CrossMark

Stereotype Threat and Gender Differences in Biology

Gita Taasoobshirazi[1] · Christina Puckett[1] · Gwen Marchand[2]

Abstract

Stereotype threat theory (STT) presents a potential explanation for differences in achievement in math and science for women. Specifically, STT postulates that the perceived risk of confirming a negative stereotype about an individual's identity group acts as a psychological burden that negatively impacts performance. This study is the third in a series of studies examining the impact of ST on gender differences in science. The present study tested the impact of stereotype threat (ST) on gender differences in biology achievement, self-efficacy, motivation, and domain identification using a four-group, quasi-experimental design. 83 introductory-level college biology students were randomly assigned to 1 of 4 ST conditions including an explicit ST condition, an implicit ST condition, a reverse ST condition, and a nullified ST condition. Results indicated that there were no gender differences by ST condition; however, overall, the women identified more with the field.

Keywords Biology · Domain identification · Gender differences · Self-efficacy · Stereotype · Threat

Introduction

There is a need for nearly one million additional university graduates in Science, Technology, Engineering, and Mathematics (STEM) fields, according to the President's Council of Advisors on Science and Technology (PCAST) report (Olson & Riordan, 2012). Over the last decade there has been much concern regarding the shortage of STEM employees needed to meet the demands of the labor force in professions such as engineering, biotechnology, nanotechnology, software

✉ Gita Taasoobshirazi
gtaasoob@kennesaw.edu

[1] Department of Statistics and Analytical Sciences, Kennesaw State University, 1000 Chastain Rd, Kennesaw, GA 30144, USA

[2] Educational Psychology and Higher Education, University of Nevada, Las Vegas, Las Vegas, NV, USA

⚠ Springer

development, and data science. There is also concern regarding a shortage in the STEM work force needed to sustain U.S. innovation enterprise, advances in medical, engineering, and space technology, global competitiveness, and the national security necessary for economic and social progress (Xue & Larson, 2015). Increasing the supply side of the STEM pipeline has been a significant challenge (Wang, 2013). This pipeline has repeatedly been referred to as leaky as it is frequently unable to retain students from secondary school all the way to STEM careers, and a large number of the students lost from the pipeline are women (Blickenstaff, 2005).

Although more women are attending public and private colleges in the United States (Rocheleau, 2016) and are pursuing degrees in the sciences compared to previous decades, they are still underrepresented in many of the sciences and in more advanced science degrees (Hill, Corbett, & St Rose, 2010; National Science Foundation (NSF), 2015). A multitude of explanations have been presented to shed light on the gender disparity in science achievement and participation including: male and female differences in motivation, differences in strategy use, lack of role models for women, and differences in parent and teacher attention and encouragement (e.g. Desouza & Czemiak, 2002; Enman & Lupart, 2000; Greene & DeBacker, 2004; Mattern & Schau, 2002; She, 2001; Shin & McGee, 2002; Tenenbaum & Leaper, 2003). A recent review of explicit and implicit gender stereotypes in science across 66 countries found that even when explicit gender stereotypes were weak, implicit stereotypes were associated with reduced participation of females in science in higher education (Miller, Eagly, & Linn, 2015). Stereotype threat (ST), or when a situation poses a risk by which one's behavior could be interpreted as confirming a negative stereotype about a person's social group, has also been hypothesized to contribute to gender differences in science achievement and participation (Marchand & Taasoobshirazi, 2013; Smith, 2004; Steele, 1997). Although there is substantial research on the impact of ST on the performance of women in mathematics (Johns, Schmader, & Martens, 2005; O'Brien & Crandall, 2003; Schmader, 2002; Spencer, Steele, & Quinn, 1999), comparatively fewer studies have examined the role of ST on gender differences in science (Marchand & Taasoobshirazi, 2013).

Recent research on gender disparities in STEM fields has provided a more nuanced view of science that has not been disentangled within the stereotype threat literature. Cheryan, Ziegler, Montoya, and Jiang (2017) present a sophisticated review of female underrepresentation between different STEM fields. These researchers make the point that females are no longer underrepresented at the level of undergraduate majors and attainment in the fields of biology, chemistry, and mathematics compared to engineering, computer science, and physics. They present a model to explore the differences in gender representation within STEM that includes three primary explanatory components: culture within fields, differences in early experiences for females in fields more dominated by males, and differences in self-efficacy amongst gender groups in different fields. Their review finds that taking this more comprehensive view on gendered participation amongst fields honors the complexity of understanding why it is that some fields are less likely to recruit and retain females (see also Wang & Degol, 2017). Recent empirical work comparing high school students' intentions to major in STEM fields lent additional support for the model proposed by Cheryan et al. (2017), finding that localized contexts effected intentions to persist in engineering and computer science, with negative influence stemming from biased male peers and positive influence from confident female peers, but that in biological and physical science, such contextual factors were not instrumental in

persistence decisions with only perceived confidence in one's own science ability predicting persistence intentions (Riegle-Crumb & Morton, 2017). These researchers note that peers may exert a more negative influence in fields that are not normative for females to enter (Riegle-Crumb & Morton, 2017), though those more normative fields may still present challenges for female participation (Cheryan et al., 2017). For example, recent research has found that females underperformed on science exams in introductory biology courses compared to their male peers (Ballen, Wieman, Salehi, Searle, & Zamudio, 2017; Cotner & Ballen, 2017). In one study, mediation analyses found that test anxiety had a negative impact on performance for female students only and topic interest had a positive effect for females (Ballen et al., 2017). These studies did not directly test stereotype threat, but reflect cues that might trigger stereotypes in STEM fields (e.g. Riegle-Crumb & Morton, 2017) and outcomes that may reflect stereotype threat processes at work (e.g. Ballen et al., 2017).

