Nevertheless, She Persisted:

An intervention to increase women's persistence intentions in CS

Katie Stolee, PhD

North Carolina State University

Collaborators!

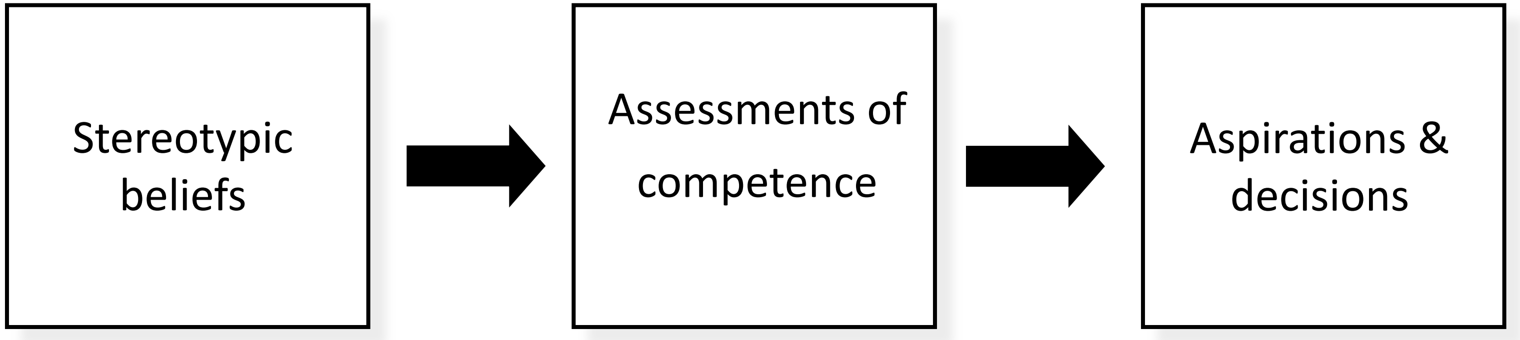
Susan Fisk
Associate Professor
Department of Sociology
Kent State University

Kathryn T. Stolee
Associate Professor
Dept. Computer Science
North Carolina State University

Lina Battestilli
Teaching Associate Professor
Dept. of Computer Science
North Carolina State University

Gender stereotypes, self-assessments of ability, and career choice

- There are many factors that influence career choice and contribute to women’s underrepresentation in CS
- Individuals must believe they have adequate ability for a given career in order to pursue that career
- Gender stereotypes influence self-assessments of ability (Correll 2004)
- Women are stereotyped as having less ability in male-typed, STEM fields (Fisk and Ridgeway 2018)


Stereotypic beliefs
Assessments of competence
Aspirations & decisions

The Role of Uncertainty

Ambiguous feedback ➡️ Uncertainty

Uncertainty ➡️ greater effects of gender stereotypes

How can we **improve feedback** to lessen the impact of gender stereotypes on self-assessed ability?

Intervention

Current Feedback Mechanism

Exam 1 Feedback External Inbox ×

Katie Stolee
to me

11:55 AM (1 minute ago)

Dear Kathryn,

You got a 86% on the test.

Intervention

Exam 1 Feedback External Inbox ×

Katie Stolee
to me

11:50 AM (5 minutes ago)

Dear Kathryn,

You got an 86% on exam 1! 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! Your score places you in the top quarter of all grades on this test! Keep working hard! I know that you have what it takes to be successful in computer science!

