doi.org/10.19173/irrodl.v19i3.3393

Watson, W. R., Yu, J. H., & Watson, S. L. (2018). Perceived attitudinal learning in a self-paced versus fixed-schedule MOOC. *Educational Media International*, *55*(2), 170–181. https://doi.org/10.1080/09523987.2018.1484044

White, J. L., & Massiha, G. H. (2016). The retention of women in science, technology, engineering, and mathematics: A framework for persistence. *International Journal of Evaluation and Research in Education*, *5*(1), 1–8. https://doi.org/10.11591/ijere.v5i1.4515

Wholey, J. S., Hatry, H. P., & Newcomer, K. E. (Eds.). (2010). *Handbook of practical program evaluation* (3rd ed.). John Wiley & Sons, Inc. https://doi.org/10.1177/0899764011420366

Wiebe, E., Thompson, I., & Behrend, T. (2015). MOOCs from the viewpoint of the learner: A response to Perna et al. (2014). *Educational Researcher*, *44*(4), 252–254. https://doi.org/10.3102/0013189x15584774

Williams, M. M., & George-Jackson, C. E. (2014). Using and doing science: Gender, self-efficacy, and science identity of undergraduate students in STEM. *Journal of Women and Minorities in Science and Engineering*, *20*(2), 99–126. https://doi.org/10.1615/JWomenMinorScienEng.2014004477

Wong, J., Baars, M., Davis, D., Van Der Zee, T., Houben, G. J., & Paas, F. (2019). Supporting self-regulated learning in online learning environments and MOOCs: A systematic review. *International Journal of Human-Computer Interaction*, *35*(4–5), 356–373. https://doi.org/10.1080/10447318.2018.1543084

Xu, D., Huang, W. W., Wang, H., & Heales, J. (2014). Enhancing e-learning effectiveness using an intelligent agent-supported personalized virtual learning environment: An empirical investigation. *Information and Management*, *51*(4), 430–440. https://doi.org/10.1016/j.im.2014.02.009

Yeomans, M., & Reich, J. (2017). Planning prompts increase and forecast course completion in massive open online courses. *ACM International Conference Proceeding Series*, *March 2017*, 464–473. https://doi.org/10.1145/3027385.3027416

Young, J. R. (2020). *Massive study of online teaching ends with surprising—and 'deflating'—result*. EdSurge. https://www.edsurge.com/news/2020-06-17-massive-study-of-online-teaching-ends-with-surprising-and-deflating-result

Yu, H., Miao, C., Leung, C., & White, T. J. (2017). Towards AI-powered personalization in MOOC learning. *Science of Learning*, *2*(1), 1–5. https://doi.org/10.1038/s41539-017-0016-3

Zhu, Y., Pan, Y., & Hu, Y. (2019). Learning desire is predicted by similar neural processing of naturalistic educational materials. *ENeuro*, *6*(5), 1–11. https://doi.org/10.1523/ENEURO.0083-19.2019

