Increasing Women's Persistence in Computer Science by Decreasing Gendered Self-Assessments of Computing Ability

Susan R. Fisk, Tiah Wingate
Department of Sociology
Kent State University
sfisk@kent.edu,twingate@kent.edu

Lina Battestilli, Kathryn T. Stolee
Department of Computer Science
North Carolina State University
lbattestilli@ncsu.edu,ktstolee@ncsu.edu

ABSTRACT

Gender stereotypes about women's computing ability contribute to the dearth of women in computing by causing women to experience gender bias. These gender stereotypes are doubly disadvantaging to women because they create gender differences in self-assessments of computing ability, decreasing the likelihood that women will persist in Computer Science (CS). This is because students need to believe they have sufficient ability in a field in order to pursue it as a career.

Building on decades of Sociological theory, we hypothesized that increasing top-performing women's self-assessments of computing ability would increase those women's intentions to persist in computing. To test this hypothesis, we conducted a field experiment in a CS1 class in which the top 50% of students were given additional performance feedback from their instructor via email. The intervention increased these women's and men's self-assessed CS ability but only increased the women's CS persistence intentions. In sum, sending a single email increased top-performing women's intentions to persist in CS by 18%. A mediation analysis found evidence for the proposed causal path; namely, that the intervention increased the women's self-assessments of computing ability, which then increased their intentions to persist in computing. This research furthers our knowledge of the processes around self-assessments of ability and career choice that contribute to the dearth of women in CS. It also provides evidence for a lightweight intervention that may increase the number of women in computing, as prior research finds that intentions to persist are highly predictive of actual persistence in STEM fields.

CCS CONCEPTS

• **Social and professional topics** → **Women**; *CS1*; Student assessment.

KEYWORDS

CS Education, CS1, Persistence, Gender inequality, Self-assessments of ability, Women in STEM

ITiCSE 2021, June 26-July 1, 2021, Virtual Event, Germany

ACM ISBN 978-1-4503-8214-4/21/06...$15.00
https://doi.org/10.1145/3430665.3456374

ACM Reference Format:
Susan R. Fisk, Tiah Wingate and Lina Battestilli, Kathryn T. Stolee. 2021. Increasing Women's Persistence in Computer Science by Decreasing Gendered Self-Assessments of Computing Ability. In *26th ACM Conference on Innovation and Technology in Computer Science Education V. 1 (ITiCSE 2021), June 26-July 1, 2021, Virtual Event, Germany*. ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3430665.3456374

1 INTRODUCTION

Despite decades of efforts, women remain underrepresented in Computer Science (CS): for instance, in 2016 women accounted for only 19% of CS majors [7]. Gender stereotypes about women's and men's abilities on male-typed tasks are a potent contributor to this disparity, as research finds that they cause women to experience significant bias from others in male-typed fields like computing [6]. But one of the most nefarious effects of these gender stereotypes is the impact they have on women themselves. Research finds that gender stereotypes cause women to self-assess their ability on male-typed tasks (like CS) lower than men of equal ability [3, 4]. These gender differences in self-assessments of ability contribute to the dearth of women in STEM fields, as they cause women to be less likely to persist in male-typed fields relative to equally capable men [3]. This is because self-assessed ability is a strong predictor of persistence, as individuals need to believe that they have sufficient ability in a field in order to pursue it as a career.

Existing research suggests that the effect of gender stereotypes on self-assessments of ability will be largest when individuals receive ambiguous feedback about their ability [8, 12, 13]. This is problematic for women in CS, because CS students–regardless of gender–often experience uncertainty about the meaning of their grades. Grades in introductory STEM courses (like CS) tend to be lower than non-STEM courses [1, 11], which can leave students wondering if a given grade indicates that they have adequate ability to persist in computing.

