How do Gender, Learning Goals, and Forum Participation Predict Persistence in a Computer Science MOOC?

R. WES CRUES, GENEVIEVE M. HENRICKS, MICHELLE PERRY, SUMA BHAT, CAROLYN J. ANDERSON, NAJMUDDIN SHAIK, and LAWRENCE ANGRAVE,
University of Illinois at Urbana-Champaign, USA

18

Massive Open Online Courses (MOOCs)—in part, because of their free, flexible, and relatively anonymous nature—may provide a means for helping overcome the large gender gap in Computer Science (CS). This study examines why women and men chose to enroll in a CS MOOC and how this is related to successful behavior in the course by (a) using k-means clustering to explore the reasons why women and men enrolled in this MOOC and then (b) analyzing if these reasons are related to forum participation and, ultimately, persistence in the course. Findings suggest that women and men have different reasons for taking this CS MOOC, and they persist at different rates, an outcome that is moderated by forum participation.

CCS Concepts: • **Social and professional topics** → **Adult education**; *Computer science education*; *Informal education*;

Additional Key Words and Phrases: Android app development, gender gap, forum participation, learning goals

ACM Reference format:
R. Wes Crues, Genevieve M. Henricks, Michelle Perry, Suma Bhat, Carolyn J. Anderson, Najmuddin Shaik, and Lawrence Angrave. 2018. How do Gender, Learning Goals, and Forum Participation Predict Persistence in a Computer Science MOOC? *ACM Trans. Comput. Educ.* 18, 4, Article 18 (September 2018), 14 pages.
https://doi.org/10.1145/3152892

1 INTRODUCTION

1.1 Women in Computer Science

Many fields within Science, Technology, Engineering, and Math (STEM) suffer from a large gender gap in which women are extremely underrepresented, and Computer Science (CS) in particular has a significantly low number of women earning degrees compared to men. The National Center for Education Statistics reports that women obtained only 18% of degrees conferred in the field during the 2013–2014 school year, and the trend is only worsening with a 19% decrease since 1983–1984 [38].

This work was funded in part by a grant to the third author from the Illinois Learning Science Design Initiative (ILSDI).
Authors' addresses: R. W. Crues, G. M. Henricks, M. Perry, and C. J. Anderson, 1310 South 6th Street, Champaign, IL, 61820; emails: {crues2, henrick2, mperry, cja}@illinois.edu; S. Bhat, 1308 W. Main Street, Urbana, IL 61801; email: spbhat2@illinois.edu; N. Shaik, Center for Innovation in Teaching & Learning (CITL), 901 W. University Ave., Suite 201, Urbana, IL 61802-2777; email: shaik@illinois.edu; L. Angrave, Siebel Center for Computer Science, 201 North Goodwin, Urbana, IL, 61801; email: angrave@illinois.edu.

https://doi.org/10.1145/3152892

The attrition rate for women in fields like CS is attributed to a number of reasons [9, 23]. Some researchers have examined persistence in the major in terms of having prior computer programming experience and claim that the relative lack of opportunities for women bodes poorly for their continuation in CS [4, 12]. Other explanations for the attrition rate point to women assessing themselves lower in their STEM abilities than men and having higher expectations for success than their male counterparts [7, 12]. Many studies argue about the influence of stereotype threat [5, 12, 36], in which a person fears confirming a negative stereotype about a group to which she belongs [37]; in this case, the person refers to a woman who fears confirming the stereotype that women are not good computer scientists. When confronted with the stereotype, anxiety will subdue the woman's working memory, which leads to decrements in performance, thereby confirming the stereotype [6]. When the discrimination is specifically addressed, however, women have shown higher rates of success in their STEM courses (see, e.g., Reference [35]). Still, others hypothesize that fields like CS do not come across as encapsulating the values and interests that women have, namely that of contributing to a social purpose [18]. This may explain why the gender gap is much smaller—if not reversed—in STEM majors focused on life sciences rather than physical sciences, within which CS falls [31]. The perception is that CS is antisocial and lacks human interaction, and, for many women, this is also the reality of their coursework experience [7, 8, 10]. A non-inclusive culture in STEM classes can lead to a feeling of social isolation [30], the sense of a "chilly climate" [20], and a lack of social belonging [21], which ultimately contributes to lack of persistence in the course. When women have access to role models and peers with whom they can identify, they tend to stay within STEM majors (see, e.g., Reference [41]).

We also note that all of these possible causes of women not persisting or not even enrolling in CS courses can be exacerbated when the classroom environment confirms these stereotypes and biases: when women find themselves the victim of doubt and discrimination, it takes extra determination to persevere. In a recent *Washington Post* article, Anderson [3] wrote "female students continue to face obstacles, including biases from classmates, teachers, and others who might cause them to doubt their potential." Although these deterrents were often blatant and deliberate in the past, currently these deterrents may be more insidious, because they are less obvious—even as seemingly innocuous as posters that portray stereotypical computer science themes [11]. Anderson went on to quote Harvey Mudd College's president and a computer scientist, Maria Klawe: "There is systematic exclusion. It used to be very deliberate, very conscious. I think it's now much less conscious." The problem that women face—of discrimination from peers and professors in their classrooms—becomes even more difficult to overcome as the discrimination becomes subtle and unconscious. With the discrimination going underground, women can be left feeling even more helpless than when the discrimination was obvious, because recognizing the issue and formulating a response requires energy and skills that may be difficult to organize and implement.

