Gender, Self-Assessment, and Persistence in Computing: How gender differences in self-assessed ability reduce women's persistence in computer science

Cynthia Hunt
chunt26@kent.edu
Department of Sociology, Kent State University
Kent, Ohio, USA

Spencer Yoder
smyoder@ncsu.edu
North Carolina State University
Raleigh, North Carolina, USA

Taylor Comment
tcomment@kent.edu
Department of Sociology, Kent State University
Kent, Ohio, USA

Thomas Price
twprice@ncsu.edu
North Carolina State University
Raleigh, North Carolina, USA

Bita Akram
bakram@ncsu.edu
North Carolina State University
Raleigh, North Carolina, USA

Lina Battestilli
lbattestilli@ncsu.edu
North Carolina State University
Raleigh, North Carolina, USA

Tiffany Barnes
tmbarnes@ncsu.edu
North Carolina State University
Raleigh, North Carolina, USA

Susan Fisk
sfisk@kent.edu
Department of Sociology, Kent State University
Kent, Ohio, USA

ABSTRACT

Are women less likely to persist in computer science because of gender differences in self-assessed computing ability? And why do gender differences exist in self-assessments among women and men who earn the same grades? We use a mixed-method research design to answer these questions, utilizing both quantitative survey data (n = 764) and qualitative interview data (n = 59) from students in introductory computing courses at a large U.S. state university. Quantitatively, we find that women self-assess their computing ability significantly lower than men who earn the same grades, and that these lower self-assessments reduce the likelihood that women enroll in future CS courses (relative to men who earn equivalent grades). Qualitatively, we explore how women and men perceive their own computing ability to understand why women self-assess their ability lower than men. Our interviews revealed that women were much less likely than men to make favorable comparative judgements about their ability relative to their classmates. Women also had higher personal performance standards than men. Lastly, women were more likely than men to experience disrespectful treatment, with an undertone of presumed incompetence, from their TAs and classmates. In sum, this research furthers our understanding of why gender differences exist in self-assessments of computing ability and how these differences can contribute to gender disparities in computing persistence. It also draws attention to the importance of feedback in computing courses and suggests that improving course feedback may reduce gender disparities in computing.

ICER 2022, August 7–11, 2022, Lugano and Virtual Event, Switzerland

ACM ISBN 978-1-4503-9194-8/22/08…$15.00
https://doi.org/10.1145/3501385.3543963

CCS CONCEPTS

• **Social and professional topics → Adult education**.

KEYWORDS

introductory course, students, self-assessments, ability, persistence, gender

ACM Reference Format:
Cynthia Hunt, Spencer Yoder, Taylor Comment, Thomas Price, Bita Akram, Lina Battestilli, Tiffany Barnes, and Susan Fisk. 2022. Gender, Self-Assessment, and Persistence in Computing: How gender differences in self-assessed ability reduce women's persistence in computer science. In *Proceedings of the 2022 ACM Conference on International Computing Education Research V.1 (ICER 2022), August 7–11, 2022, Lugano and Virtual Event, Switzerland*. ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3501385.3543963

1 INTRODUCTION

Women remain tremendously underrepresented in the field of computing. [52, 53]. While many factors contribute to this inequality, previous research suggests that gender differences in self-assessments of computing ability may contribute to the dearth of women in computing [38, 44]. This is because people need to believe that they have adequate ability to succeed in a field in order to pursue it [13], and research has found that women are less confident in their ability in male-typed fields like computing, in which cultural stereotypes hold that men are more competent than women [1, 8, 12, 12–14, 18, 19].

Yet while it is certainly plausible that gender differences in students' self-assessed computing ability cause fewer women to persist in computer science, there has unfortunately been limited study of these phenomenon within the specific field of computing. The literature that has examined gender, self-assessments, and persistence in computing finds that women do have lower self-assessments of computing ability, and that these lower self-assessments reduce women's stated intentions of persisting in computer science[19]. However, there has been no examination of how self-assessments influence women's actual persistence in computing, nor of why women might self-assess their computer science ability lower than men with objectively equal ability.

In the present research, we ask: 1) whether gender differences in self-assessments of computing ability contribute to gender disparities in persistence in computer science; and if so, 2) why do gender differences in self-assessments exist. To answer these questions, we first quantitatively determine if there are gender differences in self-assessments of ability in CS courses and whether these differences reduce women's persistence in computing using both surveys and enrollment data at a large American university. Second, we conduct qualitative interviews with a smaller sample of these students to understand what information women and men use in their self-assessments of computing ability, to determine the social-psychological mechanisms that may underpin gender differences in self-assessed ability.

Our research makes the following contributions to the literature, (1) we connect the literature on self-assessments of ability and persistence to the field of computing education; (2) we quantitatively test whether women have lower self-assessments of computing ability, and if so, whether it reduces their persistence in computer science; and (3) we increase the understanding of the mechanisms underlying gender differences in self-assessments of CS ability. In sum, this research draws attention to the importance of self-assessments of ability for gender equality in computing. It also has important implications for researchers and educators who want to design interventions to combat gender disparities in computing, as it suggests that changing students' self-assessments of computing ability may encourage more students to persist in computing, particularly those who are women.

2 GENDER STEREOTYPES, SELF-ASSESSMENTS, AND PERSISTENCE IN COMPUTING

Despite efforts to increase participation of women in fields dominated by men, women are still underrepresented in many high-status occupational fields such as science, technology, engineering and mathematics (STEM) [25]. One major factor contributing to the dearth of women in STEM fields are gender stereotypes about women's and men's abilities on male-typed tasks. Negative stereotypes about women's abilities in mathematics and science persist despite recent gains in participation in these fields during the last few decades. As early as elementary school, children are aware of these stereotypes and can express stereotypical beliefs about which science courses are suitable for boys and girls [1, 18]. A study of first grade children found that they held stereotypes that boys were better than girls at robotics and programming [12]. Further, the girls with stronger stereotypes about robotics and programming reported lower interest and self-efficacy in these domains. Thus, it follows that boys self-assess their task competency higher than girls in tasks that are considered to be masculine [13].

The relationship between self assessment and self-efficacy is reciprocal and intricate. Self assessment is defined as "a wide variety of mechanisms and techniques through which students describe (i.e., assess) and possibly assign merit or worth to (i.e., evaluate) the qualities of their own learning processes and products" [43]. Self assessment is a learning regulatory strategy whereas self-efficacy is thought to strengthen students' activation and use of these regulatory strategies, such as monitoring and evaluation [41, 46]. Self-assessment can increase perceived capability among students, which could affect students' self-efficacy [2].