Intervention Motivation

Removing ambiguity by giving students clear, unambiguous feedback

↓

Decreases impact of gender stereotypes

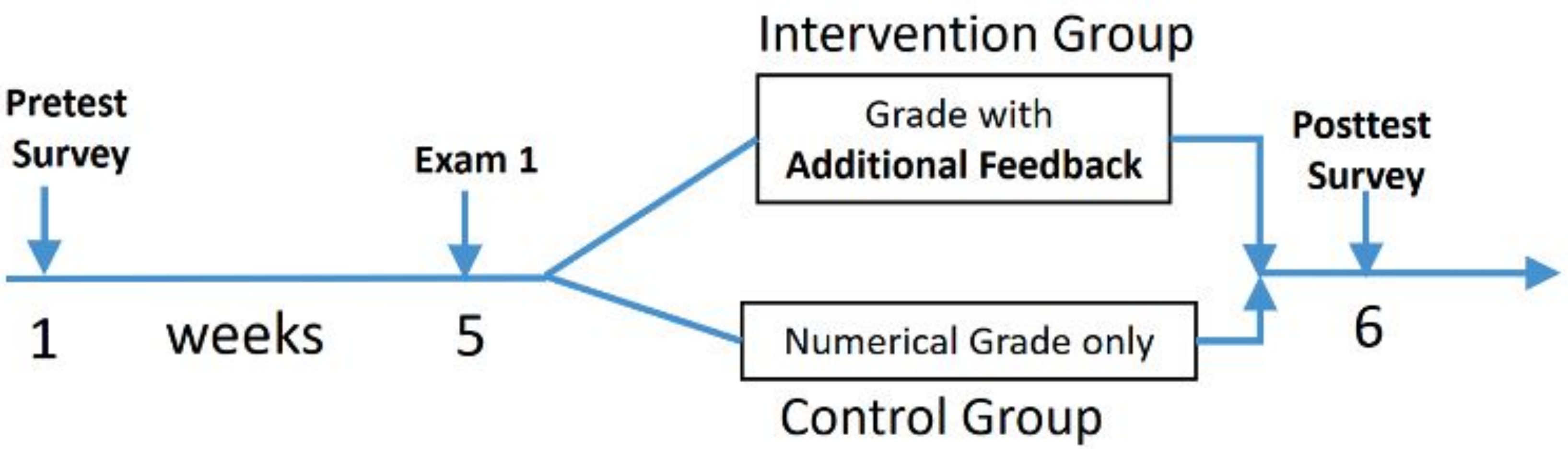
↓

Improves women’s self-assessment of ability

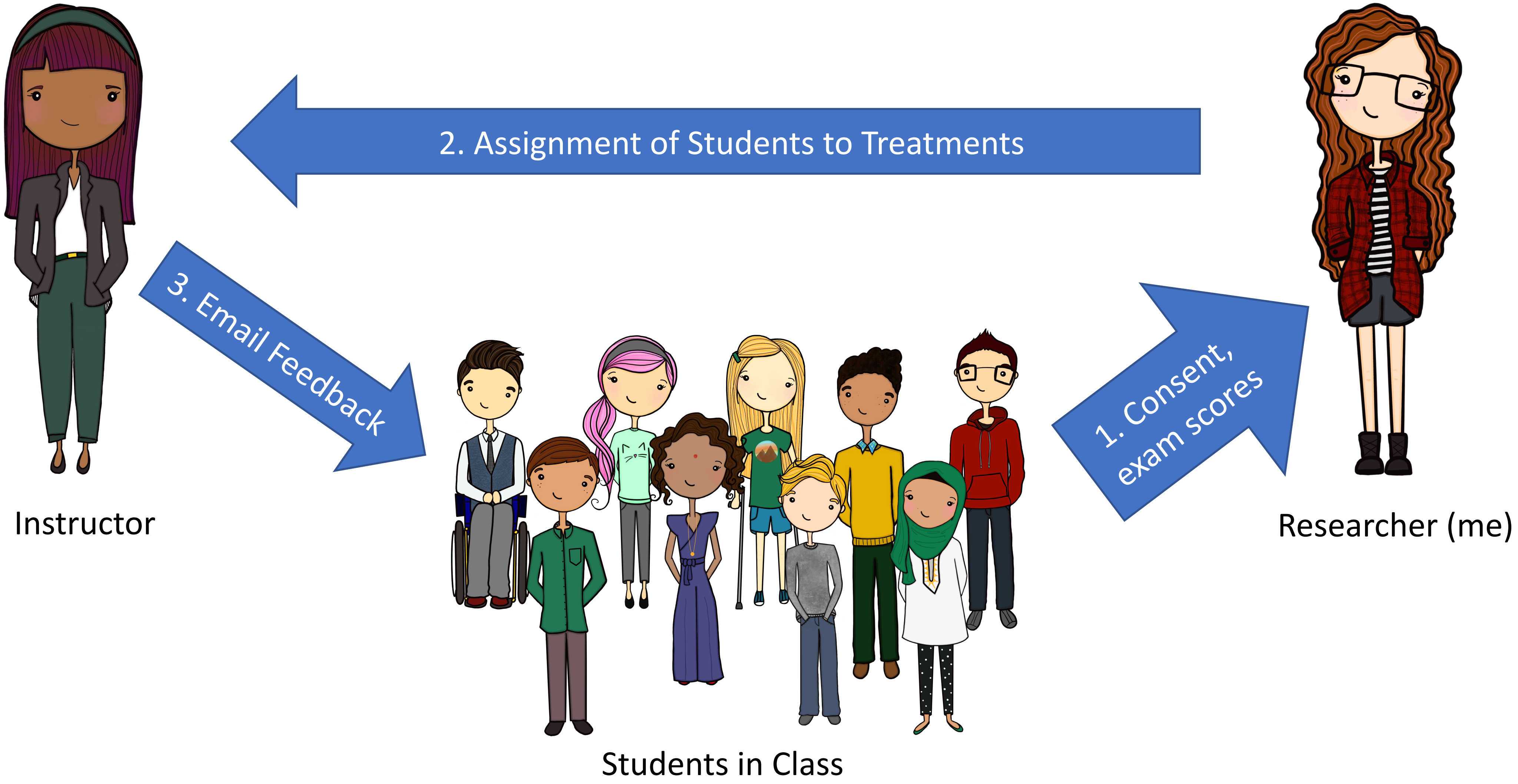
↓

Increases women’s intentions to persist in computing

Study Design - Classroom



Study Operation



Metric – Self-assessed CS Ability

Indicate to what extent you agree with the following statements regarding **Computer Science**.

	Strongly disagree	Disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Agree	Strongly agree
Computer Science is one of my best subjects	○	○	○	○	○	○	○
I get good grades in **Computer Science**	○	○	○	○	○	○	○