We hypothesized that it may be possible to reduce gender differences in self-assessments of computing ability, and in turn, decrease gender differences in CS persistence, by reducing ambiguity in grade feedback. More specifically, we hypothesized that giving top-performing CS students additional feedback about their CS performance (i.e., giving them contextual information about their grades and explicitly telling them that they are doing well) would reduce uncertainty, which should decrease gender differences in 1) *self-assessments of CS ability*, and 2) *CS persistence intentions.*

We test our hypotheses via a lightweight email intervention on 193 students in an introductory programming course (CS1) for non-majors at North Carolina State University, a large public university. We chose this course as the students have not yet committed to computing as a major, allowing us to observe changes in their intention to persist in CS. Students in the control group received their numeric test grade with no contextual information about it while students in the intervention group received their numeric test grade with additional feedback (i.e., they were explicitly told by the professor via email that they were a top performer on the exam and had enough ability to persist in CS). Students were surveyed about their self-assessments of CS ability, CS persistence intentions, and demographics in the first week of the semester (Pretest) and then again after receiving their first exam grade (Posttest).

We focus on intentions to persist in computing, as opposed to actual CS persistence, due to data limitations and the fact that intentions to persist are predictive of actual persistence in STEM fields [9, 15]. Indeed, *"...hundreds of research efforts occurring [since the late 1960s] support the contention that intention is the 'best' predictor of future behavior,"* [9, p.7-8].

In sum, this research makes the following contributions:

- Brings and applies sociological theory (about gendered, social-psychological processes around self-assessments of ability and career choice) to research on computing education,
- Examines how self-assessments of CS ability contribute to CS persistence intentions, and
- Evaluates a lightweight intervention designed to increase the persistence intentions of women in CS.

Our main findings include:

- The lightweight intervention increased self-assessed CS ability for both top-performing men and women, and increased the CS persistence intentions of women,
- Improving women's self-assessed CS ability can increase their CS persistence intentions.

2 RELATED WORK

Gender Stereotypes, Self-Assessments of Ability, and Career Choice. Widespread cultural beliefs in the United States hold that men have more ability than women in STEM subjects like CS [10]. This has many problematic implications, including the fact that gender stereotypes influence self-assessments of ability when task performance is believed to be associated with gender [2]: in female-typed fields, women have higher self-assessments of ability than men, but in male-typed fields, men exhibit higher self-assessed ability than women [3]. In a convincing demonstration of the causal mechanisms underlying this phenomenon, Correll conducted an experiment in which she determined how the gender-typing of a task impacted self-assessments of ability [4]. All participants completed a contrast sensitivity exercise (a task that participants are falsely told has correct and incorrect answers) and all participants received the same score on the task. When participants were instructed that men and women perform equally well on the task, there were no gender differences in self-assessed ability. However, when participants were told that men were better at the task, men assessed their task competence higher than women.

These self-assessments of ability influence career choices because individuals need to believe that they have sufficient ability in a field in order to pursue it as a career. Indeed, Correll has found in observational data that self-assessments of ability influenced the career-relevant decisions of both women and men, even once actual ability was controlled [3]. This finding was repeated in her aforementioned 2004 experimental work that used the contrast sensitivity task. Cech, et al. elaborate on Correll's work, finding that confidence in one's ability to professionally succeed predicted undergraduates' persistence in STEM majors and careers [2]. In this manner, gender stereotypes can influence the persistence of women and men in computing through their impact on self-assessed ability.

Grades, Uncertainty, and Gender Stereotypes. Research suggests that the effects of stereotypes are typically amplified under conditions of uncertainty [8, 12, 13], as ambiguity increases the likelihood that individuals will use stereotypes to help them make sense of the situation. Since women are stereotyped as being less competent and able than men at computing tasks [10], gender stereotypes are likely to have the largest impact when women are uncertain about their computing ability. This has problematic implications for gender equality in computing because many CS courses offer ambiguous performance feedback. Average grades in STEM courses (like CS) tend to be lower than other university courses [1], with some of the lowest grades on campus being given in introductory STEM courses [11]. These lower-than-average grades provide ambiguous signals to students about their competence; for instance, what would be a low score in a non-STEM class may equate to an average score in a STEM class. And it may be unclear to students—especially students who are women—what minimum grade is necessary to have enough ability to pursue CS as a career choice.