A number of factors have been shown to affect self-efficacy in Computer Science. In particular, Kinnunen and Simon showed that demographic factors such as race and socio-economic status affect the self-efficacy of introductory programming students [26]. Gender has also been shown to correlate with students' self-efficacy. In particular, Beyer showed that women enter into computer science with lower self-efficacy than men [6] and Lishinski et al. showed that women revise their beliefs about self-efficacy earlier than men, based on course feedback [35]. Not only do women begin with lower self-efficacy, but they lose it more quickly than men. We hoped, in our study, to gain more insight into why this is the case in order to combat these effects.

Cultural messages about who is better suited to certain subjects reinforce stereotypes about gender differences in intelligence. Men are stereotypically more associated with notions of brilliance and genius than women (e.g., [4, 5, 7, 17, 27, 30, 47, 49, 51]. Bian et al., [8] found that linking success to brilliance lowered women's interest in a range of educational and professional opportunities. Prior research has shown that interest is a crucial precursor to participation in a field [11, 12, 23, 36, 39, 54]. Women were less interested in brilliance-oriented jobs than men were; they also perceived themselves to be less similar to the people in these jobs and were less sure they could succeed in them [8]. Cultural messages about men being better suited to work in STEM fields begins at an early age and leads girls to have lower interest, lower self-efficacy, and feeling as though they do not belong in these domains which can all play a role in why women self-assess their ability lower than men throughout their education and careers [1, 8, 12, 13, 18].

Certain STEM fields have lower rates of involvement from women than others. A 2017 study introduced a model with three overarching factors to explain these larger gender gaps in participation in computer science, engineering, and physics as they occur more in those fields than in biology, chemistry and mathematics. These three factors were 1) "masculine cultures that signal a lower sense of belonging to women than men, 2) a lack of sufficient early experience with computer science, engineering, and physics, and 3) gender gaps in self-efficacy" [12]. These factors help explain why women may be self-assessing lower than men which leads to less persistence in attaining computing careers [9, 14, 33, 34].

Despite the growing popularity of the computer science major, women are still underrepresented in this field, earning only 18% of bachelor's degrees [19, 31]. A 2009 poll found that 74 percent of college-bound boys ages 13–17 said that computer science or computing would be a good college major for them compared with 32 percent of their peers who were girls [22]. Gender stereotypes create a double disadvantage for women as they create gendered differences in self-assessments of computing ability which then decrease the likelihood that women will persist in CS. Fisk and Wingate [19] hypothesized that increasing top-performing women's self-assessments of computing ability would lead to an increase in women's intentions to persist in computing as a career choice. To test their hypothesis, they conducted a field experiment in a CS1 class in which the top fifty percent of students were given additional performance feedback from their instructor via email. Their results suggest that their intervention increased women's self-assessments of computing ability which led to increased intentions to persist in computing. This is because high self-assessed ability is a strong predictor of persistence. Individuals' perceptions of their competencies are powerful motivators that affect the choices they make, the effort and persistence they put forth, and the resilience they show in overcoming obstacles [55].

Studies show that low self-efficacy is associated with higher drop-out rates from college majors and can impact career choice [21, 32, 42, 45]. Researchers have found that self-efficacy and negative beliefs about computer programming abilities may contribute to the high drop-out rates in computer science [34, 48]. Correll's 2001 study found that boys were more likely than girls of equal mathematical performance to believe that they were competent at mathematics. Women hold themselves to a higher standard than men do which results in fewer women of equal ability to men self-assessing themselves as being good at math and science, resulting in fewer women pursuing STEM careers at large [14]. This study seeks to uncover the extent to which this gendered self-assessment is taking place in computer science, why, and what effect this may have on women's persistence intentions to remain in computing careers.

3 CURRENT RESEARCH

We use a mixed-method research design [15] consisting of both quantitative survey and enrollment data (n = 764) and qualitative interview data (n = 59) from students in four introductory computing courses (Java, CS-Principles, Matlab, and Python) at a large, engineering-focused, public U.S. state university in the Fall 2020 and Spring 2021 semesters. These introductory courses provided an ideal venue to learn about students' self-assessments and intentions to persist in computing, as student typically take these courses early in their college career and only one of these courses (Java) is required for the CS major at the university. This allowed us to study students with varying and evolving self-assessments and persistence intentions, and to avoid sampling on the dependent variable (i.e., students who have already committed to majoring in computing). Students who completed the surveys were given the option of being contacted to take part in qualitative interviews, and we randomly contacted some of these students to participate in interviews (with an oversampling of students from historically underrepresented groups in computing [e.g., women, Black students, Hispanic students, etc.). The qualitative interviews took place at the end of the Fall 2020 and Spring 2021 semesters.

We first describe our statistical analyses of the survey and enrollment data to determine whether women self-assess their computing ability lower than men with equal grades, and if so, whether these lower self-assessments decrease the likelihood that women persist in computing. We then detail our analyses of the qualitative interviews with students in these courses to understand how women and men perceive their own computing ability. This provides valuable insights on *why* women self-assess their computing ability lower than men.

4 QUANTITATIVE ANALYSES

4.1 Procedures

In our quantitative analysis of survey data, we analyze students' survey responses, grade data, and enrollments in computing courses. Students were recruited to take three surveys through their instructors across four different introductory computing classes during the Fall 2020 and Spring 2021 semesters. The surveys contained a multitude of questions about students' experiences in their computing courses, their self-perceptions, and their demographics. Some questions were about their self-assessments of computing ability, which we use in our analyses. Emails were sent to every student in the participating classes containing links to the surveys. Incentives for completion were left up to instructors: instructors chose to either make the surveys optional, count them as extra credit, or include them as a required portion of students' grades. Students chose whether to consent to the usage of their survey responses, grades, and/ or enrollment data for research and no students were required to participate in the study. Student responses from three surveys were included in our data analysis, and each survey was sent to students after a major course assignment.

4.2 Participants

764 students at a large public research university completed at least one of the three surveys and consented to the use of both their grade and enrollment data. This led to an overall response rate of 69.6% over the four different courses. In the analytic sample, participants ranged in age from 17-57 years old with an average age of 20.19. 68.70% of participants who reported their race identified as White, 9.03% identified as Asian, 3.61% identified as Black or African-American, 3.27% identified as South Asian, 3.27% identified as Hispanic or Latino/Latino, and 12.09% identified as some combination of races or entered another race. At the end of the semester, (10.4%) stated they were CS majors and (8.6%) stated they were minoring in CS. Survey procedures and question wording remained the same between the two semesters and across the different courses and course sections. Five instructors participated in Fall 2020 and four participated in Spring 2021.

4.3 Metrics

4.3.1 Gender. We use the variable 'Woman' to measure gender, which was coded as '1' for participants identifying as women and '0' for participants not identifying as women. 21.39% of participants identified as women, 70.00% identified as men, 8.17% did not specify their gender, and 0.44% specified a gender other than man or woman.