The present study is the third in a series of studies examining the influence of ST on gender differences in science, with the intent to discover if there is also nuance in stereotype threat effects across disciplines within science that is concordant with the research on differences in gender representation amongst the sciences. Our first study (Marchand & Taasoobshirazi, 2013) examined the impact of ST in high school physics and compared men and women who were randomly assigned to one of three ST conditions including an explicit condition (men do better than women), implicit condition (no instructions regarding gender and performance), and nullified condition (no gender differences on the test). Results showed that the men outperformed the women on a set of physics problems in the implicit and explicit ST conditions, but that men and women performed similarly in the nullified condition. This suggested that simply being in a typical physics testing situation was enough to compromise women's performance, but that reminding students that both men and women are capable of doing well in physics removed any negative effects to women. This first study was conducted in the domain of physics where gender differences in achievement and participation are the largest of any of the sciences (NSF, 2017). For example, in 2014, only 23% of the master's degrees and 19% of doctoral degrees in physics were awarded to women (NSF, 2017).

Our second study was in the domain of chemistry (Sunny, Taasoobshirazi, Clark, & Marchand, 2017). The data on degrees awarded in chemistry suggested a closing of the gender gap, especially for less advanced degrees. For example, although women only hold 39% of the doctorates in chemistry, they hold nearly half of the bachelor's degrees awarded in chemistry (Matson, 2013). We wanted to know whether ST plays the same role in chemistry as in physics given that the comparatively less prominent gender disparity in chemistry. Our study in chemistry examined the impact of stereotype threat (ST) on gender differences in chemistry achievement, self-efficacy, and test-anxiety using a four-group, quasi-experimental design. One hundred fifty-three introductory-level college chemistry students were randomly assigned to one of four ST conditions including an explicit ST condition (men do better than women), an implicit ST condition (no instructions regarding gender and performance), a reverse ST condition (women do better than men), and a nullified condition (no gender differences on the test). This was the first study to examine ST using four experimental conditions. Results indicated that there were no gender differences by ST condition; however, overall, the men had higher self-efficacy and lower test-anxiety than the women. These differences in motivation, however, did not translate to differences in performance. The

results of our study in chemistry suggested that the negative effects of ST found in chemistry were not nearly as problematic as they were in physics.

Our third and present study examines ST in biology, where the gender discrepancy in enrollment favors women. For example, more than half of undergraduate biology majors are female (Leander, 2014), and 2008 was the first year that women earned more doctorate degrees in biology than men (Matson, 2013). In 2014, 58% of the bachelor's degrees, 56% of the master's degrees, and 52% of the doctoral degrees in the biosciences were awarded to women (NSF, 2017). Studying ST in biology is important given that the field of biology is shifting rapidly in terms of gender membership. Because ST researchers have speculated that role models and an increased sense of belonging mitigate the impacts of ST (Steele, 1997), fields like biology, where the gender discrepancy in enrollment favors women, offer unique testing grounds to better understand the impact of ST on achievement. Current research using survey methodology in a naturalistic setting with undergraduate females found that perceptions of stereotype threat were stronger in male-dominated physics courses as opposed to female-dominated biology courses. Further, women self-reported higher domain identification when they perceived lower stereotyped threat, an effect associated with more perceived opportunities for communal utility value in less stereotyped domains (Smith, Brown, Thoman, & Deemer, 2015). We questioned whether ST effects in biology would disappear or work in the opposite direction (against men). The present study is the first to experimentally examine the impact of ST in biology, though research has investigated gender differences in performance on tests in undergraduate biology courses (Ballen et al., 2017; Cotner & Ballen, 2017). It is also the first to use four experimental conditions in biology including an explicit ST condition (students are told men outperform women on the biology test), an implicit ST condition (students are not provided any information about the effect of gender on performance, but are in a traditional testing situation), a reverse ST condition (students are told women outperform men on the biology test), and a nullified condition (students are told that no gender differences in performance have been found on the biology test). We examined students' performance on a set of biology problems, motivation, domain identification, and study-condition specific self-efficacy across the four conditions.

In our previous publications on ST, we discuss the extant research on ST and ST in science (Marchand & Taasoobshirazi, 2013; Sunny et al., 2017). In order to provide the reader with some background on ST and ST in science, this research is also presented in the section below.

Stereotype Threat in Science

The origins of stereotype threat research can be traced to the studies of racial differences in performance on standardized tests (Steele & Aronson, 1995). The idea that stereotypes held about a particular group may create psychologically threatening situations associated with fears of confirming judgment about one's group, and in turn, inhibit learning and performance (Johnson, Barnard-Brak, Saxon, & Johnson, 2012), has since been extended to explore a variety of gender and racial group differences across domains such as chess, athletics, and mathematics (Schmader, 2002; Smith, 2004; Spencer et al., 1999).