Thus, we expect that for high-performing students in CS courses, additional, unambiguous feedback about their task performance (i.e., giving them contextual information about their grades and explicitly telling them that they are doing well) should increase their self-assessment of CS ability (since they are actually performing well), and this should increase their intentions to persist in CS. Moreover, given the aforementioned effects of uncertainty on the use of gender stereotypes, we expect that this feedback will have a larger positive impact on women's self-assessments of CS ability and CS persistence intentions than men's.

3 RESEARCH QUESTIONS

Our research goal was to assess the impact of additional feedback about task performance on: 1) self-assessments of CS ability, and 2) CS persistence intentions. Toward this goal, we ran a study in the context of a CS1 course for non-majors. We focus on the following research questions about *self-assessments of CS ability*:

RQ1: Are there gender differences in self-assessments of CS ability?

RQ2: Does additional performance feedback increase self-assessments of CS ability?

Related to *persistence intentions*:

RQ3: Are there gender differences in CS persistence intentions?

RQ4: Does additional performance feedback increase CS persistence intentions?

Narrowing in on gender-differentiated impacts of the intervention, we examine the following:

RQ5: Does additional performance feedback increase women's self-assessments CS ability more than men's?

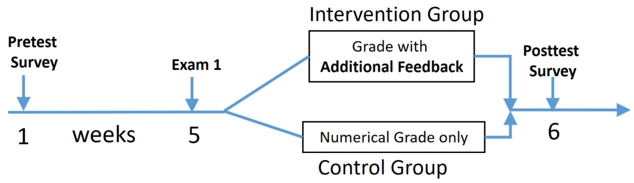

Figure 1: Timeline of the Experiment

RQ6: Does additional performance feedback increase women's CS persistence intentions more than men's?

RQ7: Does additional performance feedback increase women's CS persistence intentions *because* it increases their self-assessments of CS ability?

4 STUDY

We performed a field experiment in an introductory CS1 course for engineering students at a large public university.

Figure 1 illustrates the study timeline. Students were surveyed about their self-assessments of CS ability, CS persistence intentions and demographics in the first week of the semester (*Pretest Survey*) and then again after receiving their first exam grade (*Posttest Survey*). Students in the control group received their numeric test grade via email with no contextual information about it, while students in the intervention or treatment group received their numeric test grade and were explicitly told by the professor via email that they were a top performer on the exam (i.e., *Additional Feedback*). We ran the experiment twice, in two semesters of the same CS1 course.

4.1 Participants

This particular CS1 course is male-dominated, as around 80% of the students in the class are men. All students in the course were required to complete the Pretest Survey at the start of the course.[1] Students were offered 2 percentage points of extra credit for completing the Posttest Survey, which was given after the first exam.[2]

The study only included top performers (defined as students who received a score in the 50th percentile or above on the first exam) who consented to participate in our research. We limited our study population to top performers so that we could observe the effect of giving students honest positive feedback. In order to have a sufficient number of women in the sample, we conducted the field experiment twice (in the spring and fall semesters in 2018) and aggregated the data. All study procedures were identical between the two semesters and the course instructor was the same. Of the top performers across the two semesters, 193 students completed both the Pretest and Posttest Surveys and consented to the use of their data (response rate = 87.1%). Towards the end of the first survey, students were asked, "What is your gender?" and were given the options of: woman, man, or, "I identify as: ," with an open-response box. Of the top-performing students, 160 identified as men (82.9%) and 33 identified as women (17.1%).

[1] However, students were not required to consent to the use of their data.
[2] All students, independent of test performance and consent for data use, could earn extra credit by completing the surveys.