Appendix A

Top STEM MOOCs by Active Enrollment on Coursera

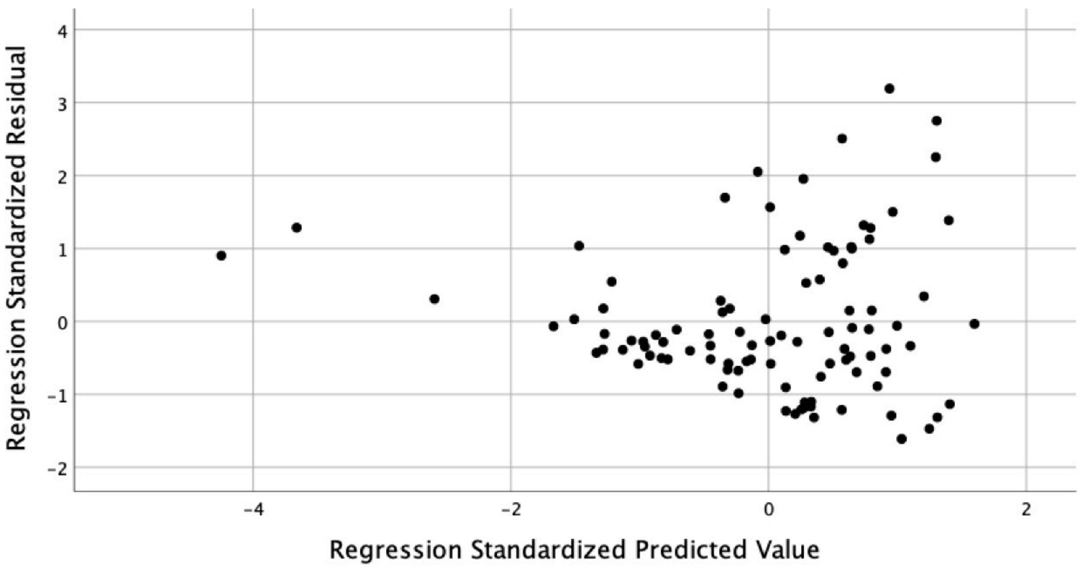
STEM MOOC URL	% of Instructor(s) Who Are Female	% Female Enrollments	% of Enrolled Females Who Complete
algorithm-design-analysis	0%	0.18	0.03
algorithmic-toolbox	0%	0.16	0.07
algorithms-divide-conquer	0%	0.17	0.07
algorithms-part1	0%	0.19	0.01
arduino-platform	0%	0.14	0.12
basic-statistics	0%	0.40	0.05
build-a-computer	0%	0.11	0.03
c-plus-plus-a	0%	0.18	0.01
calculus1	0%	0.28	0.01
ciencia-computacao-python-conceitos	0%	0.20	0.04
como-autoconstruir-tu-vivienda	0%	0.27	0.08
comparch	0%	0.17	0.00
convolutional-neural-networks	0%	0.13	0.35
crypto	0%	0.19	0.03
cryptocurrency	0%	0.17	0.01
cyber-security-domain	0%	0.21	0.04
data-cleaning	0%	0.27	0.24
data-science-course	0%	0.32	0.23
data-scientists-tools	0%	0.30	0.26
data-structures	0%	0.18	0.06
database-management	0%	0.28	0.07
datasciencemathskills	0%	0.29	0.12
decision-making	0%	0.39	0.16
deep-neural-network	0%	0.13	0.48
engineering-mechanics-statics	0%	0.18	0.04
excel-analysis	0%	0.40	0.18
excel-data-analysis	0%	0.43	0.13
exploratory-data-analysis	0%	0.26	0.26
gamification	0%	0.37	0.06
gcp-big-data-ml-fundamentals	0%	0.14	0.28
gcp-fundamentals	0%	0.13	0.27
gcp-infrastructure-foundation	0%	0.10	0.61
gis	0%	0.37	0.13
how-things-work	0%	0.37	0.02
html-css-javascript	0%	0.24	0.12
html-css-javascript-for-web-developers	0%	0.27	0.05
human-computer-interaction	0%	0.46	0.12