The research on the impact of ST on gender differences in science has been scarce (Steele, 1997). Much of what we know about gender and ST comes from the research in mathematics. Several groupings of experimental or quasi-experimental conditions have been used to test the impact of ST on gender disparities in mathematics. ST has been studied in conditions where a test is described as diagnostic (e.g. you are taking a math test) or non-diagnostic (e.g. this is a problem solving task) (Johns et al., 2005; Kiefer & Sekaquaptewa, 2007). ST has also been studied in conditions where the threat is made implicit (e.g. just being in an everyday mathematics testing situation), explicit (e.g. students are told men perform better than women on a test), or nullified (e.g. equating the groups) (Smith & White, 2002). Most commonly, implicit or explicit ST conditions are compared with a nullified condition (e.g. O'Brien & Crandall, 2003). For example, Spencer et al. (1999) compared college-level men and women's mathematics performance across two ST conditions. Students were told, prior to taking a mathematics test, that there were no gender differences on the test (a nullified condition) or were given no information regarding gender differences on the test (implicit ST condition). Results indicated that the men outperformed the women in the implicit ST condition, but that gender differences disappeared in the nullified ST condition. These findings suggested that women underperformed because of the existing and implicit stereotype that women are less capable than men in mathematics (O'Brien & Crandall, 2003; Quinn & Spencer, 2001).

Although studies in the domain of mathematics inform our understanding of gender and ST in the sciences, studies of ST in science have been scarce. Further, given that contextual cues, such as the gender of peers, gender-exclusive language, or the physical environment may trigger stereotype threat effects (Master, Cheryan, Moscatelli, & Meltzoff, 2017) and may also differ amongst STEM fields and in particular, science disciplines, it may be inappropriate to generalize gender ST effects as "STEM" effects. A thorough review of the research led to the retrieval of approximately four empirical studies examining the impact of ST on gender differences in science performance across ST conditions. Two of those studies were completed by our research team and discussed above (ST in physics and chemistry) (Marchand & Taasoobshirazi, 2013; Sunny et al., 2017). One study was conducted in engineering and the other in chemistry. Both are described below.

Good, Woodzicka, and Wingfield (2010) studied how pictures presented to high school chemistry students impacted their performance. The researchers created three conditions where students were given a chemistry text that included pictures of all male scientists, all female scientists, or a mixed group of scientists. They found that the women performed best on the chemistry test in the "all female" picture condition whereas the men performed best in the "all male" picture condition. The men and women performed similarly in the "mixed gender" picture condition.

Bell, Spencer, Iserman, and Logel (2003) assigned university-level engineering students to three ST conditions including a diagnostic condition (the test is measuring engineering aptitude), non-diagnostic condition (test responses are being used to modify and improve the test), and a gender fair condition (test is a measure of aptitude, but men and women have been found to perform equally well on the test). Results indicated that the men outperformed the women in the diagnostic condition but not the non-diagnostic and gender fair conditions. For the women, being exposed to the instructions that the test was measuring engineering aptitude negatively impacted their

performance, suggesting an existing, implicit threat that women have less ability than men in engineering.

The research on ST conditions on gender differences has been limited in examining factors beyond performance. In fact, it is only recently that researchers have considered the moderating and mediating cognitive, psychological, emotional, and motivational mechanisms by which ST may affect performance (Schmader, Johns, & Forbes, 2008; Shapiro, 2011). For example, current research has examined how goal orientation (Brodish & Devine, 2009; Deemer, Smith, Carroll, & Carpenter, 2014; Finnigan & Corker, 2017), test-anxiety (Brodish & Devine, 2009), and domain identification (Steinberg, Okun, & Aiken, 2012) mediate the effects of ST threat on performance; other research has shown that the effects of ST depend on degree of personal identification with the stereotype (Nguyen & Ryan, 2008). For example, students who identify with a given domain are more affected by ST than students who do not (Aronson et al., 1999; Keller, 2007; Steele, 1997). With a few exceptions (e.g. Sunny et al., 2017), the emerging research on ST and performance across multiple ST conditions has failed to examine the role of non-cognitive variables. The present study considers the impact of gender and ST when self-efficacy, motivation, and domain identification are both outcomes and controls.

Method

Participants

Eighty-three introductory level college biology students at a Southeastern university in the United States participated in the study. Thirty-four students were male and 49 were female; approximately 53% were Caucasian, 19% were African-American, 6% were Asian, and 11% were Hispanic. Students were randomly assigned to the four study conditions based on a cluster approach with lab groups being the unit of assignment. Students spent 1 day each week in lab, with approximately 20–25 students in each of four lab groups. The random assignment of ST conditions to lab groups resulted in 21 students in the explicit ST group, 20 students in the reverse ST group, 20 students in the nullified ST group, and 22 students in the implicit ST group. An a priori power analysis using G Power (Faul, Erdfelder, Buchner, & Lang, 2009) with the statistical test *Special Effects and Interactions* indicated that our sample size was larger than what G power recommended for a 2×4 MANOVA with four dependent variables, an effect size f^2 of .16, an alpha of .05, and a power of .80 (G Power recommended total sample size was $n = 52$). We selected the commonly used and recommended power level of .80 to correspond with Cohen's (1988) rationale that studies should be designed so that they have an 80% probability of detecting an effect given there is an effect to detect. This allows for a 20% probability of making a Type II error. In addition, we selected an effect size of .16, which corresponds to a large effect (Cohen, 1988). Unfortunately, and until more recently, the research in science education has been negligent about reporting effect sizes both a priori and post-hoc. However, our selected effect size is consistent with the substantial body of research illustrating large effect sizes for gender differences favoring males on cognitive and individual difference variables in science (e.g. Moss-Racusin, Dovidio, Brescoll, Graham, & Handelsman, 2012). This has found to be true in primary, secondary, and

post-secondary school and across the subjects of physics, engineering, chemistry, and biology (Lauer et al., 2013; Riegle-Crumb & King, 2010).