1.2 Computer Science MOOCs

Overcoming Bias. CS massive open online courses (MOOCs) are vastly popular, attracting hundreds of thousands of students, and are unlike any educational phenomena to date. They promise to democratize education by allowing people from around the globe to participate in courses taught for free from renowned professors at esteemed institutions. Moreover, MOOCs offer a potential means of overcoming some of the barriers that women face in CS. For example, they supply a free, low-stakes method for women to obtain experience and background in the field prior to taking coursework. Additionally, and pertinent to the point we just made about how hard it is for women to combat the increasingly subtle and unconscious discrimination in the CS classroom, the relatively anonymous and faceless MOOC learning space allows features of gender to remain largely unknown. This frees these traditionally underrepresented students from expectations and anticipated prejudices, minimizing the possibility of stereotype threat from occurring. Because MOOCs can reduce barriers and prevent prejudices, they can potentially allow women to form quality interactions with others, which could be difficult or even impossible in face-to-face settings. In MOOCs, this happens via the discussion forums.

Promoting Best Outcomes. Research in online learning and distance education has shown that engaging with asynchronous discussion forums leads to better academic outcomes, including longer course persistence [26], better grades [39], and enhanced critical thinking skills and knowledge construction [33, 34]. In part, these benefits can be attributed to the fact that these discussions do not occur in real time; this feature allows students to revisit and reconnect with the material at anytime during the course [16]. Moreover, students perceive that their participation in the discussion forums is helping them learn [40]. Within the discussion forums, prior research has shown that males and females participate at different rates when posting to the forums, where males post more than females [25]. Positive student outcomes are based on the instructor using best practices when implementing discussion forums as course features, including having relevant and clear goals for the forums, allowing students to share opinions and perspectives, balancing their presence as the instructor in the forums, and providing feedback [14].

Furthermore, certain characteristics of students have been linked with the skillsets that enable them to perform well in MOOCs. Specifically, Hood et al. [24] found that a MOOC student's role as either an employed data professional or a student obtaining a higher education degree predicted having better self-regulatory skills than MOOC students who were not data professionals or not obtaining a higher education degree. Robinson et al. [32] found certain terms could serve as representatives of intrinsic-value-seeking behaviors and thus led to persisting in the course, while Crossley et al. [13] found more linguistically complex forum posts are predictive of completing a MOOC. Although these results begin to provide a more fine-grained look at who is enrolled in MOOCs, it is still unclear why participants are interested in taking MOOCs. If instructors had knowledge of these various reasons, then they likely would be better equipped to help students successfully meet their specific goals.

1.3 Focal Issues for This Investigation

Although participation in the forums is touted as beneficial for all students (see, e.g., Reference [26]), we do not know whether this holds true, independent of why students enroll in the course. For example, if a student wants to get specific information about a particular CS topic, watching the course video may be sufficient and participation in the forum may be unrelated to that student's goal; if, however, someone is exploring CS as a potential major, it may be worthwhile to join in the forums, to get a sense of how others in the field ask about and respond to crucial CS issues. In addition, we do not know if men and women would enroll in a CS course for the same or for different reasons, and whether the relation between reason for enrolling and participation looks the same for men and women. Thus, one set of questions we pursue in this investigation is: Do men and women choose to enroll in a CS MOOC for similar or for different reasons? And, do the reasons students enroll in a CS MOOC relate to how they participate in the forums?

Although increased participation in forums has been linked to course retention in MOOCs and although it is clear that different course-long engagement patterns predict differential rates of forum participation [27], we do not know if men and women participate at the same rate. Thus, a second set of questions we pursue in this investigation is: Do men and women participate differentially in the forums? And does forum participation relate to course persistence? Does course persistence differ for men and women?

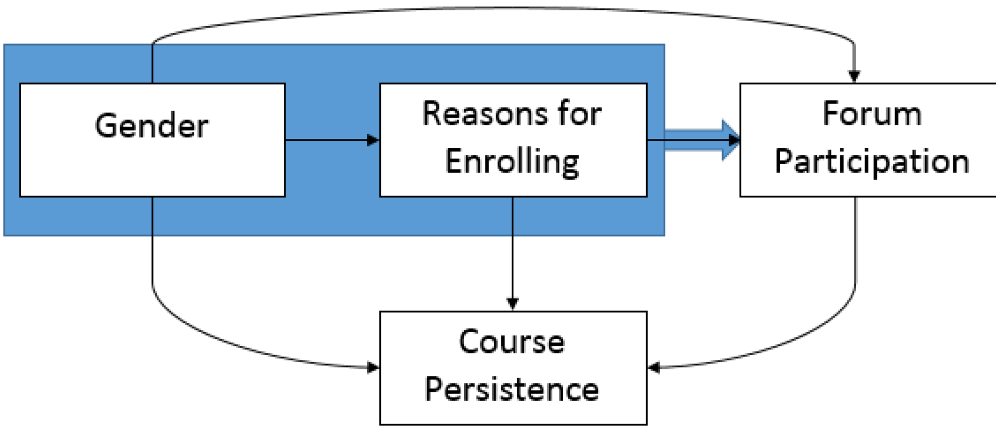

Fig. 1. The hypothesized relationship for this investigation.

In general, this investigation seeks to understand (a) if reasons for enrolling in a CS MOOC predict different types of participation in forums for both men and women and (b) whether type of reasons for enrolling and type of forum participation work together to predict course outcomes. This is displayed as an infographic in Figure 1.