Table 1: Means and Standard Deviations of Computing Ability, Self-Assessed CS Ability, and CS Persistence Intentions (Together and by Gender). Mean and (Std Dev) presented.

	All	Women	Men
Observations (n=)	386	66	320
# of Students	193	33	160
Woman	0.17	1	0
Semester	1.39	1.33	1.40
Time	1.50	1.50	1.50
Additional Feedback	0.25	0.27	0.25
Computing Ability	87.64	86.59	87.86
	(4.73)	(4.20)	(4.81)
Self-Assessed CS Ability	4.46	4.02	4.55
	(1.16)	(1.04)	(1.17)
CS Persistence Intentions	3.50	3.14	3.58
	(1.42)	(1.49)	(1.40)

4.2 Intervention Design

After the first exam, top-performing students were stratified into groups according to their test performance: top 10%, top 11-25%, or top 26-50%. Within each performance group, students were further stratified by gender. Within each group of test performance and gender, students were randomly assigned to either the control group or the intervention group. We stratified random assignment to condition to ensure that there were the same number of women in the control and intervention groups at each level of test performance.

Students in the control group received an email from the course instructor containing their numeric grade on the exam:

> *"You got an {{Grade}}% on {{Test}}."*

Students in the intervention group received an email from the professor giving them their numeric grade on the exam, as well as contextual information about their performance (i.e., *additional feedback*). This included whether they placed in the top 10%, top 25%, or top 50% of students in the course. At the end of the email message for students in the intervention group was a GIF of dancing minions, in order to affectively reinforce the positive feedback of the email message.[3] Specifically, students in the intervention group received the following email message:

> *"You got an {{Grade}}% on {{Test}}! Congratulations! Since average grades in STEM courses tend to be lower than in other university classes, I wanted to make sure that you know that you are a top performer in the class! [You scored in the top X%, and earned the X highest score in the class!/ Your score places you in the top quarter of all grades on this test!/ You scored better than half of the students in this class!] Keep working hard! I know that you have what it takes to be successful in Computer Science!"*

Emails to students in both conditions also contained a link to the Posttest Survey.

[3] Minions are cartoon characters from the children's movie *Despicable Me.* As defined by Edwards, "Minions are a species of tiny yellow henchmen; they look like unusually dressed Mike and Ike candies," [5].

4.3 Metrics

Table 1 shows the means and standard deviations for all the metrics for all participants, with a breakdown by gender.

Controls. There were 193 students, with 33 women (assigned a value of '1' for the *Woman* category) and 160 men ('0' for *Woman*). *Semester* takes on a value of '1' for the spring and '2' for the fall. *Time* takes on a value of '1' for the Pretest survey and '2' for the Posttest survey. *Additional Feedback* takes on a value of '0' for all observations at time 1 (as no students had received additional feedback at this time), and takes on a value of '1' at time 2 if the student was in the treatment group. We controlled for *Computing Ability* using student scores on the first exam.[4] For the students included in this study (who received a score in the top-50th percentile), the average score on the first test was 87.64 (SD = 4.73), with a minimum score of 80 and a maximum score of 99. The average grade for women in the study was 86.59 (SD = 4.20) and for men in the study was 87.86 (SD = 4.81), a difference that was not statistically significant.

Self-Assessment of CS Ability. Self-assessments of CS ability were measured using 7-point Likert scales adapted from the National Educational Longitudinal Study of 1988 (NELS-88) (used by Correll [3]). Students were asked to what extent they agreed/ disagreed with the following two statements: 1) *Computer Science is one of my best subjects* and 2) *I get good grades in Computer Science* (wherein 1 = strongly disagree, 4 = neither agree nor disagree, and 7 = strongly agree). Students were also asked to describe their ability in Computer Science using a 7-point Likert scale (wherein 1 = considerably below average, 4 = average, and 7 = considerably above average). These responses were combined for a three-item self-assessment CS ability index that could range from 1 to 7 and which had an alpha of 0.85.