Measures

Biology Achievement Students were given seven multiple-choice biology problems to solve (Table 1). Questions 1–3, 5, and 6 were developed by a senior lecturer in the biology department who created the lab course curriculum. Questions 2 and 7 were from a test bank from Reece, Taylor, Simon, Dickey and Hogan's (2014) Campbell's Biology text. The seven problems were based on major topics in biology including cellular respiration, photosynthesis, the cell cycle, and cell division, and were scored one point for each correct answer. The professor confirmed that the students had not seen the problems previously, that the problems were at the appropriate level, and that the students had learned the material assessed by the problems.

Self-Efficacy The following seven items were derived from the Motivation Strategies for Learning Questionnaire (MSLQ) (Pintrich, Smith, García, & McKeachie, 1993), revised to focus on biology, and were given to students to assess their study-condition specific self-efficacy:

- Even if the test is hard, I can do it
- I believe I can get an excellent grade on the test
- I believe I have the skills to do well on the test
- I expect to do well on the test
- I am certain I can figure out how to do the most difficult problem on the test
- I can do the problems on this test if I do not give up
- I can do even the hardest problem on this test if I try

Students responded to the items using a 7-point Likert scale that ranged from 1 = "not at all true of me" to 7 = "very true of me." The MSLQ has extensive evidence of reliability and construct validity (Pintrich et al., 1993). For our students, reliability as assessed by Cronbach's alpha was .89.

Motivation The 30 item Biology Motivation Questionnaire was used to assess biology motivation (BMQ) (Glynn, Taasoobshirazi & Brickman, 2009). Composite scores on the inventory have been used to assess students' overall motivation to learn (e.g. Glynn et al., 2009). The 30 BMQ items assess six important components of student motivation including intrinsically motivated biology learning, extrinsically motivated biology learning, relevance of learning biology to personal goals, self-determination for learning biology, self-efficacy in learning biology, and anxiety about biology assessment. Students responded to the 30 randomly ordered items on a 5-point Likert scale ranging from 1 (never) to 5 (always). The five anxiety about biology assessment items were reverse scored when added to the total. Previous findings (Taasoobshirazi & Sinatra, 2011) indicate that the BMQ is reliable ($\alpha = 0.84$), and valid in terms of positive correlations with college students' self-reported grade. The construct validity of the BMQ has also been assessed using exploratory factor analysis

Table 1 Biology Problems

_____ 1. DCIP (2, 6-dichlorophenolindophenol) was used in the Cell Respiration experiment (lab 6) for which of the following?
a. as an artificial electron acceptor
b. to accept electrons from FADH2
c. because the reduced form shows a corresponding color change from blue to colorless.
d. because its reduction can be measured spectrophotometrically.
e. all of the above

_____2. In the cellular respiration lab, cuvettes 1–4 were prepared containing DCIP, P2 suspension, KCN, sodium succinate and the other reagents needed to examine a reaction in the citric acid cycle. What was present in the P2 suspension?
a.Succinate Dehydrogenase
b.BSA
c.Mitochandria
d.A and C
e.None of the above

_____3. In the photosynthesis lab, the property of the photosynthetic pigments that was used to separate them during silica gel chromatography was
a.The wavelength they will absorb
b.color
c.classification of pigment
d.relative polarity
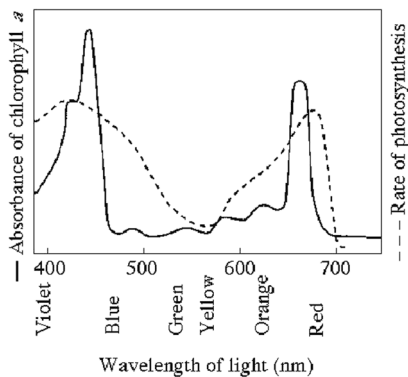
_____ 4. Fig. 10.1 shows the absorption spectra for chlorophyll a and the action spectra for photosynthesis. What wavelength of light in the figure is most effective in driving photosynthesis?
a.475 c. 675
b.420 d. 730

_____ 5. Cells spend approximately 90% of their time in which phase?
a.Mitosis c. Meiosis
b.Dividing d. Interphase

Use the following information to answer the questions below.
A student is looking through his light microscope (~450 X) at a squashed and stained onion root tip. Some, but not all, of the cells have clearly visible chromosome strands.

_____ 6. When a cell is in anaphase of mitosis, which of the following will he see?
a.Individual chromatids separating c. chromosomes clustered tightly at the center
b.Chromosomes clustered tightly at the poles d. formation of vesicles at the midline

_____ 7. In some cells, chromosomes are visible and aligned along the center of the cell. These cells are in which stage of mitosis?
a.telophase c. prophase
b.anaphase d. metaphase

(Taasoobshirazi & Sinatra, 2011). For this study, Cronbach's alpha was found to be $\alpha = 0.89$.