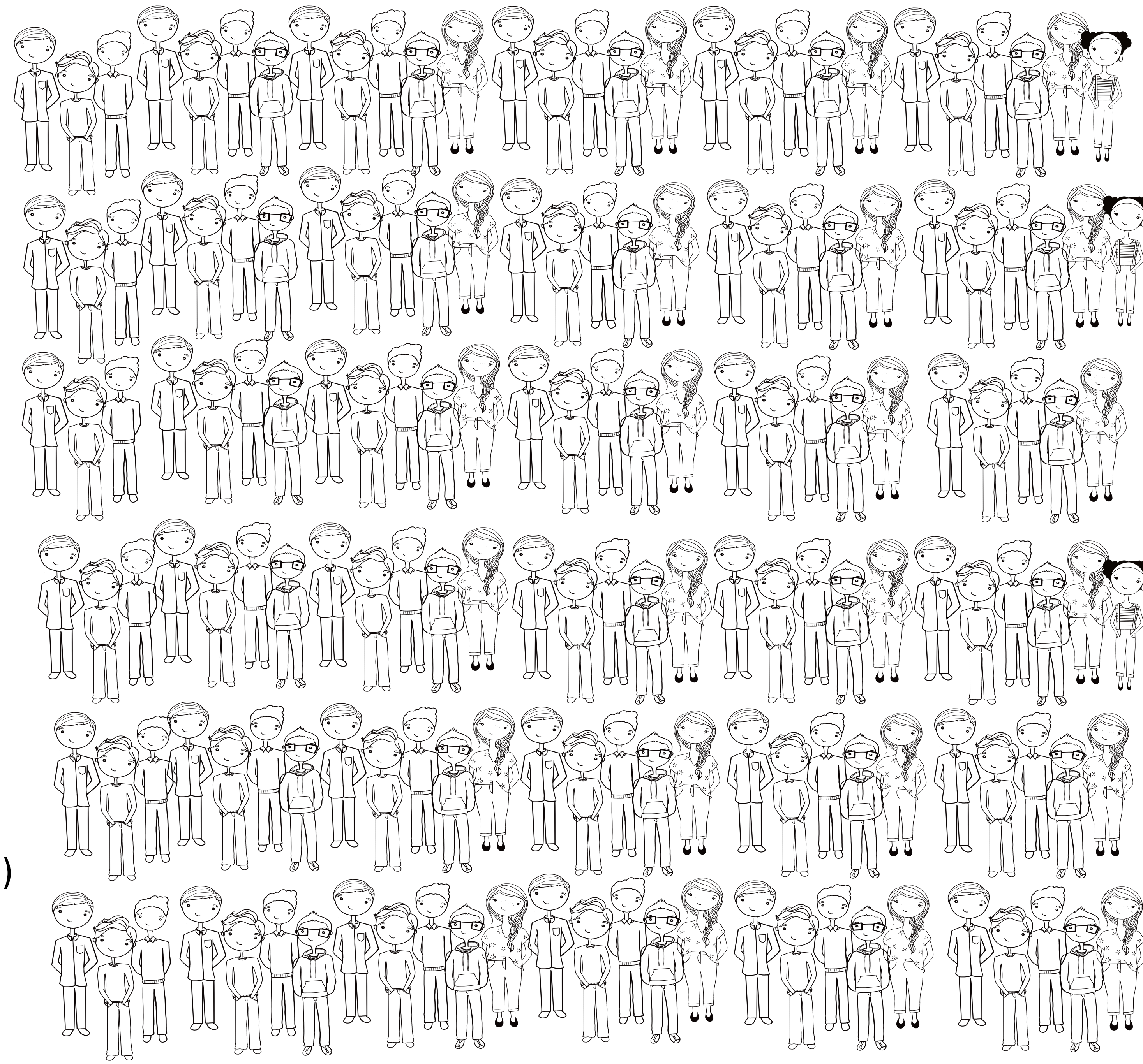
Metric – CS Persistence Intentions

How likely would you be to do the following:

	Highly unlikely	Moderately unlikely	Slightly unlikely	Neither likely nor unlikely	Slightly likely	Moderately likely	Highly likely
Take another course in **Computer Science**	○	○	○	○	○	○	○
Minor in **Computer Science**	○	○	○	○	○	○	○
Major in **Computer Science**	○	○	○	○	○	○	○
Apply to graduate programs requiring high levels of **Computer Science** ability	○	○	○	○	○	○	○
Apply for high-paying jobs requiring high levels of **Computer Science** ability	○	○	○	○	○	○	○

Population

- CS1 for non-majors
- Introduction to Computing – MATLAB
- Two (non-pandemic) semesters
- Top-performing Students (top 50% exam 1 score)
  - Two semesters
  - 160 men
  - 33 women



Results

- The intervention increased both women’s and men’s self-assessed CS ability

- Women self-assess their CS abilities to be lower than Men’s by 10%. The intervention does not decrease this gap.

- The intervention increased women’s CS persistence intentions by 18%. It did not increase men’s CS persistence intentions

Conclusions

- Preliminary evidence that giving students explicit feedback about their CS performance can increase women's self-assessments of ability and CS persistence intentions
  - Key limitation is small sample of women
  - While this does not solve the underrepresentation of women in STEM fields, this is a lightweight intervention that could keep more high-ability women in the STEM pipeline

Lingering Questions

- What is it about the intervention that worked so well?
  - That it was an encouraging email sent by the instructor?
  - That it was positive, granular feedback given by the instructor?
  - That it was positive, granular feedback in general?
  - …

Intervention – Part 2

Exam 1 Feedback External Inbox ×

Katie Stolee
to me

11:55 AM (0 minutes ago)