4.3.2 Course. We use a series of binary variables to control for differences between courses ('1' if the student completing the survey was enrolled in the course, '0' if the student was not enrolled in the course). CSC 110 is a dummy variable for a block-based CS Principles course for non-majors (11.32% of sample), CSC 111 is a dummy variable for a Python-based course largely for non-CS STEM majors (22.76% of sample), CSC 113 is a dummy variable for a MATLAB-based course largely for non-CS STEM majors (53.14% of sample), and CSC 116 is a dummy variable for a Java-based course for CS majors and minors (12.78% of sample).

4.3.3 Grades. A students' grade was calculated by averaging two major grades that were collected at two different periods in the semester. This resulted in a single value for each participant. While this is not a perfect measure of a student's grade in a course, it is an external measure, and there is no reason to believe that it is more systematically biased in favor of women (or men). The average grade was 85.65 (out of a maximum of 100), with an average score of 84.42 for women and 85.89 for men, a difference that was not statistically significant.

We asked course instructors to choose which assessments to use as the two major grades. Due to differences in structure and exam policies across courses, there was variation in what assessments were used (e.g., a test grade, a major assignment grade, or an aggregate grade). For this reason, we include controls for course in all of our models. We chose not to normalize grades by course because numeric grades (which correspond to letter grades) provide better information about a students' ability than standardized grades. This is because letter grades correspond to the level of mastery of the material (e.g., a grade of 'A' demonstrates a high level of mastery of the material even if it is the average grade in the course) versus being a measure of relative performance (as are standardized grades). In addition, students understand the meaning of letter grades differently than relative grades (i.e., an 'A' means something different to students than a 'B,' even if both constitute an average grade in their respective courses).

4.3.4 Self-Assessments of Computing Ability. Self-assessments of CS ability were measured using a 7-point Likert scale adapted from the National Educational Longitudinal Study of 1988 (NELS-88) (used by [13] and [19]. Participants were asked to what extent they agree or disagree with the following statements: 1), Computer Science is one of my best subjects and 2) I get good grades in Computer Science. Additionally, students were asked to describe their CS ability using a 7-point Likert scale ranging from 1=far below average and 7=far above average. The responses to these questions were combined across the three surveys, to create a nine-item self-assessment CS ability index with high scale reliability (Cronbach's alpha = 0.903), an average value of 4.64 and a standard deviation of 1.04.[1]

4.3.5 Computing Persistence. We obtained rosters from the university registrar to determine which students enrolled in CS 116 (the introductory Java course for CS majors/minors) or CS 216 (the subsequent CS majors/minors course) the semester after they took the surveys. These courses were chosen because they are the sequential core courses required for a major or minor in CS. A total of 116 students (18.6%) were enrolled in a sequential course the semester after completing the survey.

4.4 Path Analysis

We use a path model to understand the connection between gender, self-assessed computing ability, and computing persistence. Path models, a type of Structural Equal Model (SEM), consist of a series of linear regression equations. However, unlike linear regression analysis (which focuses on the effects of independent variable(s) on a single dependent variable), path models test the strength of complex, hypothesized relationships (i.e., a single variable can be both a dependent and independent variable) while controlling for potential confounds. In other words, a path model allows one to test how well your hypothesized relationships between variables–including the directionality of the relationship–are supported by the data. Path models also allow indirect effects to be teased apart from direct effects.

A path model was ideal for our analysis because it allowed us to test our hypotheses that gender influences self-assessments of computing ability, and that in turn, self-assessments of computing ability influence persistence in computing (while controlling for course and grades).[2]

Following the steps for path model construction outlined by Kline [28], we specified our model according to our hypotheses and added paths between the controls (i.e., grades and course) and the variables of interest (i.e., self-assessments of ability and persistence intentions). We then respecified the model to produce a well-fitting model by removing statistically insignificant paths (i.e., paths that may appear appropriate but which do not have a statistically meaningful relationship), using $p < 0.10$ as the cutoff for a path to be included in the refined model. Specifically, the following paths were removed due to a lack of statistical significance: 1) gender -> grades; and 2) grades -> persistence. CSC 110 is not included as a variable in the model as it is the reference category. The resultant model is quite well-fitting ($\chi^2 = 0.71$, $p = 0.40$; CFI = 1.000; RMSEA = 0.000; SRMR = 0.004; CD = 0.471).[3]

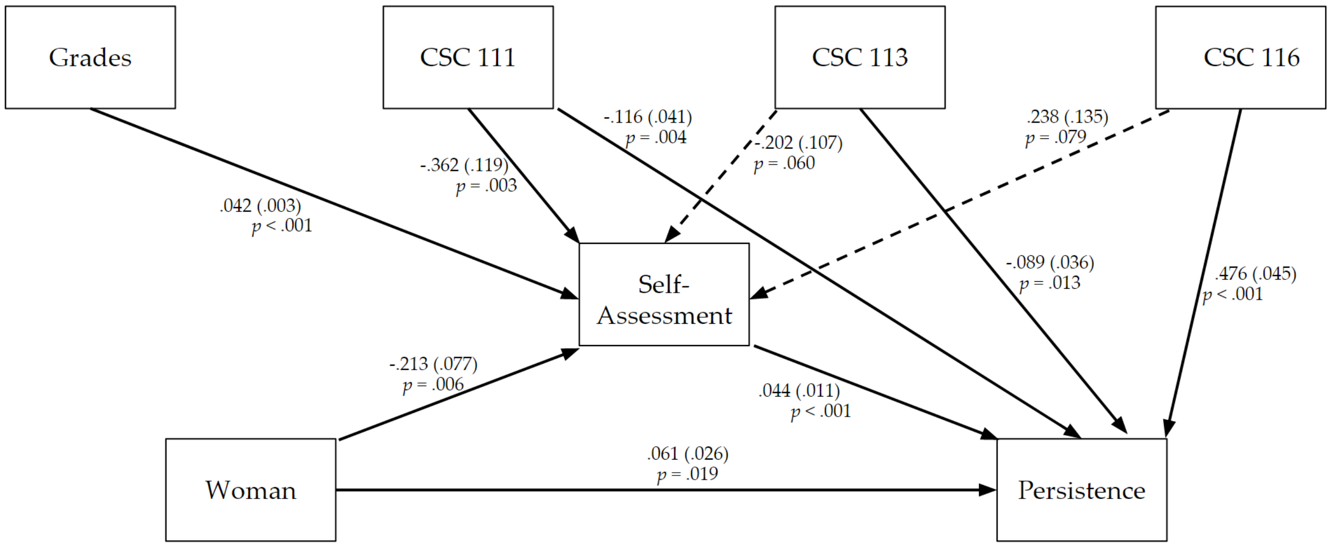
4.5 Quantitative Results

4.5.1 Gender Differences in Self-Assessed Computing Ability. We first determine if there are gender differences in self-assessed computing ability, controlling for course and grades. We find that women self-assess their computing ability 0.21 points lower than men who earn equal grades. This equates to women assessing

[1]The mean of a student's non-missing responses were used for students who did not complete all three surveys.