2 METHOD

2.1 Description of Sample & Data Collection Methods

The data analyzed in this study include (1) a survey collected at the beginning of the Android MOOC course, Creative, Serious, and Playful Science of Android Apps, and (2) records of student interaction with the course, including visiting or posting to the forum.

2.1.1 The Course and Forum Participation in the Course. The course used a week-based activity schedule, during a 12-week time period. This course introduced fundamental computer science principles that power apps. Students learned to create their own Android apps using Java and standard software development tools. This course was designed to be a novice-friendly introduction to computer science and programming Android apps for smart-phones and tablets. No prior programming knowledge was assumed by the instructor. Students were aware that they would use the programming tools that Android software developers use and build a complete and useful app during the course. Along the way, the instructor introduced fundamental computer science principles and programming ideas that power smart-phone and tablet apps.

The course intentionally encouraged students to watch the instructor create or fix Android apps of increasing complexity and then reproduce the same steps. Short quizzes reinforced simple factual knowledge and know-how that were directly related to the lecture content. Longer assignments were used to support students' application of these ideas.

App design, development, and debugging are complex and technically demanding and challenging tasks, with many technical pitfalls that would normally be covered in "office hours" or as part of a lab in a traditional on-campus course. In a MOOC, in the absence of being able to call or email the instructor, learners largely had to rely on each other for help, and the forums were often full of activity associated with the assignments. As such, forums were a critical part in helping remote students succeed in the course. Forums were designed to help novices overcome these challenges, and to provide opportunity to help with missing skills or more advanced ideas that were not

Table 1. Number of Men and Women Grouped by Non-U.S. Citizen and U.S. Citizen

	Citizenship		
Gender	non-U.S. Citizen	U.S. Citizen	Unknown
Women	888 (66.023%)	454 (33.75%)	2 (0.2%)
Men	3,398 (76.14%)	1,062 (23.8%)	3 (0.06%)

Note: Row percentages sum to 100%.

Table 2. Number of Men and Women by Age Groups

	Age Groups							
Gender	≤17	18–24	25–29	30–39	40–49	50–59	≥60	Unknown
Women	34 (2.53%)	267 (19.85%)	323 (24.01%)	335 (24.91%)	203 (15.09%)	122 (9.07%)	45 (3.35%)	16 (1.19%)
Men	115 (2.58%)	988 (22.14%)	918 (20.57%)	1,209 (27.1%)	662 (14.84%)	350 (7.84%)	193 (4.33%)	27 (0.6%)

Note: Row percentages sum to 100%.

covered in the lecture videos. The intent was that the forums would foster community, offer the chance to obtain clarification, and, especially for advanced students, to share ideas.

Forum participation was never graded nor required; however, students were strongly encouraged to use the forums by several methods:

- The course included clickable hyperlinks to topic-based forums in the weekly list of activities.
- A suggested discussion area was connected to each video.
- A discussion area was connected to each assignment.
- Weekly activities suggested that students use the forums as part of their path through the course.
- Weekly summary emails were sent to students that discussed upcoming assignments, near future deadlines, and also mentioned forums and highlighted some of the lively topics and discussions.
- The home page of the course included links to the forum areas including specific-topics and general MOOC help forums.

2.1.2 The Beginning-of-the-Course Survey. Participants were asked demographic questions and their expectations for the course, among other issues (which were not examined in this investigation). The survey was optional, where n = 5,807 participants who provided information about their gender, listed a reason for taking the course, and had data available about forum participation; before removing those who did not have this information, the number of participants was N = 146,970. There are no known data quality problems; however, we did exclude those who had missing gender, reasons for taking the course, or forum participation data for our analysis, thus arriving at 5,807 participants. These 5,807 participants were located across the globe and spanned a wide swath of ages. Table 1 gives the distribution of the 5,807 participants in terms of gender and U.S. citizenship, while Table 2 gives the distribution of the participants in terms of gender and age groups.

2.2 Clustering Reasons for Enrolling

As part of the survey, participants were asked "Why are you taking this course? What do you hope to get out of it?" To cluster the text into mutually exclusive categories for answers to this inquiry, we used k−means clustering to group students' reasons into relevant categories. Using this approach is in direct response to Dringus and Ellis' [15] suggestion to use data mining as a technique for discovering more nuanced patterns to discussion forums.

To prepare the text for analyses, we undertook several pre-processing tasks using the "tm" package in R [19]. Note that for these preliminary analyses, we used unigrams and a bag-of-words approach to glean insight into these texts. Before analysis, we removed punctuation and numerals, and changed the case of all words to lowercase. Furthermore, we removed a list of English stop words provided with the "tm" package. These are non-content words, such as *and* and *the*. It is noted however, that not all responses were in English, but because the majority were written in English, we only removed English stop words.

We made the decision to only use the stems of words, which we implemented in the "tm" package by way of the Porter stemming algorithm. Excess white space was removed and documents (in this case, responses to a query about reasons for taking the course) were converted to plain-text documents.

To select the final terms to be used in $k-$means clustering, we employed Luhn's theory of establishing relevant sentences, but with only the unigrams, i.e., individual words [29]. That is, we removed very frequent and very infrequent terms from the analysis. Once all of these pre-processing steps and selection strategies were utilized, 139 terms were used to group responses into clusters. Of the 139 terms, examples include *career*, *code*, *comput*, *creativ*, *curious*, *degre*, *design*, *educ*, *expand*, *hobbi*, *opportun*, *play*, *softwar*, *student*, and *write*, among others.