Dear Kathryn,

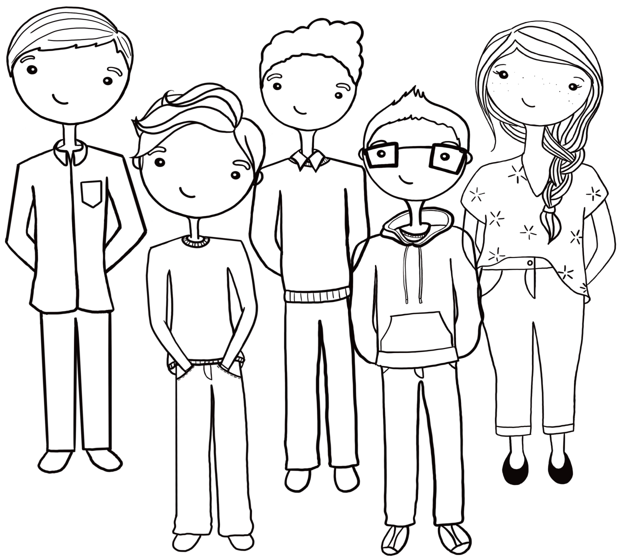
You got a 62% on the test. Remember that average grades in STEM courses tend to be lower than in other university classes and that many people do not perform well on their first computer science test.

I also believe that if you put in the time and work hard, you can improve the grade on your next test. Research shows that passion, dedication, and self-improvement – and not simply innate talent – are the road to genius and contribution. Indeed, research finds that on average, students who excel at STEM courses spend more time and energy preparing for class, studying, and trying to improve themselves.

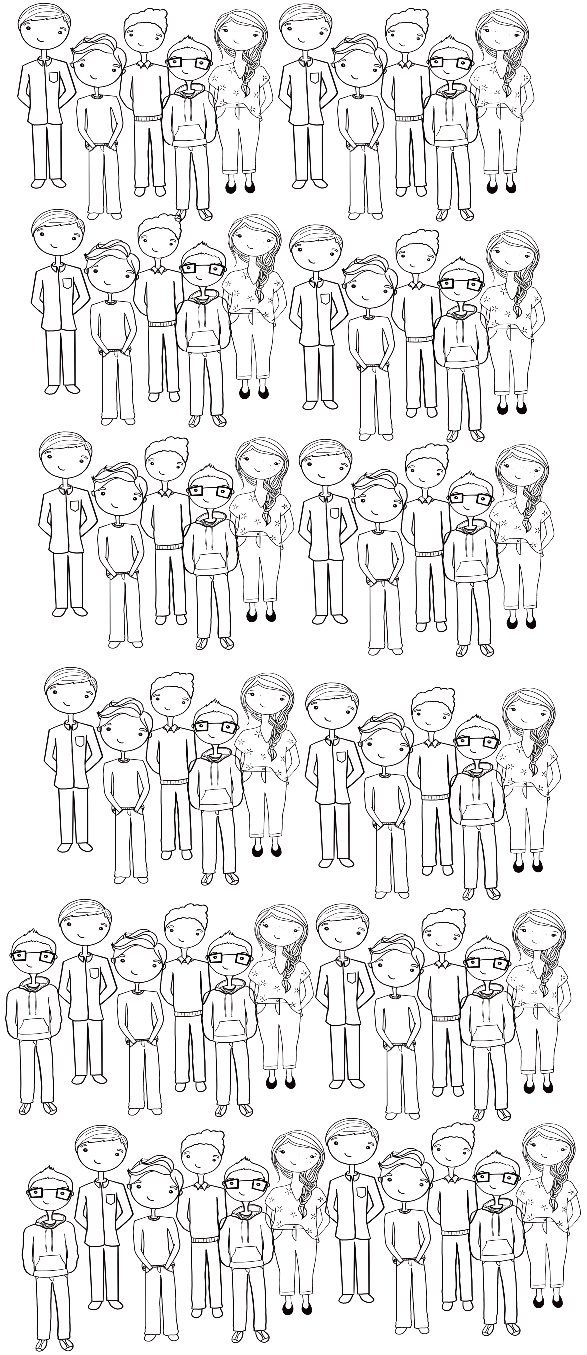
Here are some resources that might be helpful to you:
- *Office Hours (led by TAs and) and Study Hours (led by Study Hour leaders). Specific times and locations are posted on the Google Calendar on Moodle.*
- *Video of the Lectures (recorded by < Pro f essor > for Engineering Online, requires University login)*
- *The online interactive textbook*

Again, I believe that you have what it takes to be successful in this course if you work hard. Please reach out at any time if there is anything else I can do to help you succeed.