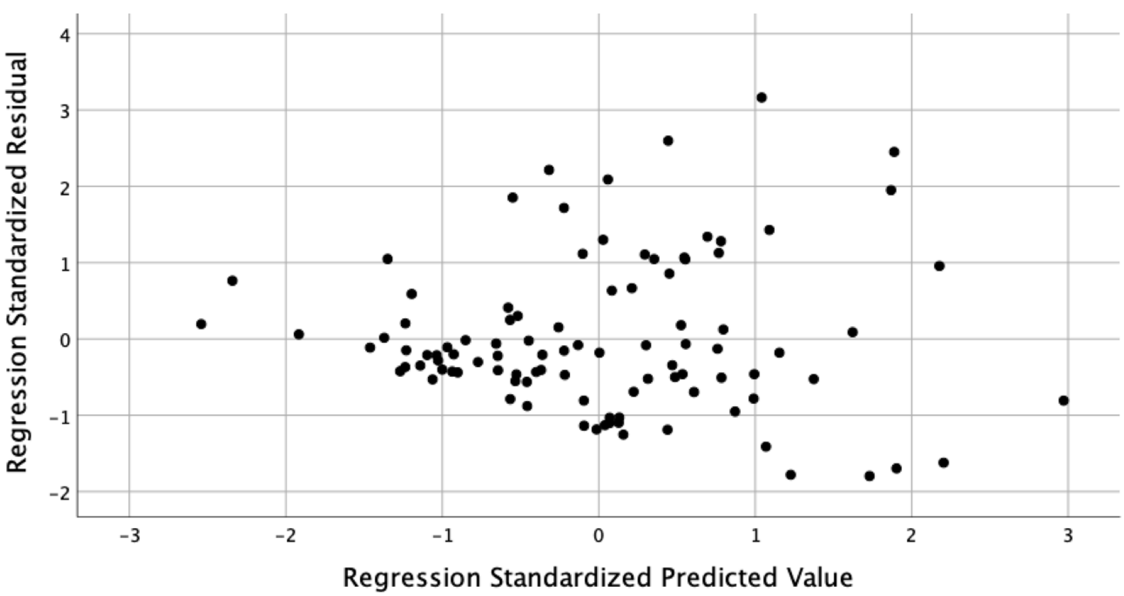
STEM MOOC URL	% of Instructor(s) Who Are Female	% Female Enrollments	% of Enrolled Females Who Complete
interactive-python-1	0%	0.24	0.06
internet-history	0%	0.28	0.10
intro-data-science-programacion-estadistica-r	0%	0.32	0.13
introduction-tensorflow	0%	0.12	0.17
iot	0%	0.18	0.12
linear-algebra-machine-learning	0%	0.17	0.11
logic-introduction	0%	0.29	0.01
machine-learning	0%	0.18	0.07
machine-learning-projects	0%	0.12	0.58
mathematical-thinking	0%	0.29	0.01
matlab	0%	0.29	0.04
neural-networks	0%	0.14	0.02
neural-networks-deep-learning	0%	0.15	0.28
nlp-sequence-models	0%	0.13	0.35
open-source-tools-for-data-science	0%	0.23	0.53
practical-machine-learning	0%	0.21	0.18
progfun1	0%	0.12	0.10
python	0%	0.27	0.22
python-data	0%	0.24	0.44
python-data-analysis	0%	0.22	0.11
python-databases	0%	0.20	0.37
python-for-applied-data-science-ai	0%	0.24	0.34
python-machine-learning	0%	0.19	0.15
python-network-data	0%	0.20	0.34
python-programming-introduction	0%	0.26	0.10
r-programming	0%	0.31	0.12
regression-models	0%	0.27	0.17
robotica-inicial	0%	0.17	0.01
robotics-flight	0%	0.10	0.04
sql-data-science	0%	0.26	0.30
statistical-inference	0%	0.26	0.20
vvedenie-mashinnoe-obuchenie	0%	0.21	0.05
what-is-datascience	0%	0.26	0.41
erasmus-econometrics	22%	0.28	0.02
duke-programming-web	25%	0.32	0.10
java-for-android	25%	0.17	0.04
java-programming	25%	0.25	0.09
hadoop	33%	0.20	0.09
website-coding	33%	0.32	0.14

STEM MOOC URL	% of Instructor(s) Who Are Female	% Female Enrollments	% of Enrolled Females Who Complete
analytics-business-metrics	50%	0.37	0.23
analytics-excel	50%	0.35	0.04
analytics-mysql	50%	0.36	0.11
analytics-tableau	50%	0.35	0.07
android-app	50%	0.23	0.00
big-data-introduction	50%	0.25	0.14
computer-networking	50%	0.21	0.40
electronics	50%	0.17	0.02
excel-para-negocios	50%	0.45	0.04
html	50%	0.35	0.25
intro-to-big-data	50%	0.21	0.22
learn-to-program	50%	0.30	0.07
ml-foundations	50%	0.17	0.14
ml-regression	50%	0.17	0.15
research-methods	50%	0.44	0.11
technical-support-fundamentals	50%	0.26	0.34
algebra-basica	67%	0.40	0.03
object-oriented-java	67%	0.22	0.05
a-programar	100%	0.29	0.06
how-to-create-a-website	100%	0.43	0.01
intro-programming	100%	0.42	0.05
probability-intro	100%	0.30	0.07
sql-for-data-science	100%	0.34	0.08

Appendix B

Residual Plot for Avg Time and Log of Avg Time on Percentage Females Who Complete





Note. The top plot is the residuals from the linear regression model of the average time needed to complete each course on the percentage of females who complete. The bottom plot is the residuals from the linear regression model of the log of the average time needed to complete on the percentage of females who complete.