[2]Given that persistence in computing was measured using a binary variable instead of a continuous variable, we violated a statistical assumption of SEM. However, we chose to present these results, as studies on binary variables in SEM, "...suggest that results based on categorical variables approximate those of their continuous counterparts, except in the extreme case where dichotomous variables were skewed in opposite directions...," [24]. This is because, "Research consistently says the correlations (and corresponding parameter estimates) are attenuated (i.e., underestimated), and standard errors and χ^2 values overestimated (Schumacker & Beyerlein, 2000), which is good news because all these results err in the statistically conservative direction...," [24]. In addition, we ran the same path model using a continuous, attitudinal measure of persistence (in place of the binary measure of persistence) as a robustness check, and obtained substantively similar results.

[3]Unlike most statistical tests, the χ^2 test for a path model is a test of statistical insignificance. In other words, a well-fitting path models has a statistically insignificant χ^2 value. It is also generally agreed that well-fitting path models have CFI values over 0.95, SRMR values under 0.05, and RMSEA values under 0.05. For a primer on structural equation models and fit indexes, please see [24].



Notes: Fit statistics: $\chi^2 = .71$, $p = .10$; CFI = 1.00; RMSEA = .00; SRMR = .004. $N = 764$.

The first number on a path is the coefficient for that path, the number in parentheses is the standard error for the coefficient for the path, and the p-value is the precise statistical significance of the path.

A solid line means that the path is statistically significant at the p < .05 level; a dotted line indicates that the path is statistically significant at the p < .10 level.

CSC 110 is the reference category for course.

Figure 1: Path analysis examining the effect of gender on self-assessments and persistence in computing

their ability about 5% lower than men, given that the average self-assessment score of men in the sample is 4.70.

4.5.2 Effect of Self-Assessments on Persistence. We next determine if self-assessments of ability influence students' persistence in computing controlling for course and student gender. We find that self-assessments are predictive of computing persistence: for every one unit increase in self-assessed computing ability (measured on a 7-point Likert scale with a mean value of 4.64), students are 4.4 percentage points more likely to enroll in a computing course the following semester ($p < 0.001$). This is a large effect given that only 15% of students took a computing course the semester after being surveyed (recall our sample is largely non-majors). It is also worth noting that self-assessments are more important than grades in predicting a students' persistence in computing, as grades do not have a statistically significant direct effect on computing persistence. [4] However, there is an indirect effect of grades on persistence intentions through self-assessments of computing ability. In other words, grades influence students' self-assessments, which in turn influence their persistence in computing. While this effect is statistically significant ($p < 0.001$), it is small (the coefficient for this indirect effect is 0.001, which means that for every 1-point increase in grades, students are 0.1 percentage points more likely to enroll in a computing course the following semester).

[4]This is why there is no direct path between grades and computing persistence in the model, as it was found to be statistically insignificant during re-specification and removed.

4.5.3 Gender, Self-Assessments of Ability, and Computing Persistence. Lastly, we determine if the observed gender differences in self-assessed ability are large enough to reduce women's persistence in computing. In other words, we statistically test whether there is an indirect effect of gender on computing persistence *through* self-assessments of ability (controlling for grades and course). As predicted, we find evidence that women's computing persistence is reduced by their lower self-assessments of computing ability (indirect effect = -.009, $p = 0.022$). This means that women's persistence in computing is lowered by one percentage point because of their lower self-assessments of computing ability (controlling for their grades and course enrollments). Given that 14.1% of men take a computing course the semester after the survey, this means that women's lower self-assessments reduce their persistence in computing by 7% relative to men of equal ability.

4.5.4 Gender and Computing Persistence. It is worth noting that both the direct effect (coef. = 0.061, $p = 0.019$) and total effect (i.e., direct effect plus indirect effect: coef. = 0.052, $p = 0.048$) of being a woman on computing persistence is positive. This means that while gender differences in self-assessments reduce women's persistence in computing, the magnitude of the effect is not large enough to overwhelm women's greater levels of persistence in computing. In other words, women would be even more likely to persist in computing (6.1 percentage points more likely to persist than men, versus the observed 5.2 percentage points [controlling for grades and courses]) if it weren't for their lower self-assessments of computing ability. This is likely because women who take computing classes are particularly committed to computing [38].

5 QUALITATIVE ANALYSES

5.1 Participants

In order to understand why women self-assess their computing ability lower than men who earn the same grades, we conducted fifty nine interviews with undergraduate college students taking these courses. Of the 59 interview participants, there were 34 men and 25 women. Women were disproportionally represented in our qualitative sample as we purposefully over-sampled them.

5.2 Procedures

Students were asked a series of open-ended, semi-structured questions related to their past experiences in computing and their perceptions of performance in the course. Additionally, they were asked to reflect on their relative performances, their goals, and their experiences with group and partner work, as well as what significance grades have (i.e., what is a "good" or "bad" grade in their eyes?).

Written informed consent was secured from each participant prior to recruitment and a verbal confirmation of consent was received directly before commencing the interview. Interviews were held over a video conferencing application and lasted approximately 45 minutes. Both interviewers were affiliated with a different institution than the students; thus students could develop more comfort knowing that they were not sharing their thoughts with researchers directly affiliated with their institution. Further, interviews were completed after final grades for the courses were submitted. All students were enrolled in the university at the time of the interview. Participants received a $25 Amazon e-gift card via email after the conclusion of their interview. Pseudonyms are used in all write-ups to protect participants' identities.

For the qualitative interviews, all audio recordings ($n = 59$) were transcribed after the interviews were completed with the use of a transcription service, Otter. ai (with manual checking from a research assistant to assure fealty of the transcription). After transcribing, our qualitative expert used a qualitative data analysis software (NVivo) to conduct open-ended coding, followed by line-by-line coding. This qualitative analysis process involved an abductive approach as well. By utilizing an abductive approach to qualitative research [50], this allowed us to focus on the language that participants were using during the interviews and employ a "double engagement" of theory and specific methodology [50]. We were able draw upon our existing, interdisciplinary knowledge of inequalities, teaching and learning, and persistence in computing to inform the qualitative data collection, analysis, and discussion process. Line-by-line coding encourages researchers to think about the material and interviews that may differ from participants' interpretations [10]. Moving from line-by-line open coding into more focused coding, we began creating analytical categories which allow for a more focused coding approach and categorical evaluations and analyses.

Our coding revealed three key themes about gender differences in students' self-assessments of computing ability. While the three main themes captured in the interviews are related— particularly through their gendered nature— each one is distinct and captures the deep intricacies of how participants discussed their CS experiences at large. We describe the three overarching themes and describe patterns in students' reports of events, perceptions, and feelings in the following subsections.