Population



Bottom-performing Students (65 total)
- Only the first semester



Results

- No change in self-assessed abilities
- No change in persistence intentions

Lingering Questions

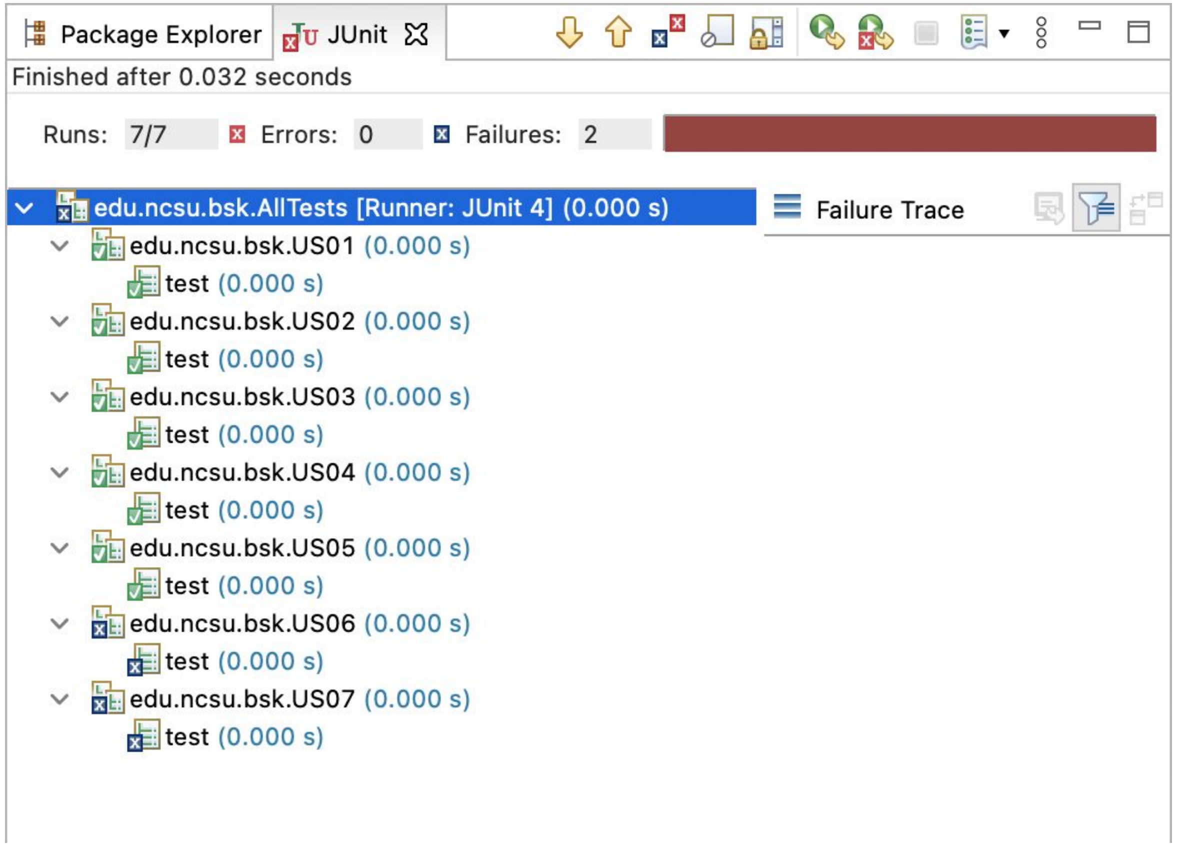
- What is it about the intervention that worked so well?
  - That it was an encouraging email sent by the instructor?
  - That it was positive, granular feedback given by the instructor?
  - That it was positive, granular feedback in general?
  - …

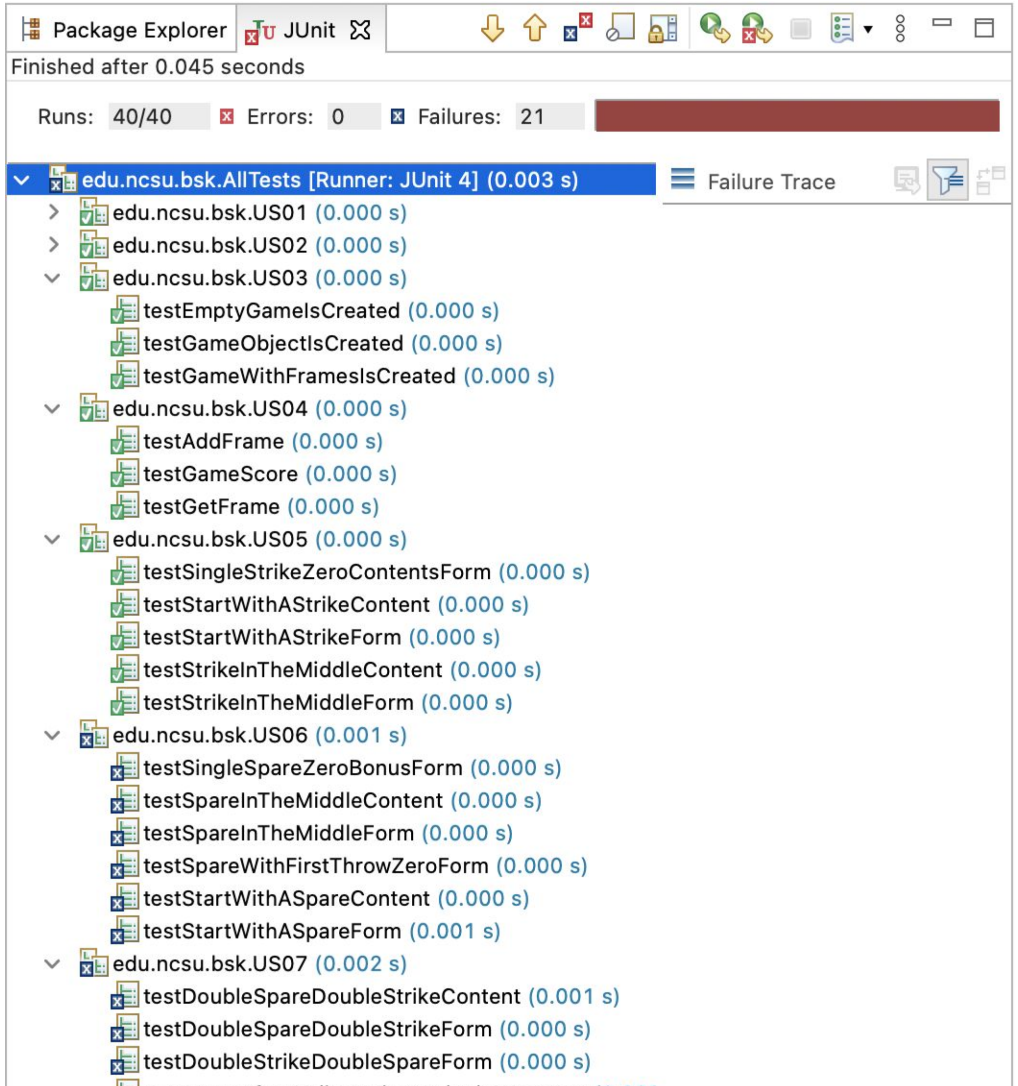
Lingering Questions

- What is it about the intervention that worked so well?
  - ~~That it was an encouraging email sent by the instructor?~~
  - That it was positive, granular feedback given by the instructor?
  - That it was positive, granular feedback in general?
  - …

How else do CS students receive feedback?