Domain Identification Five items scored on a five-point Likert scale (ranging from 1 = not at all to 5 = very much) from Smith and White's (2001) Domain Identification Measure were revised slightly to focus on the domain of biology and administered to students:

- How much do you enjoy biology?
- How likely would you be to take a job in a biology related field?
- How important is biology to the sense of who you are?
- How important is being a biology student to you?
- How important is it to you to be good at biology?

Construct validity evidence, as assessed by an exploratory factor analysis, has been established by Smith and White (2001) and internal consistency of the items for the present study was Cronbach's alpha = .84.

Procedures

Study materials were administered during the last week of class before the final exam. Students were given a packet with the study materials. The instructions varied across the four conditions in just the following ways:

> *Implicit ST condition*: You will be given seven biology problems to solve. These problems are based on biology material that you may have already covered
> *Explicit ST condition*: You will be given seven biology problems to solve. These problems are based on biology material that you may have already covered. This test has shown gender differences with males outperforming females on the problems
> *Nullified ST condition*: You will be given seven biology problems to solve. These problems are based on biology material that you may have already covered. No gender differences in performance have been found on this test
> *Reverse ST condition*: You will be given seven biology problems to solve. These problems are based on biology material that you may have already covered. This test has shown gender differences with females outperforming males on the problems

Before solving the biology problems and after the ST instructions, students were given the self-efficacy items with the instructions "In order to better understand how you feel about this upcoming biology test, please respond to each of the following statements." In addition, motivation. and domain identification information was collected from students before they were given any instructions, problems, or surveys regarding the biology test questions. Students had approximately 1 hour to complete the survey questions and biology problems. After the students completed the study and packets were collected, they were debriefed about the study.

Results

A 2 × 4 MANOVA was used to determine if differences in student performance, test-specific self-efficacy, motivation, and domain identification differed by gender (two groups: male and female) and ST condition (four groups: implicit ST, explicit ST, nullified ST, and reverse ST). Prior to analysis, we assessed the data for homogeneity of variance (Leven's test p values for all four dependent variables were greater than .05), that the observed covariance matrices for the dependent variables are equal across groups (Box's M p value = .27), and for normality of the data (skewness and kurtosis statistics for each dependent variable by each gender was less than absolute value of two); these results indicated that all assumptions for MANOVA were met. Results of the MANOVA indicated a significant effect by gender Wilks' Lambda = .865, F = 2.78, p =.03, partial eta squared = .14. Neither the ST condition effect nor the interaction between ST condition and gender were significant; a non-significant test result as part of the MANOVA indicates no effect and thus no differences by ST condition on the dependent variables and no gender differences across the ST conditions on the dependent variables (Johnson & Wichern, 2004). There was a main effect for gender on domain identification (p = .004, partial eta squared = .11) with females showing greater identification (M = 19.03) with biology than the males (M = 17.09). When domain identification, self-efficacy, and motivation were included as covariates instead of dependent variables, ST effects on the biology problems were still non-significant. Tables 2 and 3 report descriptive for men and women for the four dependent variables across the four ST conditions.

Discussion

This is our third study in a series of studies that examines the role of ST in science, with a focus on different science domains. This is the first study be done in biology and the first to compare four ST conditions on biology achievement, self-efficacy, motivation, and domain identification. Results indicated that there were no differences by ST condition or difference for ST condition by gender; rather we found a significant main effect for gender with women reporting greater domain identification with biology.

Table 2 Descriptive statistics for women across groups for the four dependent variables

	Implicit ST condition			Explicit ST condition			Nullified ST condition			Reverse ST condition		
	Mean	SD	n	Mean	SD	n	Mean	SD	n	Mean	SD	n
Performance	4.00	.95	12	3.76	1.39	17	3.30	1.25	10	4.33	.71	9
Self-Efficacy	3.33	11.36	12	33.23	9.43	17	38.10	8.41	10	38.56	5.48	9
Identification	55.67	9.81	12	57.41	12.25	17	62.60	11.16	10	64.33	8.11	9
Motivation	112.50	19.72	12	108.35	17.37	17	120.80	13.77	10	117.44	15.43	9

Table 3 Descriptive statistics for men across groups for the four dependent variables

	Implicit ST condition			Explicit ST condition			Nullified ST condition			Reverse ST condition		
	Mean	SD	n	Mean	SD	n	Mean	SD	n	Mean	SD	n
Performance	3.30	1.33	10	4.75	.96	4	4.10	1.20	10	4.20	1.40	10
Self-Efficacy	34.25	3.03	10	42.00	5.48	4	35.80	5.27	10	37.40	5.19	10
Identification	50.90	11.53	10	55.00	8.28	4	53.40	10.25	10	50.20	12.26	10
Motivation	105.90	12.22	10	117.25	6.60	4	109.10	9.42	10	107.30	20.93	10

Implications

Our results indicated that the ST instructions did not impact the students' performance on biology problems. Thus it appears that the negative effects of ST about females that are found in physics, engineering, and mathematics, all male-dominated fields, are not an issue in biology with this sample of undergraduate students. This study adds to the discussion of gender balance within STEM fields (Cheryan et al., 2017) by providing empirical investigation of possible motivational and performance outcomes when gender disparities within science are extinguished. This study extends a line of research that maps stereotype threat effects against gender proportionality across three different science disciplines; in this latest study, the most broadly gender equitable field, there were few differences between males and females. Given that females are beginning to dominate biology, it might be expected that reverse effects could occur for males, but we also did not find any reverse ST effects with males performing more poorly than women in the reverse ST or implicit conditions.