5.2.1 Gender Differences in Comparative Assessments. Firstly, we found that women were much less likely than men to make favorable comparative judgements about their ability relative to their classmates. When asked to reflect on how well they performed relative to their classmates, men were much more likely to report doing better than their classmates, "I think I did pretty well, early on I felt a little bit behind… but as the course went on, I, I think I caught up if not exceeded the average" (Marco). Even when reflecting on the virtual nature of the course, which led to less peer interaction and more ambiguity in course performance, men drew upon the limited information and came up with more positive assessments of their performance. Men were less likely to draw upon specific mechanisms (i.e., grades, test scores, etc.) when responding to this question. Rather, responses from men tended to be rather short, such as "Pretty well." or "Um, I didn't look at the grade distribution, but I think I did decently well. I would say, at least, at least top 50%, probably around top 30%." (Waylon) and "I think I outperformed most of my classmates," (Oliver). Of course, this was not completely uniform across our sample of men as some did draw upon their prior experiences or discussions with others to formulate their comparative assessments,

> "Um, I think I did pretty much along with a bunch of my classmates, I feel like the class was really geared towards helping everyone no matter what their level was… because I had taken a previous computer class and I kind of understood the basic logic of I kind of understand how computers work… So I feel like I did kind of as well as everyone else who was in the course" - Brett
>
> "I mean, I don't know. That's a great question. I mean, I know that I performed pretty well, in the time that I spent in breakout rooms, meeting people. Probably... I mean... probably near, probably in the upper half at the very least. But I don't know, there were a lot of of very, very talented people in that class that I got the opportunity to work with. So I don't want to make a big judgment on that. " - Chad

However, as a whole, many more women than men (12% more) could not come up with an answer to this question. Though these students were all taking the same remote courses, women's responses point to much more ambiguity in their feelings of course performance. For example, Elizabeth noted,

> "I think because it was online, I couldn't really tell. Like sometimes when I was in labs, and things like that, or in office hours, there were people that seemed to be more confused than me, but I could never really pick up on how people were doing."

Yet, for women who did give a clear answer to this question, they self-assessed much lower than men did. 17.1% more women than men reported that they did poorer than their classmates.

At times, women made these evaluations in short, negative self-assessments,"Not as well," (Joan) and "Um, I would say probably on the lower end of performance, so um not very good." (Loraine).

Women were much more apprehensive about giving a comparative assessment, a few noting that they do not want to make a claim about their relative performance.

> "Um, I mean, that's kind of difficult, just because of how the course was set up this semester, being online and everything, it was hard for me to, like, communicate with other classmates. And I'm not quite sure what the average was in the course. So I feel like I did well. But the thing is, I have no idea. I don't really want to compare myself to other classmates, just because I don't know how they did. So yeah, that question is kind of hard for me to answer." - Catherine

Others gave longer responses in which they drew upon several specific points of reference to formulate their responses.

> "Ah, I don't, I don't know. It was kind of, um, probably very average. Probably, like, smack dab in the middle because, um, a couple of people that like I had worked with in the class they'd been coding forever and like, this one kid had been coding since the fourth grade... And I...like, I, that was like, my first time taking the course. And I just like, had absolutely nowhere but it was like weird because with our teacher, like, our grades weren't, like, it just didn't kind of make sense sometimes. Because like this, like that kid that had been coding since fourth grade. Um, yeah. So like he he would get like, 90s on his test, but then like, 50s on our projects, and we had like, five projects in the semester, and they're worth like, together, I think they made up like 25% of our grade. So like, but like, I would get, like 70s on the test, and then like 80s on the projects. So it was just like really weird of like that. But in terms of understanding and concepts, I think I'm very average" - Lindsey

While researchers typically do not quantitatively enumerate qualitative interview themes due to the small sample size, we felt that doing so would be quite illustrative in this case. We found that 44% of women reported that they performed better than their classmates, 24% reported that they performed the same (average) as their classmates, 20% reported that they performed worse than their classmates, and 12% reported that they were unsure of their course performance relative to their peers. As for men, 76.5% reported that they outperformed their peers, 20.6% said that they performed the same, 2.9% reported that they performed worse, and not a single man reported that they were unsure about their course performance relative to their classmates. It is important to note that many of the women who reported being unsure of their performance or poor performance also mentioned that they received As and Bs in the courses, whereas many men who reported that they performed average in the course reported receiving lower grades than As or Bs.

Thus, 32% of women reported that they performed worse or were unsure of their comparative performance, compared to only 2.9% of men, despite receiving similar letter grades. In sum, these findings suggest women are much more likely than men to experience ambiguity in their course performance relative to their peers.

5.2.2 Gender Differences in Personal Performance Standards. Secondly, women were more likely to have higher personal performance standards when compared to men who, on average, had lower personal standards and relied on overall class averages to gauge their performance. Men also used fewer sources of assessment (e.g., grades, group work, interactions with classmates, relative coding experience, etc.) to come to conclusions on their overall performance, but they were also more likely to have increased leniency when it came to grade expectations or general assessments of what a "good" or "bad" grade is. The following are examples of men reflecting on their perceptions of grades:

> "Um, I'd probably say, like, for me, it's probably a B or an A is what I consider a good grade. C is okay. And then a D and an F are bad grades." - Adam
>
> "I would say a bad grade, like, grade wise, just on an assignment is anywhere 70 or below." - Monroe
>
> "Um, well, typically, it's been, a C is where bad grades start. But, um, this year, I think I learned that that can actually be good, because everything averages out in the end. So as long as you keep the rest of the things high, then one C or one D is fine. So as long as I think it's more about consistency than anything, because if you're getting consistently bad grades, then you're gonna do bad, but consistently good grades, and a few bad grades sprinkled in isn't anything to worry about." - Omar

On the other hand, women were far more self-critical when it came to grades. Similar to their perceptions of comparative self-assessments, most women in this sample held themselves to very high standards when it came to their actual grades received on assignments, tests, or the course as a whole. For example, Charisse noted that she did "pretty good" in the course and elaborated to say that as long as you receive an A or higher (meaning an A+) that is "good". She went on to say,

> "Honestly, this is gonna sound so stupid, but anything below an A [is bad]. Like, I know that's like, not like the right answer. But I think for me, as I said, again, like, doing good in school is very important to me... so anything that goes to B+ or a B, definitely stresses me out. Like I have definitely had like my fair share of anxiety situations where, like, got an a B, or a B+ and I've been like, really stressed about it… Yeah, I think because high school is… you were like, a straight A student, and then you go to college... And you're like, wow, I'm really not that smart. What was I thinking?"

Similarly, other women noted many of the same feelings in the perceptions of their own grades,

> "As a perfectionist, I haven't really received very bad grades. So like, in general, I know that B's and A's are good, like are designated as good grades. Like, personally, my standards are pretty high for myself… And while I received B's on like tests, just like not

> overall classes, like, I understand that they're not terrible, but like, in my head, it's like I could have done better." - Regina

Included in this leniency that men seem to give more often in their grade evaluations is the perspective that as long as they are staying within the assumed average grades of the class at large, they do not need to panic.