An excursion into software testing





How does the JUnit test structure impact persistence intentions?

Study

- TDD project in Eclipse IDE. Given tests, write code to make them pass.
- 90 minute lab session
- Pre- and post- surveys with persistence questions
- A/B study: multiple-assertion and single-assertion
- 33 total students:
  - 19 graduate students
  - 14 undergraduate sophomore students

RQ1: Do more granular test cases assist students better in implementing programs?

Nope, no differences in code quality

RQ3: Do more granular test cases improve students’ persistence intentions in CS?

Nope, no impact on self-assessed ability or persistence

Why the non-result?

Possibilities:

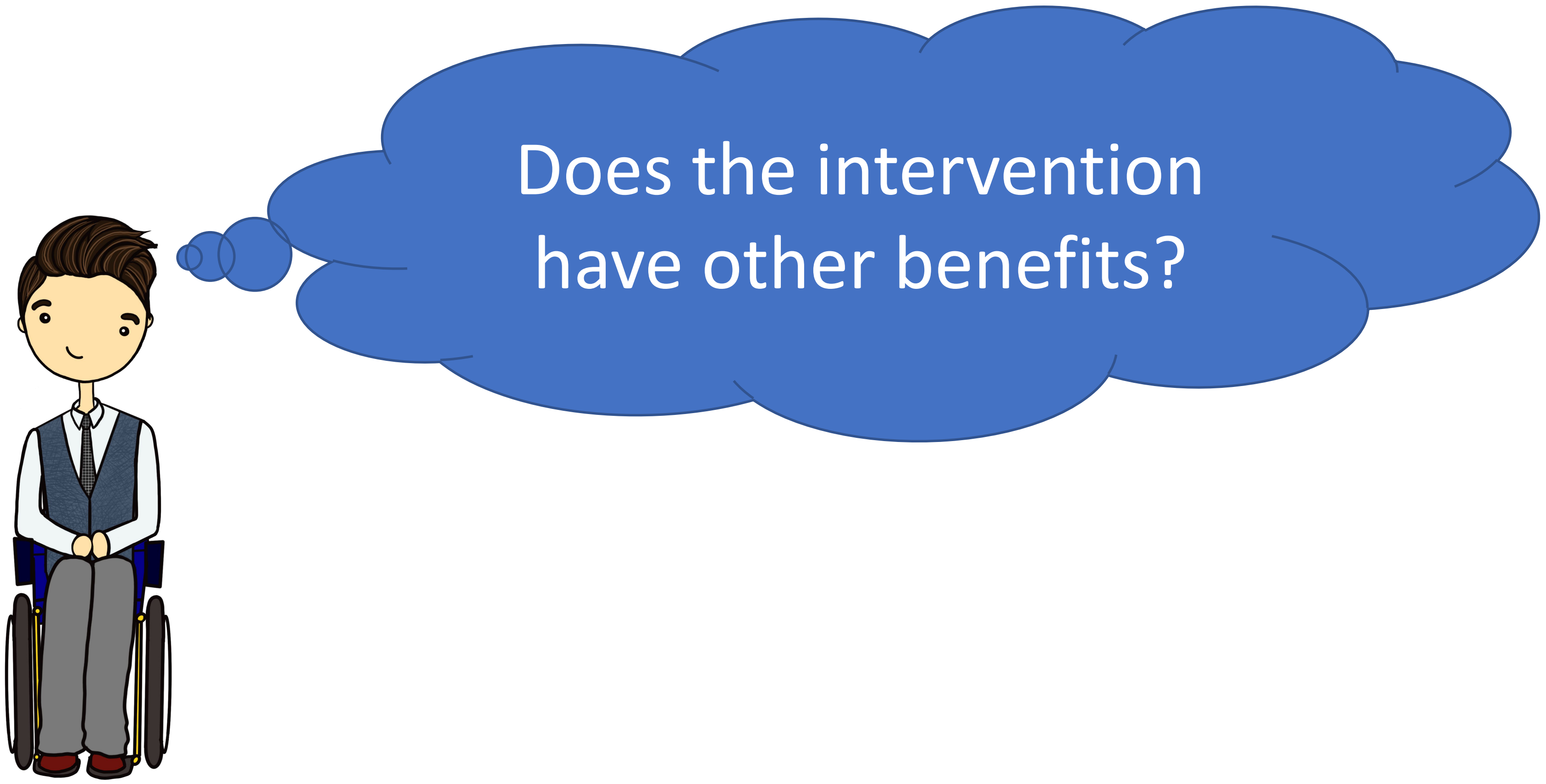
- Small study (33 participants)
- Testing feedback is *temporary* while grades are *lasting*
- Credibility of feedback source*: instructor vs. automated system*
- The students were *already committed* to their degree program

Lingering Questions

- What is it about the intervention that worked so well?
  - ~~That it was an encouraging email sent by the instructor?~~
  - That it was positive, granular feedback given by the instructor?
  - That it was positive, granular feedback in general?
  - …

Lingering Questions

- What is it about the intervention that worked so well?
  - ~~That it was an encouraging email sent by the instructor?~~
  - That it was positive, granular feedback given by the instructor?
  - ~~That it was positive, granular feedback in general?~~
  - …


Does the intervention
have other benefits?