Appendix C

Logic Model of Proposed Intervention

<table>
<tr><th>Context</th><th colspan="3">Processes</th><th>Outcomes</th></tr>
<tr><td rowspan="4">Female learners in STEM MOOCs on Coursera display, on average, lower self-efficacy, less autonomy, and diminished alignment to the material.

These factors contribute to their lower engagement and persistence in these online courses.

The platform can surface in-course messages to help support these learners and prepare them for greater success.</td><th>Inputs</th><th colspan="2">Outputs</th><td rowspan="2"><u>Short-Term</u>
Increase competence, autonomy, and relatedness, respectively, across treatment groups
<u>Intermediate</u>
Increase motivation and engagement through meeting the psychological needs of SDT
<u>Long-Term</u>
Increase first-week and course completion rates
Increase STEM skill development
Increase enrollments in subsequent MOOCs</td></tr>
<tr><td>Backend platform system to automatically surface messages to newly enrolled learners in STEM MOOCs
A/B testing system that creates random-assignment groups of learners as they enroll into certain MOOCs
Four variants of the activity based on previous self-determination theory (SDT) literature. Learners will be placed in one of the four treatments or the control
Legal approval that this experiment falls within the existing research policies of Coursera as a company</td><td><u>Activities</u>
Learners in the treatment groups engage with in-course messages at the start of the MOOC and across the first few weeks of content, based on their activity
Some learners respond to the prompts and indicate the perceived helpfulness of the messages</td><td><u>Participation</u>
All learners who enroll in the identified 150 STEM MOOCs on Coursera from December 8, 2021, to March 20, 2022</td></tr>
<tr><td colspan="2">Assumptions
Learners will continue enrolling in high numbers in STEM MOOCs on Coursera during the months of this intervention study
A meaningful subset of the learners in the treatment groups will engage with the online prompts shown</td><td colspan="2">External Factors
Learners’ competing priorities of home and work responsibilities
Natural disasters or other crises leading to loss of internet or financial ability to continue in the MOOC on Coursera</td></tr>
</table>

Appendix D

Process Evaluation Plan for Intervention to Support Female Learners in STEM MOOCs

Process Evaluation Question	Process Evaluation Component	Indicator	Data Source	Data Collection	Data Analysis	Frequency
To what extent did the intervention reach the target learner group?	Reach (Baranowski & Stables, 2000)	Total number of learners and the percentage female	Learners	As reported by the backend of the Coursera platform and as assigned by the experimentation tool	Descriptive statistics and statistical tests by treatment/control groups and by gender	Assessed monthly after initial implementation and at experiment close
To what extent did learners find the prompt helpful?	Exposure (Baranowski & Stables, 2000)	Percentage of learners who report the prompt as "helpful"	Learners	Learners' self-reported helpful/unhelpful rating of the prompt itself collected at the time of pop-up on the Coursera platform	Descriptive statistics and statistical tests by treatment/control groups and by gender	Assessed monthly after initial implementation and at experiment close

Appendix E

Outcome Evaluation Plan for Intervention to Support Female Learners in STEM MOOCs