> "I guess I would kind of say that a bad grade is… I guess lower than some sort of statistical average. I like to think that if we're learning a certain concept, and no one gets it, I don't think that a bad grade is... I guess it's relative based on everyone else. That's why I like classes with curves and stuff. Because I know that sometimes it's just hard for students to grasp the concepts or there's outside forces that are messing with people somehow. So I like to say a bad grade would be based on the other participants in the class." - Brett

Whereas women seem to push for, or aim for, higher grades regardless of how others are performing,

"Um, I definitely like to stay in the A/B range as much as possible. And when I get a C that definitely kind of kickstarts me to try and do better. Um, whereas like I knew if I get kind of a B on an assignment, but still, like, there's still hope for me to really push for an A in the class by making up for it in other areas. So that's that's kind of what my personal expectations of myself are." - Amber

However, both women and men did point to the type or discipline of the class as a factor in their grade expectations and eventual evaluations. For instance, Rhianna noted,

> "Um, I mean, I try to shoot for B's or A's. So I guess, like, anything below an 80, I consider, it really depends on the class, honestly. And relative to my performance in that class, because there's some classes where I'm like, oh, my gosh, I got a 60 on this exam. "Depends on the course, in organic chemistry, I've been happy with the 60. in chemical engineering, I've not been happy with the 73. So I think it depends on the rigor of the course and how well I expected to know the material. Like if I'm going into a chemical engineering exam, I practice a lot more than I did for organic chemistry, possibly. So I expect a 60 in organic chemistry, but in chemical engineering, I'd expect like a 90."

5.2.3 Gender Differences in Disrespectful Treatment. Thirdly, we found that women were more likely to experience disrespectful treatment, with an undertone of presumed incompetence, from their TAs and classmates. Several women reflected on their gender as a possible reason for experiencing this disrespectful treatment. For example, Joan noted,

> "Um, well, here's the positives: the professor was great. The classmates were great… And all my lab partners were great. Did have an absolutely not great TA. He gave me some pretty bad vibes… I don't know if it was because I was a female or what? … All my lab partners were guys, I think there were three females in our entire lab section out of 40… Like, I don't really care as a female [in] STEM, you're kind of used to it… Our TA came in the first time. And like, my partner asked the question, and like, he helped us specifically with the code and was like, 'Oh, yeah, this is where the problem is, like, this is what you need to do to fix it and stuff like that.' And like, we got help. And then like, the second time, I asked the question, he was like, 'Yeah, your code is wrong' and then left. And it was like that every single time."

One woman, after noting that she may not pursue CS any further, remarked on the importance of feeling welcomed in a community and that "...STEM careers are less accepting of others". When asked about the prospect of participating in a CS environment in the future, Rose noted that the concern comes from being a girl in CS since it's not as common. The following conversation ensued,

> "Yeah, yeah. I mean, there was just like some people who made me a little bit uncomfortable... just like being a girl in like a very male dominated environment is… there's like sort of more of a spotlight on you. There aren't really… specific things that made me like super uncomfortable, but just in general, like the way people talk to you and things like that is a little bit unnerving… it's hard to explain a bit. They were just a little bit more inclined to listen to like, my male counterparts, or there are some times I felt if I was working with a partner, it was less of a, like, professional relationship, I guess and more of like, interrogating a little bit."

Though the questions were focused on the nature of the current semester's CS course, many students told stories of prior STEM courses that still shape their views of the STEM field at large today. For instance, Amber reflected on her physics experience and related this to her feelings on the coding environment,

> "Yeah, the other underlying factor was that I was the only girl in my physics lab, which definitely kind of compounded the pressure that like, I felt like I was upholding a stereotype that girls can't code and there were all these guys in this class. And so it was already a little bit of a toxic environment. And so then, to have that feeling of failure associated with the coding was a little bit traumatic."

She also went on to note that in one physics lab, a student who is a man physically grabbed the mouse from her hand when they were working together. She continued,

> "Yeah, I definitely felt disrespected… there was another incident that was like, a little smaller, but not related to coding… we had group roles that were assigned. And without talking to the rest of the group, the guy automatically gave the other guy in the group the leadership role for that day. And so it was just an uncomfortable experience overall."

On the other hand, no men mentioned experiencing disrespectful treatment in the course. Overall, men did not remark on gender at all aside from one man who noted his position of privilege in the class as a straight, white man and two men who noted that there were women in their peer groups who contributed and meshed well with the men in the groups.

6 DISCUSSION

As predicted, our quantitative results demonstrate that women in computer science self-assess their ability lower than men with equivalent grades, and that these lower self-assessments of computing ability reduce the likelihood that women persist in computing. Our qualitative interviews illustrated three key mechanisms that drive gender differences in self-assessed ability: gender differences in comparative assessments, gender differences in personal performance standards, and gender differences in disrespectful treatment from TAs and classmates.

While we often think of grades as an objective measure that students will use to accurately self-assess their own ability, our interviews illustrated that grades are filtered through a gendered lens that reduces women's persistence in computing. Our interviews revealed that letter grades are often quite ambiguous, as the meaning associated with a given letter grade may differ for students and educators (e.g., a professor may see a 'B' as a good grade, while a student may not), especially for newer college students who are attempting to differentiate "normal" high school grades with the "new normal" college grades. We found that this ambiguity was amplified by gendered processes in that women were less likely to make positive comparative assessments relative to their classmates and more likely to personally hold themselves to a high grade standard.

In addition, we found that women were much more likely than men to experience disrespectful treatment that presumed a lack of competence. While only a minority of women experienced such treatment, no men mentioned any sort of disrespectful treatment. This treatment from students and TAs indirectly, and perhaps unknowingly, indicated to women that others perceived them to be "bad" at computing and/ or to be unworthy of respect in computing. These sorts of negative messages have been found to dampen women's persistence in computing [3, 8, 12, 29, 37]. And indeed, we found that these experiences were quite salient and had lasting impacts, as the women who received disrespectful treatment were more likely to internalize these messages about their lack of competence, and consequently, less likely to want to persist in computing.

While it is common for introductory CS students to perceive themselves to be bad at computing (this occurs for for a variety of reasons, many of which are common experiences in computing; for instance, struggling to fix errors or using resources to look up syntax [20]), our research illustrates the ways in which perceptions of computing ability are gendered. Comparative judgements, personal performance standards, and disrespectful interactions with TAs and other classmates combine to cause women to self-assess their computing ability lower than men with similar grades. These lower levels of self-assessment of computing ability directly reduce women's persistence in computing.