Context

- 23.2% of Computer Science faculty in the US are Women [Taulbee Survey 2020]
- Student Evaluations of Teaching (SETs) are common for evaluating faculty
- Research shows gender bias in SETs

Research Questions

- **RQ1:** Does the intervention **increase students' perceptions of the woman professor's likability**?

- **RQ2:** Does the intervention **increase SETs for a woman professor**?

Study

CS1 course for engineering students (non-majors)

Fall 2018 - 185 students

1) **Field Experiment - causal data**

INTERVENTION

2) **SETs - observational data**

following semester

Spring 2019 - 264 students

SETs - observational data

same female instructor

Metric – instructor likability

How much do you like the instructor of this class?

- ○ Greatly dislike
- ○ Dislike
- ○ Somewhat dislike
- ○ Neither like nor dislike
- ○ Somewhat like
- ○ Like
- ○ Greatly like

Metric – Student Evaluations of Teaching

	Question
1	The **instructor's** teaching aligned with the course's learning objectives/outcomes
2	The **instructor** was receptive to students outside the classroom
3	The **instructor** explained material well.
4	The **instructor** was enthusiastic about teaching the course
5	The **instructor** was prepared for class
6	The **instructor** gave useful feedback.
7	The **instructor** consistently treated students with respect
8	Overall, the **instructor** was an effective teacher
9	The course materials were valuable aids to learning
10	The course assignments were valuable aids to learning
11	This course improved my knowledge of the subject
12	Overall, this course was excellent
Average (all questions equally weighted)	

RQ1 - Likability

- **Intervention increases top-performing student ratings of professor likability** by .33 points ($p < 0.05$), 5.8% increase
- **Intervention did not increase bottom-performing students professor likability ratings**

RQ2 – SETs

- **SETs were significantly higher in the intervention semester**, despite the fact that the **intervention was given to only half the students**.
- Biggest differences were seen in **Qs related to the professor, her overall effectiveness, the usefulness of her feedback, and treatment of students**.

DISCLAIMER

- Women should not have to alter their behavior to correct for the gender biases of others $\longrightarrow$ instead we need institutional change
- We hope that this intervention may prove to be a valuable, ”survival strategy,” for women working to be successful in computing education.
- Indeed, many women already report adjusting their behavior to take into account the gender biases of observers [Williams, 2018]

Summary of Major Findings

- A simple email intervention significantly increased the CS persistence intentions of top-performing women

- A simple email intervention significantly increased the student evaluations of teaching for a woman professor

Takeaway Messages

- As educators, we do more than teach material: we are shaping students' career ambitions
- Efforts to increase retention in computing often focus on improved learning outcomes, but *beliefs* about ability are typically more predictive of persistence
- Our feedback has a powerful influence on students' self-assessments and persistence

Something fun – diverse clipart!

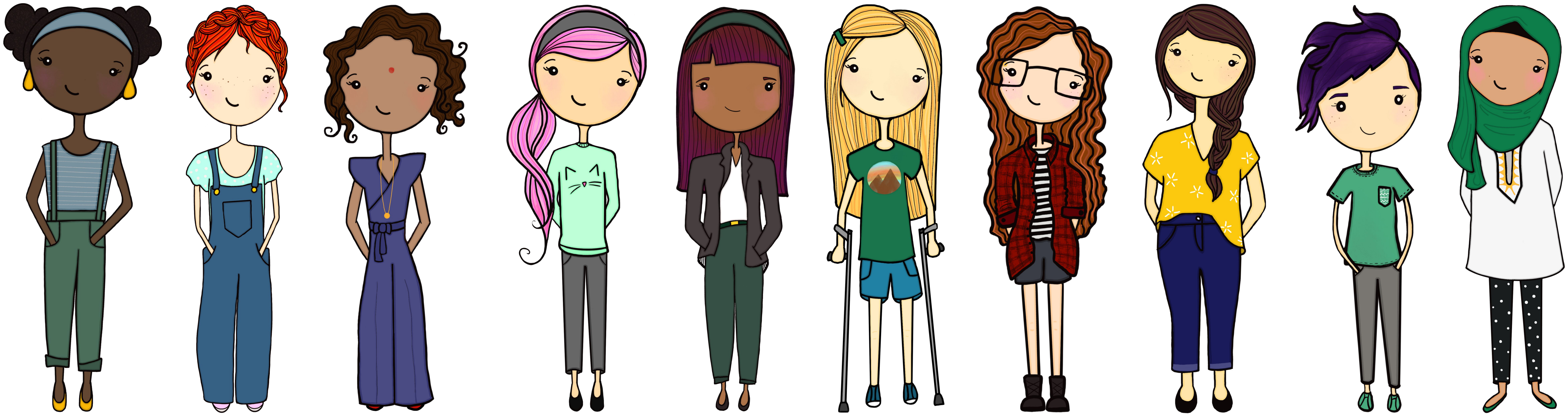


Table 1: Means and Standard Deviations of Computing Ability, Self-Assessed CS Ability, and CS Persistence Intentions (Together and by Gender). Mean and (Std Dev) presented.