2.3 Description of Forum Participation Variable

Given that one of our main interests was the way students interacted with the forum, we derived a variable to indicate each student's behavior for the forum. We considered three types of forum participation, ordinally: no views or posts in the forums, viewing the forum (but not posting to the forum), and viewing and posting in the forum. For the participants in the course, we had n_{none} = 1,693 who neither viewed nor posted to the forum during the course; n_{view} = 2,158 who viewed the forum at least once during the course; and, finally, n_{post} = 1,956 who posted at least once in the forum during the course.

2.4 Description of Persistence Variable

To estimate our models and determine what would be predictive of our outcome—students persisting in the course—we considered only those students who signed up for the course before or during the very first week.

We considered any activity during a week to be indicative of being active in the course. This activity could be viewing the course lecture, viewing the forums, posting to the forum, viewing the Wikis, visiting course webpages, or taking a quiz, all of which indicated some active engagement in the course. We then considered three levels of persistence: active sometime during course weeks 0 (i.e., before the course officially started), 1, or 2 (but not beyond), which we call low persistence; active early in the course (i.e., during weeks 0, 1, or 2) **and** sometime during course weeks 3, 4, or 5, which we call medium persistence; and finally, active early and midway through the course **and** sometime during course weeks 6, 7, or 8, which we call high persistence. Of the participants we analyzed, n_{low} = 2138 had low persistence, n_{medium} = 1717 had medium persistence, and n_{high} = 1952 had high persistence.

3 RESULTS

First, we discuss the results of the clustering of reasons for enrolling, then we report the results stemming from the analytic model, attempting to predict course persistence, the outcome of interest, and whether any of these results show differential patterns for men versus women.

3.1 Clusters of Reasons for Enrolling

We used the text that students provided for open-ended responses and used k-means clustering, where the number of centroids was five, as this solution was the most interpretable after experimenting with the number of centroids. One cluster was the largest and represents students with multiple or unique reasons for taking the course, or those who did not present reasons that fit into any of the remaining four clusters. The remaining four clusters of reasons were validated both logically (i.e., when examining the complete corpus, the verbatim reasons could be labeled distinctly) and empirically (i.e., students who fell into one cluster behaved differently than students who fell into other clusters). Although we could have employed other methods for coding students' reasons for taking the course, the method we chose was both efficient (relative to hand coding and obtaining inter-rater reliability) and useful.

The clusters are mutually exclusive; therefore, each student belongs to only one of the five clusters. We now turn towards our interpretation of the clusters and provide samples of student responses.

Computer Science Student Cluster. We identified a group of students in the course who were interested in being students of computer science, were formerly computer science students, or had a more clearly defined interest in the academic field of computer science. Here is an example of a student response:

> "I am preparing myself to go back to a school for a Master's degree in Software Engineering so I would like to refresh my information in the field of Computer Science (in general) before getting back to the school."

Here is another example, where the student wrote:

> "Hope to take this course in my spare time (now that I currently own a smartphone). Hope to learn more about Android development in the field of computer science, which was my college major."

Understanding and Learning Cluster. Some participants in this MOOC identified an interest in creating an understanding of Android app development and an understanding the machinery behind applications they use. An example indicating this interest is:

> "I have been curious about the apps for tablets & phones. I think this class will give me a start on learning and understanding them."

Another example of a student who was interested in understanding Android apps:

> "Looking to gain a better understanding to what it takes to build an android app."

Programming is Cool Cluster. Some participants identified reasons that indicated their general fascination with programming. Here is an example of a student's response that fell in this cluster:

> "I love programming, and I think it will be always useful to know how to create Apps for a variety of purposes."

Here is another example of a response from a student in this cluster:

> "I'm curious to know how mobile apps are developed on Android. I'll also like to be able to develop one."

Table 3. Number of Men and Women in Each of the Five Empirically Derived Clusters of Reasons for Enrolling in the Course

	Cluster				
Gender	"Other"	Student	Understand	Cool	Career
Women	987 (73.38%)	71 (5.28%)	92 (6.84%)	88 (6.54%)	107 (7.96%)
Men	3,463 (77.61%)	203 (4.55%)	258 (5.78%)	256 (5.74%)	282 (6.32%)

Note: Row percentages sum to 100%.

Career and Entrepreneurial Activities Cluster. Finally, a portion of the participants in this MOOC identified being interested in career development or using app development in an entrepreneurial way. Here is an example of a student who identified career interests:

> "So that I can write something on my resume in future when I am required to get a job."

Another example, this one with a more entrepreneurial spirit, is presented below.

> "I want to be able to develop an app/game that I can share with friends and family and/or eventually sell on the marketplace as an indie game."

"Other" Cluster. The final cluster contained the vast majority of students enrolled in the course. An example of a response that did not fit into the other four clusters.

> "I'm hoping to see if it's a field I want to go into. I like Android, but not sure if I'd enjoy making apps for it. I love science. I thought it might be a fit."

Another example of a response in the other cluster is below. Note that here, a very specific reason is given.

> "I want to make and give an app customized for my kids."

3.2 Gender's Relation with Reasons for Enrolling

We began by looking for men's and women's stated reasons for enrolling in this CS MOOC, to see whether gender predicted reasons for enrolling. The vast majority of both men and women fell into the catch-all "other" cluster, but using a χ^2 test of independence, we determined that gender and reasons for enrolling in the course shared a dependent relationship, χ^2 (4) = 10.75, $p < 0.03$. That is, we found a significant association between gender and reasons for enrolling. We note that although we uncovered a significant statistical relationship, the patterns found for men and for women were not remarkably different, as seen in Table 3. Further, when we removed the "other" column from the table, the dependent relationship was no longer present, where χ^2 (3) = 0.4, $p > 0.9$.