It worth noting that we found these gendered effects even though this research was conducted during the COVID-19 pandemic, during which time all courses in our sample were conducted remotely. This should have made gender less salient, as gender is not as easily discernible nor as omnipresent in a remote setting, as one can always turn off their camera. This suggests that the effects of gender on self-assessments may be larger in settings in which students meet in person.

7 LIMITATIONS

Our results have a few important limitations. First, the data was collected during the COVID-19 pandemic when courses were conducted remotely. While we hypothesize that this suppressed the effects of gender (implying that we would have found larger gender differences in self-assessments and negative course experiences in a face-to-face setting), we cannot be certain. Future research is needed to address this question. Second, as we studied students enrolled in computing courses at a large research university in the US, our results may not be generalizable to other populations. Lastly, there may be a non-response bias, given that poorly performing students were less likely to participate in our study because students were recruited through their courses using course credit and extra credit. However, we do not believe that this non-response bias systematically skewed our results, given the high response rate and the fact there is no reason to believe that there would be gender differences in non-response among low performing students.

8 CONCLUSION AND FUTURE WORK

This work makes significant contributions to the field of computing education by highlighting the importance of gender differences in self-assessed ability for gender equality in computing and by furthering our understanding of why gender differences in self-assessed computing ability exist. In addition, by specifying the causes of gender differences in self-assessments of computing ability (i.e., comparative judgements, personal performance standards, and disrespectful interactions with TAs and classmates), this research lays the foundation to reduce such disparities. If women are holding themselves to higher standards in computing because others do the same, interventions can be implemented to force decision-makers to evaluate women and men similarly. Other interventions could focus on making the criteria for success clear to everyone by removing ambiguity.

Future research should investigate the impact of remote learning on gendered experiences and explore interventions to reduce disrespectful interactions and the ambiguity of grade feedback. If the effects of gender were lessened during the COVID-19 pandemic due to the remote nature of learning, what will these students experience when they move back to in-person courses? Or when they attend in-person courses for the first time at the college level? In addition, future research should focus on interventions that reduce the impact of "bad actors" on women in computing courses. Perhaps more rigorous training and monitoring procedures could be implemented to ensure that computing TAs treat all students with respect. Lastly, interventions that reduce the ambiguity of feedback in computing and other male-typed fields should be studied. In a field that is facing a shortage of workers [16, 40], it is more important now than ever to explore such interventions that keep students in computing.

In conclusion, it is our hope that this research will help researchers and educators improve women's self-assessments of computing ability, and in turn, their persistence in computing.

ACKNOWLEDGMENTS

This research was supported by NSF #402130 and #2021396.