Our study design did not investigate test-specific anxiety, in which recent research has revealed gender differences in biology exam performance (Ballen et al., 2017). However, in line with models of ST processes and mechanisms (Schmader et al., 2008), we did test for motivational and domain identification effects. Our finding that females had a stronger domain identification with biology than males in this sample is not surprising given current research on domain identification in fields with greater parity amongst males and females (e.g, Cheryan et al., 2017). While we did find that women reported a stronger identification with biology, there was no interaction with stereotype threat condition, in contrast with other STEM research (e.g. Wout, Danso, Jackson & Spencer, 2008). Previous research suggests that moderate levels of domain identification are required for ST threat activation (Nguyen & Ryan, 2008); therefore the greater identification of females with biology in this sample suggests that if the ST effect were present, it would have been prompted. Work by Smith et al. (2015) found that females perceived less stereotype threat in female-dominated biology courses and, concurrently, higher domain identification. Based on our findings and this previous work, although not directly tested, a plausible story emerges that when gender parity exists in science disciplines, females develop greater domain identification on a systematic level, thus weakening possible stereotype threat effects in performance situations to be undetectable or not present. The experimental manipulation of stereotype threat conditions in performance situations and concurrent investigations of domain identification extends the current research on stereotype threat effects and gender broadly, while also adding

to the unfolding knowledge base on the possible positive effects of gender parity within the sciences. Future research that combines experimental stereotype threat performance manipulations in biology with perceived stereotype threat, motivational variables, and domain identification with a more robust and diverse sample would be a substantial benefit to further disentangling when stereotype threats are present, how they develop, and for whom.

Despite the lack of ST effects related to gender in these biology courses, it's possible that local contexts could still create conditions by which gender differences may be prompted. For example, persistent negative gender stereotyped messages from an instructor may lead to the development of a threat. Our study only looked at a single point in time and the messages that were delivered by a researcher with whom students did not have a relationship and who was unrelated to the course. Moreover, it's not clear how students interpret their discipline specific experiences, such as biology, within a broader domain, such as science. Our study did not find reverse ST effects for men. One plausible explanation is that males are still in a privileged position within the science domain; thus they may interpret contextual cues, such as proportionality in the immediate context, differently than students who are members of historically underrepresented groups within STEM fields.

Limitations and Cautionary Notes

This study is not without limitations that should be considered when interpreting the findings. Although the study was sufficiently powered to examine ST effects across conditions, the sample size was relatively small, particularly within conditions. We recommend ST in biology continue to be explored using large sample sizes. For example, research in science education and on ST has recently begun to explore how motivational and cognitive variables exasperate or ameliorate the impact of ST. Some of the variables implicated in contributing to ST include self-regulation, goal orientation, ST endorsement, and domain-identification (Schmader et al., 2008; Smith, 2004; Steinberg et al., 2012). Path models with multiple cognitive and motivational variables can provide information about how such variables may mediate the ST-performance relationship and how those relationships change with time.

Although a larger sample in this study size may yield a statistically significant effect, the effect is likely to be small and lacking in practical implications. Further, the sample was relatively homogenous and geographically limited. Additional research with a more diverse sample is recommended to determine if there might be gender by ethnicity interactions or subgroups who may in fact be subject to stereotypes related to biological sciences (Smith et al., 2015). We encourage researchers studying ST with larger and diverse samples to confirm the soundness of the psychometrics of the cognitive and individual difference inventories being used in their studies. For example, a confirmatory factor analysis can assess the convergent and divergent validity of individual difference survey items. For cognitive measures in ST and gender studies, differential item functioning may be useful for ensuring that test items are not measuring different abilities for separate subgroups. Finally, due to classroom time constraints, we were unable to pre-test the biology students at the beginning of the semester, prior to the intervention. Pre-testing results between males and females as well as conditions would allow researchers to study whether any pre-existing differences exist across genders and

conditions. For example, if males have lower pre-test performance than females overall, but these gender differences disappeared in the reverse ST condition, this would be a shift that would be very meaningful. Unfortunately, the failure to pre-test is a common design flaw in the research on ST and gender and racial group differences in mathematics and science. We strongly recommend that researchers studying ST include pre-testing into their study design.

Conclusion

This study represents a non-significant ST finding. We would like to point out that null findings are typically interpreted as a true non-effect. Thus for the present study, a conclusion is made that ST effects do not exist in biology. A better explanation is that the data do not support the alternative hypothesis. It is possible that with other (larger) samples, that the null hypothesis would be disproven and significant effects would be present. Additional work on ST in biology would help to determine if the results of the present study hold true.

A positive message from this study is that with gender proportionality could come a remediation of overt and contextual effects that negatively impact performance. While typically such findings are not prevalent in the literature, this study offers a unique contribution to the growing understanding of differences in student experience within the sciences. Such differentiation within the sciences is important to guide policy and practice when funders, researchers, scholars, and policy-makers consider support for the future of STEM fields.