	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)

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)

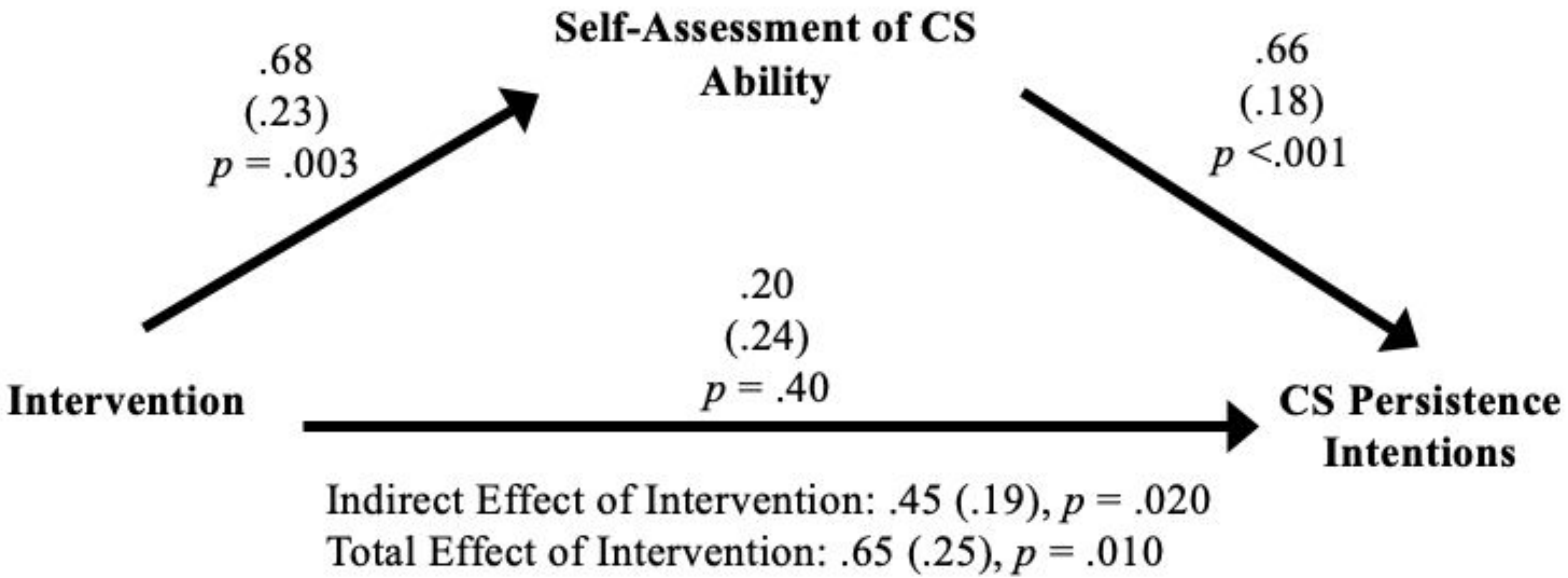
†$p \leq .10$ *$p \leq .05$ **$p \leq .01$ ***$p \leq .001$.
n = 386 observations from 193 students

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)

$^{\dagger}p \leq .10$ $^{*}p \leq .05$ $^{**}p \leq .01$ $^{***}p \leq .001$.

n = 386 observations from 193 students


Self-Assessment of CS Ability
.68
(.23)
p = .003
.66
(.18)
p <.001
.20
(.24)
p = .40
Intervention
CS Persistence Intentions
Indirect Effect of Intervention: .45 (.19), p = .020
Total Effect of Intervention: .65 (.25), p = .010

	Question	Intervention (Fall 2018) Mean	stdev	n	Control (Spring 2019) Mean	stdev	n	t-statistic	p-value
1	The **instructor's** teaching aligned with the course's learning objectives/outcomes	4.40	0.61	80	4.09	0.76	103	-3.10	0.002 **
2	The **instructor** was receptive to students outside the classroom	3.96	0.85	71	3.60	0.98	98	-2.51	0.013 **
3	The **instructor** explained material well.	3.90	0.89	80	3.63	1.08	103	-1.85	0.067 †
4	The **instructor** was enthusiastic about teaching the course	4.29	0.70	80	4.21	0.75	103	-0.69	0.492
5	The **instructor** was prepared for class	4.31	0.70	80	4.20	0.68	103	-1.05	0.294
6	The **instructor** gave useful feedback.	3.94	0.90	78	3.61	1.08	99	-2.22	0.028 *
7	The **instructor** consistently treated students with respect	4.24	0.75	79	3.95	1.00	102	-2.22	0.028 *
8	Overall, the **instructor** was an effective teacher	4.00	0.86	79	3.63	1.08	103	-2.57	0.011 *
9	The course materials were valuable aids to learning	3.90	1.09	80	3.91	1.00	102	0.08	0.940
10	The course assignments were valuable aids to learning	4.38	0.86	80	4.14	0.97	100	-1.71	0.088 †
11	This course improved my knowledge of the subject	4.44	0.71	78	4.36	0.74	102	-0.67	0.504
12	Overall, this course was excellent	3.70	1.00	79	3.87	1.04	102	-1.16	0.248
Average (all questions equally weighted)		4.13	0.28	12	3.92	0.22	12	-5.84	< 0.001 ***