To understand the relationships in Table 3, we analyzed the standardized Pearson residuals. Upon computing these residuals, we found the observed frequencies for women and men in the "other" cluster were quite different than their expected frequencies, where the standardized residuals were 3.2 for women and −3.2 for men, suggesting lack-of-fit under the hypothesis assumed when conducting the χ^2 test of independence with all five clusters [1]. This further supports our contention to remove the "other" category and analyze Table 3 without this category.

3.3 Gender's Relationship with Participating in the Forums

Next, we examined the relationship between gender and participation in the forums. Using a χ^2 test of independence, we determined that gender and forum participation shared a dependent

Table 4. Number of Men and Women at Each of the Three Levels of Participation in the Forums

	Level of Participation in the Forums		
Gender	No views (no visits)	Views (no posts)	Posts
Women	395 (29.37%)	435 (32.34%)	515 (38.29%)
Men	1,298 (29.1%)	1,723 (38.62%)	1,441 (32.29%)

Note: Row percentages sum to 100%.

relationship, χ^2 (2) = 22.0, $p < 0.0001$. In general, the data indicate that the proportion of women who post compared to those who only view the forum, and the proportion who view compared to those who do not view the forum, was different for men versus women. Namely, the modal response for men was different: men had a higher proportion of viewing forum posts than women, and women had a higher proportion of posting to the forum or not viewing the forum, as seen in Table 4.

For a more thorough analysis, we again computed Pearson standardized residuals for Table 4. We found the residuals were greater than four for men and women in both the views and posts levels of participation. That is, the number of women and men we expected to view or post in the forum when using the χ^2 test of independence was quite different from the number of men and women we observed who viewed or posted in the forum [1]. Specifically, for women, we observed more women viewing the forum than we expected, but less women posted in the forum. The reverse was true for men: More men than expected posted to the forum, but less men than expected viewed the forum.

3.4 Reasons-for-Enrolling's Relation to Participating in the Forums

In our next step, we looked at the effects of reasons for enrolling to uncover how reasons for enrolling in the MOOC might predict forum participation. To do this, we began by estimating successive ordinal logistic regression models, because our response variable, forum participation, has a natural ordering. The final model included only two statistically significant predictors of forum participation. A likelihood ratio test comparing the model with no predictors and the two statistically significant predictors revealed a test statistic χ^2 (2) = 13.78, $p < 0.01$. This supports our contention of including these two predictors.

These reasons were *computer science student* and *career and entrepreneurial activities*. Holding all else constant, the odds that students (men and women, combined) who volunteered the reasons that fit in the *computer science student* cluster (that they were a *computer science student*) would post in the forum versus just view the forum (or just view the forum versus not ever visit the forum) was 1.41 times the odds of a student who listed another reason. Holding all else constant, the odds that students (men and women, combined) who volunteered the reason that fit in the cluster of *career and entrepreneurial activities* would post in the forum versus just view the forum (or just view the forum versus not ever visit the forum) was 1.26 times the odds of a student who listed another reason. Note that the proportional odds assumption was verified for this and for all ordinal logistic regression models presented.

3.5 Gender's and Reasons-for-Enrolling's Combined Relation to Predict Participating in the Forums

Because we found a dependent relationship between student gender and stated reasons for enrolling in this MOOC, we conducted ordinal logistic regression separately for men and women. Thus, in our next step, we looked at the relationship between gender and reasons for enrolling to

Table 5. Number of Men and Women at Each of the Three Levels of Persistence in This CS MOOC

	Persistence in the MOOC		
Gender	Low	Medium	High
Women	591 (43.94%)	380 (28.25%)	374 (27.81%)
Men	1,547 (34.67%)	1,337 (29.96%)	1,578 (35.37%)

Note: Row percentages sum to 100%.

uncover how gender and reasons for enrolling in the MOOC might work together to predict how students participate in the forum. Likelihood ratio tests comparing each of these models to a null model confirmed these models explain the data better than models with no predictors.

Women. We initially included all of the reasons-for-enrolling and obtained a final model to predict forum participation, however, the only statistically significant covariate to predict forum participation was when they indicated they took the course due to being a student and wanting to learn (i.e., being identified in the computer-science-student cluster). Namely, the odds that these women would post in the forum versus just view the forum (or just view the forum versus not every visit the forum) was 2.21 times the odds of a woman who listed another reason.

Men. We followed a similar procedure for men as we did for women. However, the only statistically significant covariate to predict forum participation was when men indicated they took the course due to *career aspirations*, i.e., were in the career-and-entrepreneurial-activities cluster. Namely, the odds for a man who indicated taking this course due to *career aspirations* would post in the forum versus just view the forum (or just view the forum versus not ever visit the forum) was 1.28 times the odds of a man who listed another reason.

3.6 Predicting Persistence

We initially included all covariates discussed (i.e., gender, reasons for enrolling, and forum participation) to predict persistence. The final model included gender and forum participation as statistically significant predictors of persistence. A likelihood ratio test comparing this model to a model with no predictors revealed the final model significantly explains the data better than a model with no predictors. Although reasons for enrolling were significantly related to both gender and forum participation, when including gender, reasons-for-enrolling, and forum participation, the reasons-for-enrolling were not statistically significant predictors of persistence. Thus, we describe results for both gender and forum participation as predictors of persistence.