References

Aronson, J., Lustina, M. J., Good, C., Keough, K., Steele, C. M., & Brown, J. (1999). When white men can't do math: Necessary and sufficient factors in stereotype threat. *Journal of Experimental Social Psychology, 35*(1), 29–46.

Ballen, C. J., Wieman, C., Salehi, S., Searle, J. B., & Zamudio, K. R. (2017). Enhancing diversity in undergraduate science: Self-efficacy drives performance gains with active learning. *CBE Life Sciences Education, 16*(4), ar56.

Bell, A. E., Spencer, S. J., Iserman, E., & Logel, C. E. (2003). Stereotype threat and women's performance in engineering. *Journal of Engineering Education, 92*(4), 307–312.

Blickenstaff, C. J. (2005). Women and science careers: Leaky pipeline or gender filter? *Gender and Education, 17*(4), 369–386. https://doi.org/10.1080/09540250500145072.

Brodish, A. B., & Devine, P. G. (2009). The role of performance–avoidance goals and worry in mediating the relationship between stereotype threat and performance. *Journal of Experimental Social Psychology, 45*(1), 180–185. https://doi.org/10.1016/j.jesp.2008.08.005.

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

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, MI: Lawrence Erlbaum Associates.

Cotner S., & Ballen C. J. (2017). Can mixed assessment methods make biology classes more equitable? *PLoS ONE, 12*(12), e0189610. https://doi.org/10.1371/journal.pone.0189610.

Deemer, E. D., Smith, J. L., Carroll, A. N., & Carpenter, J. P. (2014). Academic procrastination in STEM: Interactive effects of stereotype threat and achievement goals. *The Career Development Quarterly, 62*(2), 143–155. https://doi.org/10.1002/j.2161-0045.2014.00076.x.

Desouza, J. M. S., & Czemiak, C. M. (2002). Social implications and gender differences among preschoolers: Implications for science activities. *Journal of Research in Childhood Education, 16*, 175–188. https://doi.org/10.1080/02568540209594983.

Enman, M., & Lupart, J. (2000). Talented female students' resistance to science: An exploratory study of post-secondary achievement motivation, persistence, and epistemological characteristics. *High Ability Studies, 11*(2), 161–178. https://doi.org/10.1080/13598130020001205.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*, 1149–1160.

Finnigan, K. M., & Corker, K. S. (2017). Do different types of stereotype threat moderate the effect of performance-avoidance goals on female math performance? *Journal of Research in Personality, 63*, 36–43.

Glynn, S. M., Taasoobshirazi, G., & Brickman, P. (2009). Science motivation questionnaire: Construct validation with nonscience majors. *Journal of Research in Science Teaching, 46*(2), 127–146.

Good, J. J., Woodzicka, J. A., & Wingfield, L. C. (2010). The effects of gender stereotypic and counter-stereotypic textbook images on science performance. *The Journal of Social Psychology, 150*(2), 132–147. https://doi.org/10.1080/00224540903366552.

Greene, B. A., & DeBacker, T. K. (2004). Goal and orientations toward the future: Links to motivation. *Educational Psychology Review, 16*, 91–120. https://doi.org/10.1023/B:EDPR.0000026608.50611.b4.

Hill, C., Corbett, C., & St Rose, A. (2010). *Why so few? Women in science, technology, engineering, and mathematics*. Washington, DC: American Association of University Women.

Johns, M., Schmader, T., & Martens, A. (2005). Knowing is half the battle teaching stereotype threat as a means of improving women's math performance. *Psychological Science, 16*(3), 175–179.

Johnson, H. J., Barnard-Brak, L., Saxon, T. F., & Johnson, M. K. (2012). An experimental study of the effects of stereotype threat and stereotype lift on men and women's performance in mathematics. *The Journal of Experimental Education, 80*(2), 137–149. https://doi.org/10.1080/00220973.2011.567312.

Johnson, R.A. & Wichern, D.W. (2004). Multivariate analysis. *Encyclopedia of Statistical Sciences, 8*.

Keller, J. (2007). Stereotype threat in classroom settings: The interactive effect of domain identification, task difficulty and stereotype threat on female students' maths performance. *British Journal of Educational Psychology, 77*(2), 323–338.

Kiefer, A. K., & Sekaquaptewa, D. (2007). Implicit stereotypes and women's math performance: How implicit gender-math stereotypes influence women's susceptibility to stereotype threat. *Journal of Experimental Social Psychology, 43*(5), 825–832. https://doi.org/10.1016/j.jesp.2006.08.004.

Lauer, S., Momsen, J., Offerdahl, E., Kryjevskaia, M., Christensen, W., & Montplaisir, L. (2013). Stereotyped: Investigating gender in introductory science courses. *CBE-Life Sciences Education, 12*(1), 30–38.

Leander, S. (2014). More than half of biology majors are women, yet gender gaps remain in science classrooms. *ScienceDaily*. Retrieved December 20, 2015 from https://www.sciencedaily.com/releases/2014/09/140902171429.htm.

Marchand, G.C., & Taasoobshirazi, G. (2013). Stereotype threat and women's performance in physics. *International Journal of Science Education, 35*(18), 3050–3061. https://doi.org/10.1080/09500693.2012.683461.