CS Persistence Intentions. Students were asked about their CS persistence intentions using 7-point Likert scales adapted from Correll [4]. Students were asked to state how likely they were to: 1) take another course in CS, 2) minor in CS, 3) apply to graduate programs requiring high levels of CS ability, and 4) apply for high-paying jobs requiring high levels of CS ability (wherein 1 = highly unlikely, 4 = neither likely nor unlikely, and 7 = highly likely). This led to a four-item CS Persistence Intentions index that could range from 1 to 7 and which had an alpha of 0.85.[5]

4.4 Analysis

We use linear regression models to determine the impact of our intervention. Our linear regression models predict a given outcome (i.e., the dependent variable) using our key independent variable of interest (i.e., our intervention) while simultaneously controlling for other variables that could influence the relationship between the intervention and the outcome variable. This allows us to control for any student differences that systematically vary by student gender or condition (e.g., differences in initial interest in persisting in computing, differences in objective computing ability). These controls, in combination with the random assignment to condition, give us more confidence that any observed differences between the treatment and control groups can be attributed solely to the email intervention. We use linear mixed models with repeated measures because they take the non-independence of observations into account, as there are two observations from each student. Within-group errors were modeled to have an autoregressive structure with a lag of 1, given the time-lag between observations.

We predict self-assessments of computing ability in Table 2 and CS persistence intentions in Table 3. In these tables, the models have a nested structure. Model A predicts the dependent variable controlling only for gender, semester, and time (allowing us to determine if there are mean gender differences in our sample). Model B predicts the same dependent variable as Model A, and adds computing ability to the controls used in Model A (allowing us to determine if gender differences exist once computing ability is controlled). Nested models allow us to determine how variables mediate and moderate the effects of other variables: for instance, if we observed a gender difference in self-assessments of ability in Model A, but there was no gender difference in Model B, that would show that the observed gender difference in self-assessments of CS ability in Model A was caused by gender differences in computing ability.

For evaluating RQ1–RQ6, Models C and D are of particular interest as these evaluate the impact of the intervention. Model C adds additional feedback to the controls (allowing us to determine the overall effect of the intervention on the dependent variable, controlling for gender, semester, time, and computing ability). Model D adds the interaction between additional feedback and gender to the controls used in Model C (allowing us to determine if the effect of the intervention varied by student gender).

Evaluating RQ7 requires a mediation analysis, which tests whether the relationship between two variables (i.e., the intervention and CS persistence intentions) is explained by a third, intermediate variable (i.e., self-assessments of CS ability). This allows us to determine whether the email intervention increased women's intentions to persist in CS *because* it increased their self-assessments of CS ability.

5 RESULTS

5.1 Effect of Intervention on Self-Assessments of CS Ability (RQ1, RQ2, RQ5)

We find gender differences in self-assessments of CS ability, even after controlling for objective computing ability (RQ1). Women self-assess their CS ability approximately 0.5 points lower than men on a seven-point scale ($p < .05$) (Table 2, Models A-D, the coefficients for *Woman* range from -0.47 to -0.53). This means that women self-assess their ability approximately 10% lower than men with equivalent computing ability (given that the average man's self-assessment of CS ability score was 4.55, as reported in Table 1).

We also find that the intervention increases both women's and men's self-assessments of CS ability (RQ2). In Table 2, Model C, we

[4]Computing ability was assumed to be unchanged between time 1 and 2 in our models, as we could not assess computing ability at time 1 (as there were no course assessments at this time). While this is an imperfect assumption, it does provide meaningful information about a students' objective computing ability, as course performance tends to be highly correlated over time (so students with high computing ability at time 2 likely had high computing ability at time 1).

[5]In a departure from the index used in Correll 2004, we did not include students' stated likelihood of majoring in CS. This is because this particular CS1 course is for non-majors and is taken by students who have already declared a major in engineering.