Gender. To see the effects of gender on persistence, we held forum participation constant; the odds were 1.59 times the odds for a male than a female to have high persistence than medium persistence, and again, the odds were 1.59 times the odds for a male than a female to have medium persistence over low persistence. Table 5 shows persistence by gender.

Forum Participation. To see the effects of forum participation on persistence, we held gender constant; the odds were 5.90 times the odds for a person who posted to the forum than other people (who only viewed the forum or never went to the forum) to have high persistence than medium persistence, and again, the odds were 5.90 times the odds for a person who posted in the forum than other people to have medium persistence than low persistence. Furthermore, the odds were 2.43 times the odds for a person who viewed the forum versus others not viewing to have high versus medium persistence, and, likewise, a viewer of the forum had odds 2.43 times the odds than a non-viewer to have medium versus low persistence.

When we included gender along with forum participation in the model to predict persistence, we found gender to have a statistically significant effect. Because we found a difference between men and women, we conducted separate analyses for males and females. Likelihood ratio tests confirmed these models explain the data better than models with no predictors.

Forum Participation for Women. Using ordinal logistic regression, we found that the odds were 4.64 times more for a woman who posted than a woman who did not post in the forum to have high persistence than medium persistence, and, likewise, the odds were 4.64 times more for a woman who only viewed but did not post to have medium versus low persistence. Furthermore, a woman who viewed the forum had odds 2.15 times more than a woman who did not view the forum to have high persistence versus medium persistence, and likewise, the odds were 2.15 times more for a woman who viewed the forum versus a woman who did not view the forum to have medium versus low persistence.

Forum Participation for Men. Again, using ordinal logistic regression, we found that the odds were 6.39 times more for a man who posted than a man who did not post to have high persistence than medium persistence, and likewise, the odds were 6.39 times more for a man who posted versus a man who did not post to have medium versus low persistence. Furthermore, we found that the odds were 2.53 times more for a man who viewed the forum than a man who did not view the forum to have high persistence than medium persistence, and again, the odds were 2.53 times more for a man who viewed the forum versus a man who did not view the forum to have medium versus low persistence.

4 DISCUSSION

4.1 Summary of Results

The major results of this investigation suggest that men and women behave differently in the CS MOOC we examined. Men and women participated in the forums differently, and this was related to their reasons for taking their course as well as their gender: the men who actually visited the forums were less likely to post than the women; moreover, men who offered that they were taking the course to fulfill career aspirations were more likely than other men to participate more substantially in the forums. Women who offered that they were taking the course because they were a CS student were more likely than other women to participate more substantially in the forums.

Finally, we found that men and women persisted differently, and this was related both to their gender and to their forum participation: men were more likely to persist longer in the course than women, and students who posted to the forum persisted longer in the course than students who did not post; these students who went to the forum and did not post persisted longer than students who never even visited the forum.

4.2 The Significance and Implications of Women's and Men's Different Behaviors for Supporting Student Success in CS MOOCs

In general, women participate less in the field of CS [2, 30] and less in CS MOOCs than men [22]. Moreover, for MOOCs, posting to the forum is related to persevering in the MOOC [27]. We found these results to be true of the students we examined in this investigation. However, to date, we know of no reports that examine the patterns of posting to CS MOOC forums by women versus by men. Given what has been reported, we might have expected that women would post proportionately less frequently than men. Instead, we found that women posted to the forum at higher proportions than men. One striking aspect of this finding is that gender roles in face-to-face courses were *not* replicated in this online, MOOC environment, namely women contributed in productive and visible ways in this CS MOOC, contrary to what Dubosson and Emad [17] reported (although they did look at gendered participation in the course forum, they only reported the posts of men compared to women, not the percentage of, e.g., women who posted).

Our findings that women posted to the forum at higher proportions than men and that women who posted to the forum had higher odds of persisting than women who did not post to the forum (which was true for men as well) provides potential for optimism. Further, because the models were fit using maximum likelihood estimation, we used a Wald test, and found the log odds were significantly different for women and men ($z = 2.1$, $p < 0.05$). Given this relatively widespread behavior among women who sign up for the course, women could be informed of this and could be encouraged (especially if they are a non-posting woman) to get involved in the course by posting to the forum. Perhaps all students—with special attention to target the women in the course—could be reminded that, as Kop et al. [28] noted, "active participation through the production of digital artifacts [leads] to positive learning outcomes as it helped participants to reflect and involved them in a creative process that stimulated their cognitive processes." We think that letting all students in on the fact that posting leads to positive outcomes—and why this is the case—could motivate students to post. Letting women know that this is not only possible, but typical of women in a CS MOOC, could inspire more women to post and ultimately succeed in new iterations of this and in other CS MOOCs. In future work, we hope to test the effect of providing this motivation and examine the effects of this on women's participation in CS MOOCs.

We reiterate that, for both men and women, posting to the forum led to more positive outcomes (i.e., persistence) than just viewing or than not even viewing the forum. Given that men were less inclined to post relative to women (who were inclined to post if they went to the forum), men, too, could be informed of the considerable benefits of posting to the forum.

However, if our goal is to hone in on how to support women in CS courses, we can learn something from the reasons women cited for taking this CS MOOC. Given that the most successful women in this course cited reasons relating to being a student and wanting to learn computer science, we suggest that interventions be designed and tested to encourage women—especially women already enrolled as students—to embrace being a student and wanting to learn, as this may validate their purpose and keep them focused on what they can control in the near term (success in the course) rather than what may be harder to attain (e.g., eventual success in a career).