Outcome Evaluation Question	Construct	Data Source	Data Collection	Data Analysis	Frequency
What differences in impact did each intervention have on week-one and course completions?	Persistence	Learners' behavior	Learners' completion rate for the first week and full course	Descriptive statistics and statistical tests by treatment/control groups and by gender	Assessed one month after the experiment closed
What differences in impact did each intervention have on course completers' performance and skill development?	Skill Development	Learners' performance	Learners' average "skill score" increase, determined by performance on in-course assessments	Descriptive statistics and statistical tests by treatment/control groups and by gender	Assessed one month after the experiment closed
How did the intervention impact female learners' self-reported reasons for dropping out of the STEM MOOCs before completing?	Dropoff	Learners' written responses	Learners' self-reported reasons for stopping the STEM MOOC before completing, taken from an open-ended survey question	Qualitative coding by theme for females by treatment/control group	Survey sent after three weeks of inactivity, assessed one month after the experiment closed
To what extent did the intervention spark learners to continue learning in other MOOCs?	Continued Learning	Learners' behavior	Average number of additional MOOCs learners enroll in after joining this experiment	Descriptive statistics and statistical tests by treatment/control groups and by gender	Assessed one month after the experiment closed

Appendix F

150 STEM MOOCs for Intervention Sample

STEM MOOC URL	Active Enrolled Learners
machine-learning	3,432,151
python	1,787,236
covid-19-contact-tracing	1,057,399
neural-networks-deep-learning	764,678
python-data	595,479
algorithms-part1	595,110
introduction-psychology	564,124
r-programming	465,923
data-scientists-tools	455,428
python-data-analysis	441,914
food-and-health	425,559
html-css-javascript-for-web-developers	424,170
cryptocurrency	378,804
python-network-data	375,186
deep-neural-network	320,600
childnutrition	316,510
html	310,495
duke-programming-web	295,282
convolutional-neural-networks	288,644
matlab	287,399
algorithmic-toolbox	277,697
physiology	264,122
psychological-first-aid	261,031
social-psychology	256,575
sql-for-data-science	249,370
introduction-psych	244,858
happiness	242,345
ml-foundations	240,619
python-databases	235,043
nlp-sequence-models	226,022
mathematical-thinking	217,779
crypto	205,873
java-programming	193,899
python-basics	188,775
introduction-tensorflow	186,888
linear-algebra-machine-learning	184,531
iot	179,101
big-data-introduction	177,066
object-oriented-java	175,386
datasciencemathskills	169,720
learn-to-program	169,632

STEM MOOC URL	Active Enrolled Learners
sciwrite	168,696
science-exercise	167,490
user-experience-design	167,103
basic-statistics	166,294
vital-signs	163,805
medical-neuroscience	161,310
analytics-business-metrics	158,737
probability-intro	157,890
engineering-mechanics-statics	149,784
research-methods	135,415
intro-programming	135,405
algorithms-divide-conquer	134,978
excel-data-analysis	134,245
data-structures	131,572
python-machine-learning	130,896
bootstrap-4	130,212
python-programming-introduction	129,951
electronics	127,947
neurobiology	122,968
introduction-cybersecurity-cyber-attacks	121,855
python-data-visualization	115,385
analytics-tableau	113,839
astro	113,609
java-for-android	112,664
web-development	112,032
programming-fundamentals	109,067
introcss	108,769
everyday-excel-part-1	108,177
database-management	106,675
c-for-everyone	105,038
data-visualization-tableau	105,029
introduction-to-ai	104,309
blockchain-basics	102,508
build-a-computer	99,226
positive-psychiatry	98,473
javascript	97,964
gis	92,346
wind-energy	90,552
introduction-to-data-analytics	90,335
everyday-parenting	90,154
bioinformatics	88,967
intro-to-deep-learning	88,101
introduction-to-calculus	87,554
agile-atlassian-jira	86,041
data-analytics-business	81,062