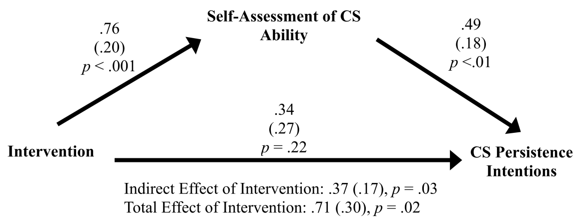

Figure 2: Mediation Analysis for Effect of Intervention on Self-Assessments of CS ability and CS Persistence Intentions, Women Only

find that the intervention increases self-assessments of CS ability by 0.25 points ($p < .05$) (controlling for gender and computing ability), which represents a 5.6% increase in self-assessments of ability (given that the average student's self-assessments score was 4.46). *However, we do not find any evidence that the intervention decreases the gender gap in self-assessments of CS ability (RQ5)*, given that the effect of the interaction between the self-assessment intervention and gender is not statistically significant (Table 2, Model D).

> There are gender differences in self-assessments of CS ability, and the additional feedback intervention increases both women's and men's self-assessments of CS ability. However, the intervention does not decrease the gender gap in self-assessments of CS ability.

5.2 Effect of Intervention on Computer Science Persistence Intentions (RQ3, RQ4, RQ6)

We find evidence that women have lower CS persistence intentions than men (RQ3), once we control for computing ability, additional feedback, and additional feedback interacted with gender (Table 3, Model D). Women score 0.55 points lower than men on the CS persistence index ($p < .05$) (Table 3, Model D). This means that women's persistence intentions are about 15% lower than equally able men (given that the average CS persistence index score for men was 3.58, as reported in Table 1).

We do not find evidence that the additional feedback intervention increases the CS persistence intentions of both women and men (RQ4), as there is no main effect of additional feedback on persistence intentions ($p > .10$) (Table 3, Models C and D).

We do find evidence that the additional feedback intervention increases the CS persistence intentions of women (RQ6), as the effect of the intervention interacted with gender is positive and statistically significant ($p < .05$) (Table 3, Model D). Moreover, the magnitude of the effect of the intervention is substantial as it increases the CS persistence intentions of women by 0.58 points, representing an increase of about 18% (given that the average CS persistence index score for women was 3.14) (Table 1).

> Women have lower CS persistence intentions than men. The intervention increased women's CS persistence intentions by 18%. It did not increase men's CS persistence intentions.

Table 2: Linear Mixed Models with Repeated Measures Predicting Self-Assessments of CS Ability

	Model A	Model B	Model C	Model D
Woman	-.53**	-.47*	-.47*	-.51*
	(0.19)	(0.19)	(0.19)	(0.20)
Semester	.01	-.03	-.04	-.04
	(0.15)	(0.15)	(0.15)	(0.15)
Time	.57***	.57***	.44***	.44**
	(0.06)	(0.06)	(0.08)	(0.08)
Computing Ability		.05***	0.05***	.05***
		(0.02)	(0.02)	(0.02)
Additional Feedback			.25*	.23†
			(0.11)	(0.12)
Woman x Add' Feedback				0.13
				(0.22)
Intercept	3.68***	-0.75	-0.53	-0.53
	(0.24)	(1.34)	(1.35)	(1.35)

†$p \leq .10$ *$p \leq .05$ **$p \leq .01$ ***$p \leq .001$.
n = 386 observations from 193 students

Table 3: Linear Mixed Models with Repeated Measures Predicting CS Persistence Intentions

	Model A	Model B	Model C	Model D
Woman	-0.44†	-0.40	-0.40	-0.55*
	(0.25)	(0.25)	(0.25)	(0.26)
Semester	-0.01	-0.03	-0.03	-.03
	(0.20)	(0.19)	(0.19)	(0.19)
Time	0.16*	0.17*	0.18†	.18†
	(0.07)	(0.07)	(0.10)	(0.09)
Computing Ability		0.03†	0.03†	0.03†
		(0.02)	(0.02)	(0.02)
Additional Feedback			-.03	-0.14
			(0.13)	(0.14)
Woman x Add' Feedback				.58*
				(0.24)
Intercept	3.34***	0.42	0.39	0.38
	(0.31)	(1.78)	(1.78)	(1.78)