4.3 Limitations

Like any investigation, this one has a set of limitations. Although we examined a large number of students, we only analyzed data from students who answered relevant survey questions. Students who chose to take the survey may be qualitatively different than those who did not take the survey, which leaves the possibility that other students might demonstrate different results. Additionally, we only studied one course, which limits generalizability. Conducting a future investigation across a range of courses and instructors would allow for a better understanding of how pervasive these results are. A future investigation may also consider using hand-coding to reexamine the "other" category as a reason for enrolling in this MOOC subcategories may emerge when using an alternative to text mining. Finally, although this investigation focused on persistence as an outcome, future investigations could analyze how different reasons for taking the course and forum participation affect other outcomes, such as quiz scores, self-perception of ability, and continuation into other computer science courses.

5 CONCLUSIONS

We believe it is worthwhile to make CS MOOCs a place where women can engage, persist, and learn, although we recognize that CS and MOOCs have been places where it is more commonplace for men to succeed. By understanding how women's behavior with CS MOOCs differ from men's, we can play to these strengths (e.g., encouraging and facilitating posting to forums) and mitigate barriers (e.g., inviting women to embrace being a student rather than focusing on how programming might be cool or attain success in a computer science career). Finding ways in which all students can succeed in computer science has the potential to make the field not only a more diverse but also a richer enterprise.

ACKNOWLEDGMENTS

The authors thank the anonymous referees for their feedback and suggestions. All opinions expressed herein are those of the authors and not the Illinois Learning Science Design Initiative (ILSDI).

REFERENCES

[1] Alan Agresti. 2002. *Categorical Data Analysis* (2nd ed.). Wiley-Interscience.

[2] American Association of University Women. 2000. *Tech-savvy: Educating Girls in the New Computer Age.* Technical Report.

[3] Nick Anderson. 2016. Women break barriers in engineering and computer science at some top colleges. *Washington Post* (2016). Retrieved September 16, 2016 from https://www.washingtonpost.com/local/education/women-break-barriers-in-engineering-and-computer-science-at-some-top-colleges/2016/09/16/538027a4-7503-11e6-be4f-3f42f2e5a49e_story.html?noredirect=on&utm_term=.0105fa716298.

[4] Lecia J. Barker, Charlie McDowell, and Kimberly Kalahar. 2009. Exploring factors that influence computer science introductory course students to persist in the major. In *Proceedings of the 40th ACM Technical Symposium on Computer Science Education.* ACM, 153–157.

[5] Maya A. Beasley and Mary J. Fischer. 2012. Why they leave: The impact of stereotype threat on the attrition of women and minorities from science, math and engineering majors. *Soc. Psychol. Edu.* 15, 4 (2012), 427–448.

[6] Sian L. Beilock, Robert J. Rydell, and Allen R. McConnell. 2007. Stereotype threat and working memory: Mechanisms, alleviation, and spillover. *J. Exper. Psychol. Gen.* 136, 2 (2007), 256.

[7] Sylvia Beyer, Kristina Rynes, Julie Perrault, Kelly Hay, and Susan Haller. 2003. Gender differences in computer science students. In *Proceedings of the 34th SIGCSE Technical Symposium on Computer Science Education.* ACM, 49–53.

[8] Maureen Biggers, Anne Brauer, and Tuba Yilmaz. 2008. Student perceptions of computer science: A retention study comparing graduating seniors with CS leavers. In *Proceedings of the 39th SIGCSE Technical Symposium on Computer Science Education.* ACM, 402–406.

[9] Jacob Clark Blickenstaff. 2005. Women and science careers: Leaky pipeline or gender filter? *Gender Edu.* 17, 4 (2005), 369–386.

[10] Lori Carter. 2006. Why students with an apparent aptitude for computer science don't choose to major in computer science. *Proceedings of the 37th SIGCSE Technical Symposium on Computer Science Education.* 27–31.

[11] Sapna Cheryan, Victoria C. Plaut, Paul G. Davies, and Claude M. Steele. 2009. Ambient belonging: How stereotypical cues impact gender participation in computer science.*J. Personal. Soc. Psychol.* 97, 6 (2009), 1045.

[12] Shelley J. Correll. 2001. Gender and the career choice process: The role of biased self-assessments. *Amer. J. Sociol.* 106, 6 (2001), 1691–1730.

[13] Scott Crossley, Danielle S. McNamara, Ryan Baker, Yuan Wang, Luc Paquette, Tiffany Barnes, and Yoav Bergner. 2015. Language to completion: Success in an educational data mining massive open online class. In *Proceedings of the 8th International Conference on Educational Data Mining*. 388–391.

[14] Vanessa Paz Dennen. 2005. From message posting to learning dialogues: Factors affecting learner participation in asynchronous discussion. *Distance Edu.* 26, 1 (2005), 127–148.

[15] Laurie P. Dringus and Timothy Ellis. 2005. Using data mining as a strategy for assessing asynchronous discussion forums. *Comput. Edu.* 45, 1 (2005), 141–160.

[16] Laurie P. Dringus and Timothy Ellis. 2010. Temporal transitions in participation flow in an asynchronous discussion forum. *Comput. Edu.* 54, 2 (2010), 340–349.