STEM MOOC URL	Active Enrolled Learners
front-end-react	77,666
excel-vba-for-creative-problem-solving-part-1	77,169
machine-learning-duke	75,686
animal-welfare	75,264
programming-languages	74,141
systematic-review	73,562
clinical-research	72,900
web-applications-php	71,980
what-is-a-proof	69,253
cloud-computing-basics	68,642
autodesk-autocad-design-drafting	68,561
agile-software-development	67,929
quantitative-methods	67,282
intro-self-driving-cars	65,960
covid-19-training-healthcare	63,435
forensic-science	63,203
algorithms-part2	62,673
uol-machine-learning-for-all	62,192
applied-data-science-capstone	60,695
java-programming-arrays-lists-data	60,007
cs-programming-java	59,446
classification-vector-spaces-in-nlp	58,634
server-side-nodejs	57,908
positive-psychology-visionary-science	57,850
weight-management-beyond-balancing-calories	57,383
clinical-terminology	55,801
excel-basics-data-analysis-ibm	50,843
understanding-visualization-data	50,793
object-oriented-design	49,260
anatomy403-1x	47,177
stanford-statistics	46,997
solar-energy-basics	42,731
computational-thinking-problem-solving	42,612
womens-health-human-rights	41,149
fundamentals-of-reinforcement-learning	41,143
sas-programming-basics	40,832
information-security-data	39,640
introduction-to-cloud	38,783
blockchain-foundations-and-use-cases	37,013
matrix-algebra-engineers	36,865
tcpip	35,699
cybersecurity-roles-processes-operating-system-security	35,443
aws-cloud-technical-essentials	35,087
foundations-of-mindfulness	34,664
software-processes	33,995

STEM MOOC URL	Active Enrolled Learners
django-database-web-apps	33,983
python-project-for-data-science	31,336
building-modern-python-applications-on-aws	27,733
introduction-to-computers-and-office-productivity-software	27,395
introduction-to-machine-learning-in-production	26,530
introduction-software-testing	25,106
data-visualization-dashboards-excel-cognos	21,804
information-systems-audit	18,498
microsoft-azure-cloud-services	18,162
ibm-exploratory-data-analysis-for-machine-learning	15,404
introduction-to-data-engineering	14,743
ethics-technology-engineering	12,543
python-programming-intro	12,227
javascript-basics	11,677
java-introduction	11,325
introduction-to-web-development-with-html-css-javacript	8,839
cybersecurity-for-everyone	4,857
machine-learning-aws-nvidia	3,966
agile-development-and-scrum	3,456

Appendix G

The “Education Research” Section of Coursera’s Terms

> ## Education Research
>
> Coursera is committed to advancing the science of learning and teaching, and records of your participation in courses may be used for education research. In the interest of this research, you may be exposed to variations in the Content Offerings. Research findings will typically be reported at the aggregate level. Your personal identity will not be publicly disclosed in any research findings without your express consent.

Note. The full Terms & Conditions for Coursera can be viewed at www.coursera.org/about/terms