†$p \leq .10$ *$p \leq .05$ **$p \leq .01$ ***$p \leq .001$.
n = 386 observations from 193 students

5.3 Why Does The Intervention Increase Women's CS Persistence Intentions (RQ7)?

We conducted a mediation analysis to determine the direct and indirect effects of the additional feedback intervention (see Figure 2)[6]. This analysis revealed that *the intervention increased women's self-assessments of CS ability ($p < .001$), and that these improved self-assessments of CS ability then increased women's CS persistence intentions ($p < .01$)* (controlling for computing ability, semester, and time (RQ7)). This means that self-assessments of CS ability mediate the relationship between the intervention and CS persistence intentions ($p < .05$): this indirect effect (.37) represents 52.1% of the

[6]Given that the data was multi-level, the GSEM and ncolm commands in STATA were used to make these predictions.

total effect of the intervention (.71) on CS persistence intentions. In sum, the intervention increased women's self-assessments of computing ability, which then increased their intentions to persist in computing.

> The intervention increased women's CS persistence intentions because it increased their self-assessments of CS ability.

6 DISCUSSION

Top performing students in the intervention condition were told by the instructor they were top performers and had sufficient ability to study computer science. We found that the intervention increased all students' self-assessments of CS ability. However, the intervention only increased women's CS persistence intentions. A mediation analysis revealed that the intervention increased women's CS persistence intentions *because* it increased their self-assessments of CS ability.

These results have important practical and theoretical implications for computing education. This research advances our knowledge of lightweight interventions that can decrease gender inequality in CS and possibly other male-dominated fields. Although not as effective as widespread organizational change, implementing this additional feedback intervention is simple and can be easily transferred across classrooms, universities, and organizational settings. In this manner, our intervention has the potential to meaningfully increase the retention of women in computing and other male-dominated fields. In addition, the results suggest that it is possible for educators to overwhelm the negative effects of gender stereotypes through providing additional feedback that reduces ambiguity and uncertainty.

6.1 Other Factors at Play

Was the intervention effective simply because it was an email from the professor? One might argue that *any* email from a professor might increase women's intentions to persist in computing. However, both control and intervention groups received an email from the professor. In addition, bottom-performing students were sent encouraging emails with resources to help them improve their grade (in a separate study performed in the same course), and we found no evidence that these emails impacted self-assessed computing ability nor intentions to persist in computing. Lastly, the mediation analysis found evidence for our proposed causal path: namely, that the intervention increased top-performing women's persistence intentions *because* it increased their self-assessed computing ability. Thus, our data strongly suggests that it was the content of the email that positively affected the top-performing women.

Did discussions between students about the intervention influence the results? The median time between the first email being sent to a student in the course and a student's completion of the survey was 1.3 hours. Thus, it appears that students tended to immediately complete the surveys upon opening the email from the professor. This would make it difficult for them to confer with classmates, especially given that the emails were sent outside of course meeting hours (when students were unlikely to be physically present with their classmates).

Was there a gender difference in the effect of the intervention because the men were objectively better at computing? This explanation could not explain our results because: 1) there was not a statistically significant gender difference in objective computing ability, and 2) we found these effects even when we controlled for computing ability.

6.2 Limitations

The ability feedback given to students was likely legitimated by the fact that it was given by an acknowledged expert in CS (i.e., the professor). Other avenues of positive feedback (e.g., through bots, online messaging systems, pull request reviews) may not be viewed as legitimate by students. Further study is required to assess their effectiveness. We studied top-performing students in a CS1 course for non-majors at a large research University in the United States: our results may not generalize to other populations. Persistence intentions may not translate to actual persistence in computing.