REFERENCES

[1] Nalini Ambady, Margaret Shih, Amy Kim, and Todd L Pittinsky. 2001. Stereotype susceptibility in children: Effects of identity activation on quantitative performance. *Psychological science* 12, 5 (2001), 385–390.
[2] Heidi L Andrade, Xiaolei Wang, Ying Du, and Robin L Akawi. 2009. Rubric-referenced self-assessment and self-efficacy for writing. *The Journal of Educational Research* 102, 4 (2009), 287–302.
[3] Sarah Banchefsky and Bernadette Park. 2018. Negative gender ideologies and gender-science stereotypes are more pervasive in male-dominated academic disciplines. *Social Sciences* 7, 2 (2018), 27.
[4] Mark Bennett. 1996. Men's and women's self-estimates of intelligence. *The Journal of social psychology* 136, 3 (1996), 411–412.
[5] Mark Bennett. 1997. Self-estimates of ability in men and women. *The Journal of Social Psychology* (1997).
[6] Sylvia Beyer. 2014. Why are women underrepresented in Computer Science? Gender differences in stereotypes, self-efficacy, values, and interests and predictors of future CS course-taking and grades. *Computer Science Education* 24, 2-3 (2014), 153–192.
[7] Lin Bian, Sarah-Jane Leslie, and Andrei Cimpian. 2017. Gender stereotypes about intellectual ability emerge early and influence children's interests. *Science* 355, 6323 (2017), 389–391.
[8] Lin Bian, Sarah-Jane Leslie, Mary C Murphy, and Andrei Cimpian. 2018. Messages about brilliance undermine women's interest in educational and professional opportunities. *Journal of Experimental Social Psychology* 76 (2018), 404–420.
[9] Erin Cech, Brian Rubineau, Susan Silbey, and Caroll Seron. 2011. Professional role confidence and gendered persistence in engineering. *American sociological review* 76, 5 (2011), 641–666.
[10] Kathy Charmaz et al. 1996. The search for meanings-grounded theory. *Rethinking methods in psychology* (1996), 27–49.
[11] Sapna Cheryan and Victoria C Plaut. 2010. Explaining underrepresentation: A theory of precluded interest. *Sex roles* 63, 7 (2010), 475–488.
[12] Sapna Cheryan, Sianna A Ziegler, Amanda K Montoya, and Lily Jiang. 2017. Why are some STEM fields more gender balanced than others? *Psychological bulletin* 143, 1 (2017), 1.
[13] Shelley J Correll. 2001. Gender and the career choice process: The role of biased self-assessments. *American journal of Sociology* 106, 6 (2001), 1691–1730.
[14] Shelley J Correll. 2004. Constraints into preferences: Gender, status, and emerging career aspirations. *American sociological review* 69, 1 (2004), 93–113.
[15] John W. Creswell and J. David Creswell. 2018. *Research design: Qualitative, quantitative, and mixed methods approaches.* SAGE Publications, Inc.
[16] Peter J Denning and Edward E Gordon. 2015. A technician shortage. *Commun. ACM* 58, 3 (2015), 28–30.
[17] Kristen C Elmore and Myra Luna-Lucero. 2017. Light bulbs or seeds? How metaphors for ideas influence judgments about genius. *Social Psychological and Personality Science* 8, 2 (2017), 200–208.
[18] Stephen J Farenga and Beverly A Joyce. 1999. Intentions of young students to enroll in science courses in the future: An examination of gender differences. *Science education* 83, 1 (1999), 55–75.
[19] Susan R Fisk, Tiah Wingate, Lina Battestilli, and Kathryn T Stolee. 2021. Increasing Women's Persistence in Computer Science by Decreasing Gendered Self-Assessments of Computing Ability. In *Proceedings of the 26th ACM Conference on Innovation and Technology in Computer Science Education V. 1.* 464–470.
[20] Jamie Gorson and Eleanor O'Rourke. 2020. Why do cs1 students think they're bad at programming? investigating self-efficacy and self-assessments at three universities. In *Proceedings of the 2020 ACM Conference on International Computing Education Research.* 170–181.
[21] Gail Hackett and Nancy E Betz. 1982. Mathematics Self-Efficacy Expectations, Math Performance, and the Consideration of Math-Related Majors. (1982).
[22] Catherine Hill, Christianne Corbett, and Andresse St Rose. 2010. *Why so few? Women in science, technology, engineering, and mathematics.* ERIC.
[23] Chris S Hulleman and Judith M Harackiewicz. 2009. Promoting interest and performance in high school science classes. *science* 326, 5958 (2009), 1410–1412.
[24] Dawn Iacobucci. 2010. Structural equations modeling: Fit indices, sample size, and advanced topics. *Journal of consumer psychology* 20, 1 (2010), 90–98.
[25] Shulamit Kahn and Donna Ginther. 2017. *Women and STEM.* Technical Report. National Bureau of Economic Research.
[26] Päivi Kinnunen and Beth Simon. 2011. CS majors' self-efficacy perceptions in CS1: results in light of social cognitive theory. In *Proceedings of the seventh international workshop on Computing education research.* 19–26.
[27] Bruce Kirkcaldy, Peter Noack, Adrian Furnham, and Georg Siefen. 2007. Parental estimates of their own and their children's intelligence. *European Psychologist* 12, 3 (2007), 173–180.
[28] Rex B Kline. 2015. *Principles and practice of structural equation modeling.* Guilford publications.
[29] Campbell Leaper and Christine R Starr. 2019. Helping and hindering undergraduate women's STEM motivation: Experiences with STEM encouragement, STEM-related gender bias, and sexual harassment. *Psychology of Women Quarterly* 43, 2 (2019), 165–183.
[30] Aaron Lecklider. 2013. *Inventing the egghead: The battle over brainpower in American culture.* University of Pennsylvania Press.
[31] Kathleen J Lehman, Linda J Sax, and Hilary B Zimmerman. 2016. Women planning to major in computer science: Who are they and what makes them unique? *Computer Science Education* 26, 4 (2016), 277–298.
[32] Robert W Lent and Gail Hackett. 1987. Career self-efficacy: Empirical status and future directions. *Journal of vocational Behavior* 30, 3 (1987), 347–382.
[33] Robert W Lent, Matthew J Miller, Paige E Smith, Bevlee A Watford, Robert H Lim, and Kayi Hui. 2016. Social cognitive predictors of academic persistence and performance in engineering: Applicability across gender and race/ethnicity. *Journal of Vocational Behavior* 94 (2016), 79–88.
[34] Karyn L Lewis, Jane G Stout, Noah D Finkelstein, Steven J Pollock, Akira Miyake, Geoff L Cohen, and Tiffany A Ito. 2017. Fitting in to move forward: Belonging, gender, and persistence in the physical sciences, technology, engineering, and mathematics (pSTEM). *Psychology of Women Quarterly* 41, 4 (2017), 420–436.
[35] Alex Lishinski, Aman Yadav, Jon Good, and Richard Enbody. 2016. Learning to program: Gender differences and interactive effects of students' motivation, goals, and self-efficacy on performance. In *Proceedings of the 2016 ACM Conference on International Computing Education Research.* 211–220.
[36] Charles A Malgwi, Martha A Howe, and Priscilla A Burnaby. 2005. Influences on students' choice of college major. *Journal of education for business* 80, 5 (2005), 275–282.
[37] Allison Master, Sapna Cheryan, Adriana Moscatelli, and Andrew N Meltzoff. 2017. Programming experience promotes higher STEM motivation among first-grade girls. *Journal of experimental child psychology* 160 (2017), 92–106.
[38] Carolina Milesi, Lara Perez-Felkner, Kevin Brown, and Barbara Schneider. 2017. Engagement, persistence, and gender in computer science: Results of a smartphone ESM study. *Frontiers in psychology* 8 (2017), 602.
[39] Carolyn Morgan, James D Isaac, and Carol Sansone. 2001. The role of interest in understanding the career choices of female and male college students. *Sex Roles* 44, 5 (2001), 295–320.
[40] Adams Nager and Robert D Atkinson. 2016. The case for improving US computer science education. *Available at SSRN 3066335* (2016).
[41] Frank Pajares. 2008. Motivational role of self-efficacy beliefs in self-regulated learning. *Motivation and self-regulated learning: Theory, research, and applications* 111139 (2008).
[42] Frank Pajares and M David Miller. 1994. Role of self-efficacy and self-concept beliefs in mathematical problem solving: A path analysis. *Journal of educational psychology* 86, 2 (1994), 193.
[43] Ernesto Panadero, Gavin TL Brown, and Jan-Willem Strijbos. 2016. The future of student self-assessment: A review of known unknowns and potential directions. *Educational psychology review* 28, 4 (2016), 803–830.
[44] Keith Quille, Natalie Culligan, and Susan Bergin. 2017. Insights on Gender Differences in CS1: A Multi-institutional, Multi-variate Study.. In *Proceedings of the 2017 acm conference on innovation and technology in computer science education.* 263–268.
[45] Dale H Schunk. 1989. Self-efficacy and achievement behaviors. *Educational psychology review* 1, 3 (1989), 173–208.
[46] Dale H Schunk and Peggy A Ertmer. 2000. Self-regulation and academic learning: Self-efficacy enhancing interventions. In *Handbook of self-regulation.* Elsevier, 631–649.
[47] Seth Stephens-Davidowitz. 2014. Google, tell me. Is my son a genius. *The New York Times* 19 (2014).
[48] F Boray Tek, Kristin S Benli, and Ezgi Deveci. 2018. Implicit theories and self-efficacy in an introductory programming course. *IEEE Transactions on Education* 61, 3 (2018), 218–225.
[49] Joachim Tiedemann. 2000. Gender-related beliefs of teachers in elementary school mathematics. *Educational Studies in Mathematics* 41, 2 (2000), 191–207.
[50] Stefan Timmermans and Iddo Tavory. 2012. Theory construction in qualitative research: From grounded theory to abductive analysis. *Sociological theory* 30, 3 (2012), 167–186.
[51] Sandra Upson and Lauren F Friedman. 2012. Where are all the female geniuses? *Scientific American Mind* 23, 5 (2012), 63–65.
[52] Anna Vitores and Adriana Gil-Juárez. 2016. The trouble with 'women in computing': a critical examination of the deployment of research on the gender gap in computer science. *Journal of Gender Studies* 25, 6 (2016), 666–680.
[53] Kelly Widdicks, Alice Ashcroft, Emily Winter, and Lynne Blair. 2021. Women's Sense of Belonging in Computer Science Education: The Need for a Collective Response. In *United Kingdom and Ireland Computing Education Research conference.* 1–7.
[54] Allan Wigfield and Jacquelynne S Eccles. 1992. The development of achievement task values: A theoretical analysis. *Developmental review* 12, 3 (1992), 265–310.

[55] Amy L Zeldin, Shari L Britner, and Frank Pajares. 2008. A comparative study of the self-efficacy beliefs of successful men and women in mathematics, science, and technology careers. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching* 45, 9 (2008), 1036–1058.