Appendix H

Helpfulness Rating Across Learners for Each Treatment by Message

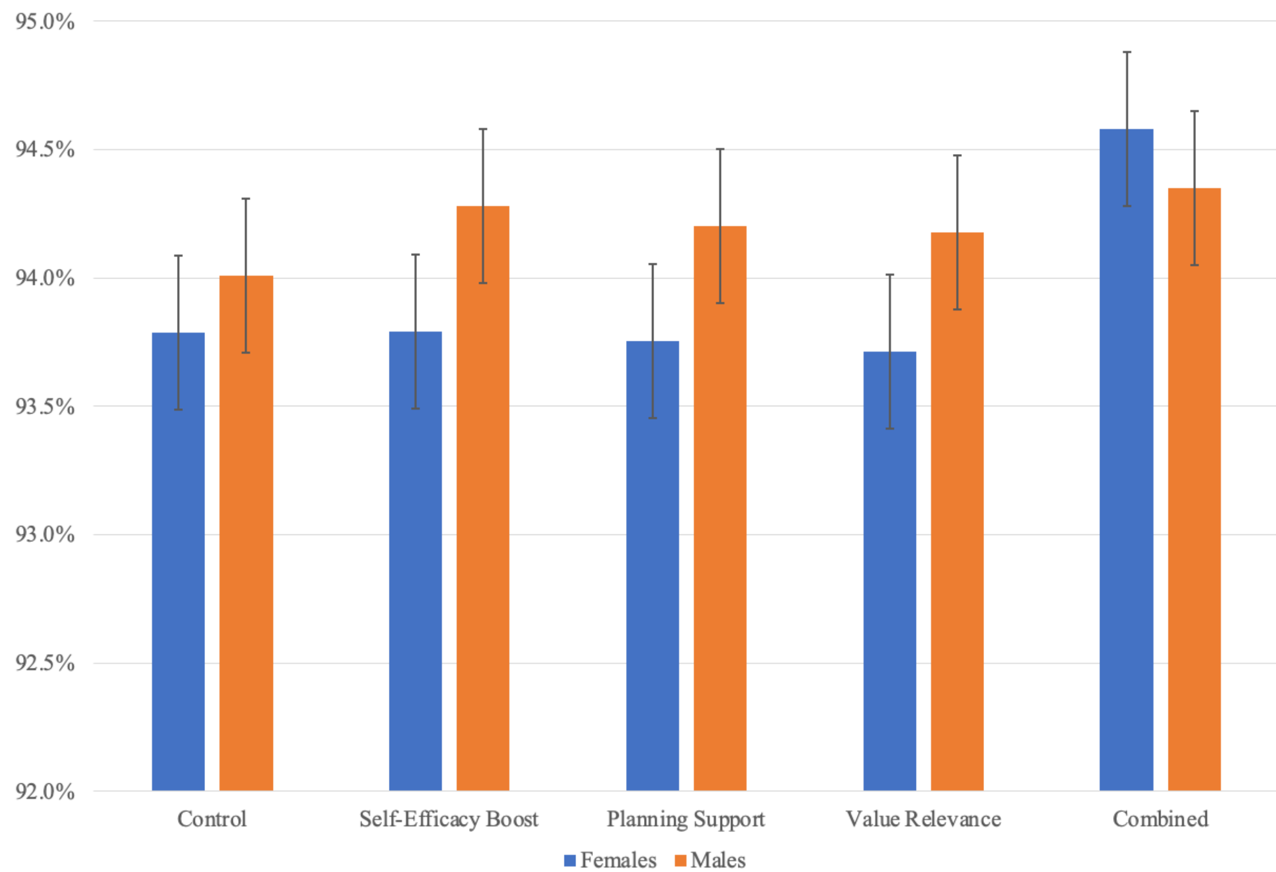
Treatment	Message Title	Reported Helpful	Reported Not Helpful	% Helpful	Avg. % Helpful by Treatment
Self-Efficacy Boost	Celebrate this victory	2,730	244	91.8%	91.2%
	Changes in your brain	2,409	197	92.4%	
	Nicely done!	4,462	303	93.6%	
	On your way!	28,598	2,854	90.9%	
	Take your time!	566	96	85.5%	
	Way to keep trying!	1,361	136	90.9%	
	You can do it	744	107	87.4%	
Planning Support	Celebrate your progress	3,154	317	90.9%	86.5%
	Enrolling is the first step!	22,685	3,761	85.8%	
	Make a plan to stay on track!	3,704	764	82.9%	
	When will you learn next?	3,263	299	91.6%	
Value Relevance	This learning is for you	7,155	479	93.7%	88.5%
	What are your goals?	2,803	399	87.5%	
	What is most important to you?	20,374	3,146	86.6%	
	Your learning is working!	2,757	258	91.4%	
Combined	Changes in your brain	1,896	179	91.4%	86.9%
	Make a plan to stay on track!	3,621	674	84.3%	
	Nicely done!	3,401	283	92.3%	
	Take your time!	533	102	83.9%	
	Way to keep trying!	1,148	140	89.1%	
	What are your goals?	2,243	456	83.1%	
	What is most important to you?	20,321	3,203	86.4%	
	When will you learn next?	2,362	312	88.3%	

Helpfulness Rating Across Female Learners for Each Treatment by Message

Treatment	Message Title	Reported Helpful	Reported Not Helpful	% Helpful	Avg. % Helpful by Treatment
Self-Efficacy Boost	Celebrate this victory	602	39	93.9%	91.7%
	Changes in your brain	506	42	92.3%	
	Nicely done!	920	71	92.8%	
	On your way!	6174	581	91.4%	
	Take your time!	104	9	92.0%	
	Way to keep trying!	309	32	90.6%	
	You can do it	142	20	87.7%	
Planning Support	Celebrate your progress	630	59	91.4%	86.7%
	Enrolling is the first step!	4593	745	86.0%	
	Make a plan to stay on track!	755	145	83.9%	
	When will you learn next?	676	69	90.7%	
Value Relevance	This learning is for you	1535	95	94.2%	89.4%
	What are your goals?	605	77	88.7%	
	What is most important to you?	4410	636	87.4%	
	Your learning is working!	567	39	93.6%	
Combined	Changes in your brain	402	43	90.3%	86.6%
	Make a plan to stay on track!	718	130	84.7%	
	Nicely done!	721	63	92.0%	
	Take your time!	109	16	87.2%	
	Way to keep trying!	243	29	89.3%	
	What are your goals?	399	97	80.4%	
	What is most important to you?	4374	706	86.1%	
	When will you learn next?	390	53	88.0%	