Another limitation of this study was that there were only 33 women in our sample. To reduce the likelihood of a Type 1 error, we used linear mixed models to control for existing differences between students in the control and intervention group (as the small number of women reduces the likelihood that random assignment will ensure equivalent treatment and control groups). In addition, we presented a statistically-significant mediation analysis that supports our proposed causal chain (e.g., that the intervention increased women's self-assessments of computing ability, which then increased their intentions to persist in computing).

7 CONCLUSION AND FUTURE WORK

We found evidence that our lightweight intervention increased top-performing women's CS persistence intentions by increasing their self-assessments of CS ability. Future research should study other populations, as well as the effect of this and other self-assessment interventions on actual CS persistence. Future research should also examine the effect of this self-assessment intervention on Black and Hispanic students, as there is theoretical reason to believe that this intervention would also be beneficial to them (as they also face negative stereotypes about cognitive ability [14]). Unfortunately, there were too few Black and Hispanic students in our sample to test this hypothesis (although preliminary analyses were promising).

Women face challenging environments and negative gender stereotypes in many CS environments. It is our hope that this additional feedback intervention will allow more of them to persist.

ACKNOWLEDGMENTS

This research was supported in part by NSF SHF-#1749936, NSF DUE-#2021396, and NSF DUE-#2021330.

REFERENCES

[1] Alexandra C Achen and Paul N Courant. 2009. What are grades made of? *Journal of Economic Perspectives* 23, 3 (2009), 77–92.
[2] 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.
[3] 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.
[4] Shelley J Correll. 2004. Constraints into preferences: Gender, status, and emerging career aspirations. *American sociological review* 69, 1 (2004), 93–113.
[5] Phil Edwards. 2015. Minions, Explained. (2015). https://www.vox.com/2015/7/10/8928069/minions

[6] Susan R Fisk and Cecilia L Ridgeway. 2018. Framing gender. In *Handbook of the Sociology of Gender*. Springer, 157–171.

[7] National Science Foundation. 2019. Women, Minorities and Persons with Disabilities in Science and Engineering (NSF Report No. 19-304). (2019). https://ncses.nsf.gov/pubs/nsf19304/

[8] Elizabeth H Gorman. 2006. Work uncertainty and the promotion of professional women: The case of law firm partnership. *Social Forces* 85, 2 (2006), 865–890.

[9] David M Merolla, Richard T Serpe, Sheldon Stryker, and P Wesley Schultz. 2012. Structural precursors to identity processes: The role of proximate social structures. *Social Psychology Quarterly* 75, 2 (2012), 149–172.

[10] Brian A Nosek, Frederick L Smyth, Natarajan Sriram, Nicole M Lindner, Thierry Devos, Alfonso Ayala, Yoav Bar-Anan, Robin Bergh, Huajian Cai, Karen Gonsalkorale, et al. 2009. National differences in gender–science stereotypes predict national sex differences in science and math achievement. *Proceedings of the National Academy of Sciences* 106, 26 (2009), 10593–10597.

[11] Kevin Rask. 2010. Attrition in STEM fields at a liberal arts college: The importance of grades and pre-collegiate preferences. *Economics of Education Review* 29, 6 (2010), 892–900.

[12] Cecilia L Ridgeway. 2011. *Framed by gender: How gender inequality persists in the modern world*. Oxford University Press.

[13] Muzafer Sherif. 1935. An experimental study of stereotypes. *The Journal of Abnormal and Social Psychology* 29, 4 (1935), 371.

[14] Claude M Steele. 2011. *Whistling Vivaldi: How stereotypes affect us and what we can do.* WW Norton & Company.

[15] Jan E Stets, Philip S Brenner, Peter J Burke, and Richard T Serpe. 2017. The science identity and entering a science occupation. *Social science research* 64 (2017), 1–14.