[17] Magali Dubosson and Sabine Emad. 2015. The forum community, the connectivist element of an xMOOC. *Univ. J. Edu. Res.* 3, 10 (2015), 680–690.

[18] Jacquelynne S. Eccles. 1987. Gender roles and women's achievement-related decisions. *Psychol. Women Quart.* 11, 2 (1987), 135–172.

[19] Ingo Feinerer and Kurt Hornik. 2015. *Tm: Text Mining Package.* R package version 0.6-2.

[20] Faye Flam. 1991. Still a "chilly climate" for women? Women in astronomy and physics say they face not so much overt discrimination as a pattern of "micro-inequalities." The remedy: More women. *Science* 252, 5013 (1991), 1604–1607.

[21] Catherine Good, Aneeta Rattan, and Carol S. Dweck. 2012. Why do women opt out? Sense of belonging and women's representation in mathematics. *J. Personal. Soc. Psychol.* 102, 4 (2012), 700.

[22] Philip J. Guo and Katharina Reinecke. 2014. Demographic differences in how students navigate through MOOCs. In *Proceedings of the 1st ACM Conference on Learning@Scale Conference.* ACM, 21–30.

[23] Catherine Hill, Christianne Corbett, and Andresse St. Rose. 2010. *Why so few? Women in science, technology, engineering, and mathematics.* American Association of University Women, Washington, D.C.

[24] Nina Hood, Allison Littlejohn, and Colin Milligan. 2015. Context counts: How learners' contexts influence learning in a MOOC. *Comput. Edu.* 91 (2015), 83–91.

[25] Jonathan Huang, Anirban Dasgupta, Arpita Ghosh, Jane Manning, and Marc Sanders. 2014. Superposter behavior in MOOC forums. In *Proceedings of the 1st ACM Conference on Learning@Scale Conference.* ACM, 117–126.

[26] Suhang Jiang, Adrienne Williams, Katerina Schenke, Mark Warschauer, and Diane O'Dowd. 2014. Predicting MOOC performance with week 1 behavior. In *Proceedings of the 7th International Conference on Educational Data Mining (EDM'14).* 273–275.

[27] René F. Kizilcec, Chris Piech, and Emily Schneider. 2013. Deconstructing disengagement: Analyzing learner subpopulations in massive open online courses. In *Proceedings of the 3rd International Conference on Learning Analytics and Knowledge.* ACM, 170–179.

[28] Rita Kop, Hélène Fournier, and John Sui Fai Mak. 2011. A pedagogy of abundance or a pedagogy to support human beings? Participant support on massive open online courses. *Int. Rev. Res. Open Distrib. Learn.* 12, 7 (2011), 74–93.

[29] Hans Peter Luhn. 1958. The automatic creation of literature abstracts. *IBM J. Res. Dev.* 2, 2 (1958), 159–165.

[30] Jane Margolis and Allan Fisher. 2002. *Unlocking the Clubhouse: Women in Computing.* MIT Press, Cambridge, MA.

[31] Ben Ost. 2010. Differences in persistence patterns between life and physical science majors: The role of grades, peers, and preparation (Electronic version). Cornell University, School of Industrial and Labor Relations site. Retrieved July 30, 2018 from http://digitalcommons.ilr.cornell.edu/workingpapers/119/.

[32] Carly Robinson, Michael Yeomans, Justin Reich, Chris Hulleman, and Hunter Gehlbach. 2016. Forecasting student achievement in MOOCs with natural language processing. In *Proceedings of the 6th International Conference on Learning Analytics & Knowledge.* ACM, 383–387.

[33] Tammy Schellens and Martin Valcke. 2005. Collaborative learning in asynchronous discussion groups: What about the impact on cognitive processing? *Comput. Human Behav.* 21, 6 (2005), 957–975.

[34] Tammy Schellens and Martin Valcke. 2006. Fostering knowledge construction in university students through asynchronous discussion groups. *Comput. Edu.* 46, 4 (2006), 349–370.

[35] Emily S. Shaffer, David M. Marx, and Radmila Prislin. 2013. Mind the gap: Framing of women's success and representation in STEM affects women's math performance under threat. *Sex Roles* 68, 7–8 (2013), 454–463.

[36] Claude M. Steele. 1997. A threat in the air: How stereotypes shape intellectual identity and performance. *Amer. Psychol.* 52, 6 (1997), 613.

[37] Claude M. Steele and Joshua Aronson. 1995. Stereotype threat and the intellectual test performance of african americans. *J. Personal. Soc. Psychol.* 69, 5 (1995), 797.

[38] U.S. Department of Education. 2015. *"Degrees and Other Formal Awards Conferred" surveys, 1970-71 through 2013-14; Integrated Postsecondary Education Data System (IPEDS), "Completions Survey" (IPEDS-C:87-99); and IPEDS Fall 2000 through Fall 2014, Completions Component.* Technical Report.

[39] Eileen Webb, Alan Jones, Philip Barker, and Paul van Schaik. 2004. Using e-learning dialogues in higher education. *Innovat. Edu. Teach. Int.* 41, 1 (2004), 93–103.

[40] Dezhi Wu and Starr Roxanne Hiltz. 2004. Predicting learning from asynchronous online discussions. *J. Asynchr. Learn. Netw.* 8, 2 (2004), 139–152.

[41] Alicja Zawistowska. 2017. Is entering STEM socially contagious? Contextual factors in women's educational decisions. *Polish Sociol. Rev.* 197 (2017), 51.

Received September 2016; revised October 2017; accepted October 2017