Appendix I

Average Final Course Grades for Future STEM Enrollments Beyond the Experiment



Note. Female learners, n = 15,394; male learners, n = 34,301

Curriculum Vitae

ALEXANDRA DOWNS URBAN

EDUCATION

Johns Hopkins University, 2019–2022 (anticipated graduation) *Online Degree Program*
Doctor of Education in Mind, Brain, and Teaching.

Harvard Graduate School of Education, 2015–2016 *Cambridge, Massachusetts*
Master's in Education in Mind, Brain, and Education, emphasis in Data Science and Statistics.

Brown University, 2011–2015 *Providence, Rhode Island*
Bachelor of Science in Educational Neuroscience, *magna cum laude,* Phi Beta Kappa.

Oxford Advanced Studies Program, Fall 2010 *Oxford, United Kingdom*
Computer Programming and Shakespeare; Oxfam Bookshop volunteer, assistant store manager.

Concord Academy, 2006–2010 *Concord, Massachusetts*
Senior Class Representative, Varsity Ski Team Captain, pioneered STEM peer tutoring program.

WORK EXPERIENCE

Summer 2015 – Present *Mountain View, California (& Remotely)*
Coursera, *Principal Learning Designer*, empower instructors in online degree creation, conduct data analysis to evaluate success, and align platform with research-backed pedagogy.

Summer 2014 *Boston, Massachusetts*
Symphony Learning, *Cognitive Development Researcher,* cataloged neuroscience used in their educational software, statistically assessed student progress, and guided teacher implementation.

Fall 2013 – Spring 2015 *Providence, Rhode Island*
Brown University, *Independent Concentration Coordinator,* on IC approval committee, head of IC Department Undergraduate Group, peer advisor in the Curricular Resource Center.

Summer 2012 *Woburn, Massachusetts*
"Wicked Cool For Kids" *Instructor,* facilitated STEM enrichment program for 1st to 4th graders.

Spring 2011 *Shanghai, China*
Concordia International High School, *Mathematics Teacher and Tutor,* classroom instructor for Algebra, Calculus, and Advanced Statistics, plus after-school tutor for STEM AP exams.

FELLOWSHIPS, AWARDS & BOARDS

Summer 2022: **National Science Foundation Award**, to participate in the International Mind, Brain, and Education Society annual conference in Montreal, Canada.

2019–2021: **Supervisory Board of Digital Learning Lab**, at the Higher School of Economics.

Spring 2016: **Harvard's Mind, Brain, Behavior Award**, to present research at interdisciplinary conference in Washington, DC.

Spring 2015: **Harvey A. Baker Fellowship**, to fund graduate study, awarded for high scholastic achievement, meaningful participation in college activities, and demonstrated student leadership.

Spring 2015: **The Jin Prize**, awarded for outstanding academic performance and commitment to public service.

Summer 2014: **Messing Family LINK Award**, to fund the summer internship at Symphony Learning.

Spring 2014: **The Nora Kahn Piore Award**, to support Brown School of Public Health primary research on support needed for low-income households in Rhode Island.

Fall 2013: **TRI-Lab Fellowship**, to sponsor collaborative research for healthy early childhood development in Rhode Island.

Summer 2013: **Brown International Scholars Program**, to conduct honors thesis primary research in New Zealand secondary school classrooms.

ADDITIONAL ACTIVITIES

- Professional Photographer
- Downhill Ski Racer
- Clean Water Activist
- Contributing Travel Writer
- Design Thinking Facilitator