Master, A., Cheryan, S., Moscatelli, A., & Meltzoff, A. N. (2017). Programming experience promotes higher STEM motivation among first-grade girls. *Journal of Experimental Child Psychology, 160*, 92–106.

Matson, J. (2013). Women are earning greater share of STEM degrees, but doctorates remain gender skewed. *Scientific American*, 308(5). Retrieved from http://www.scientificamerican.com/article/women-earning-greater-share-stem-degrees-doctorates-remain-gender-skewed/.

Mattern, N., & Schau, C. (2002). Gender differences in science attitude-achievement relationships over time among white middle-school students. *Journal of Research in Science Teaching, 39*(4), 324–340. https://doi.org/10.1002/tea.10024.

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

Moss-Racusin, C. A., Dovidio, J. F., Brescoll, V. L., Graham, M. J., & Handelsman, J. (2012). Science faculty's subtle gender biases favor male students. *Proceedings of the National Academy of Sciences, 109*(41), 16474–16479.

National Science Foundation (2015). *Women, minorities, and persons with disabilities*. National Center for Science and Engineering Statistics Directorate for Social, Behavioral and Economic Sciences Report. Retrieved December 22, 2015 from https://www.nsf.gov/statistics/2017/nsf17310/static/downloads/nsf17310-digest.pdf.

National Science Foundation, National Center for Science and Engineering Statistics (2017). *Women, minorities, and persons with disabilities in science and engineering: 2017* (Special Report NSF 17–310). Arlington, VA: Author. Available at www.nsf.gov/statistics/wmpd/.

Nguyen, H. H. D., & Ryan, A. M. (2008). Does stereotype threat affect test performance of minorities and women? A meta-analysis of experimental evidence. *Journal of Applied Psychology, 93*(6), 1314–1134.

O'Brien, L. T., & Crandall, C. S. (2003). Stereotype threat and arousal: Effects on women's math performance. *Personality and Social Psychology Bulletin, 29*(6), 782–789. https://doi.org/10.1177/0146167203252810.

Olson, S. & Riordan, D.G. (2012). *Engage to excel: Producing one million additional college graduates with degrees in science, technology, engineering, and mathematics*. Report to the President. Executive Office of the President.

Pintrich, P. R., Smith, D. A., García, T., & McKeachie, W. J. (1993). Reliability and predictive validity of the Motivated Strategies for Learning Questionnaire (MSLQ). *Educational and Psychological Measurement, 53*(3), 801–813.

Quinn, D. M., & Spencer, S. J. (2001). The interference of stereotype threat with women's generation of mathematical problem-solving strategies. *Journal of Social Issues, 57*(1), 55–71.

Reece, J. B., Taylor, M. R., Simon, E. J., Dickey, J. L., & Hogan, K. A. (2014). *Campbell biology: Concepts and connections* (8th ed.). San Francisco, CA: Benjamin Cummings.

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

Riegle-Crumb, C., & Morton, K. (2017). Gendered expectations: Examining how peers shape female students' intent to pursue STEM fields. *Frontiers in Psychology, 8*, 329–339.

Rocheleau, M. (2016). On campus, women outnumber men more than ever. *Boston Globe*. Retrieved December 20, 2016 from https://www.bostonglobe.com/metro/2016/03/28/look-how-women-outnumber-men-college-campuses-nationwide/YROqwfCPSlKPtSMAzpWloK/story.html.

Schmader, T. (2002). Gender identification moderates stereotype threat effects on women's math performance. *Journal of Experimental Social Psychology, 38*(2), 194–201. https://doi.org/10.1006/jesp.2001.1500.

Schmader, T., Johns, M., & Forbes, C. (2008). An integrated process model of stereotype threat effects on performance. *Psychological Review, 115*(2), 336–356.

Shapiro, J. R. (2011). Different groups, different threats: A multi-threat approach to the experience of stereotype threats. *Personality and Social Psychology Bulletin, 37*(4), 464–480.

She, H. C. (2001). Different gender students' participation in the high-and low-achieving middle school questioning-orientated biology classrooms in Taiwan. *Research in Science & Technological Education, 19*(2), 147–158. https://doi.org/10.1080/02635140120087696.

Shin, N. & McGee, S. (2002). *The influence of inquiry-based multimedia learning environment on specific problem-solving skills among ninth grade students across gender differences*. Paper presented at Association for Educational Communications and Technology, Dallas, TX.

Smith, J. L. (2004). Understanding the process of stereotype threat: A review of mediational variables and new performance goal directions. *Educational Psychology Review, 16*(3), 177–206. https://doi.org/10.1023/B:EDPR.0000034020.20317.89.

Smith, J. L., Brown, E. R., Thoman, D. B., & Deemer, E. D. (2015). Losing its expected communal value: How stereotype threat undermines women's identity as research scientists. *Social Psychology of Education, 18*(3), 443–466.

Smith, J. L., & White, P. H. (2001). Development of the domain identification measure: A tool for investigating stereotype threat effects. *Educational and Psychological Measurement, 61*(6), 1040–1057.

Smith, J. L., & White, P. H. (2002). An examination of implicitly activated, explicitly activated, and nullified stereotypes on mathematical performance: It's not just a woman's issue. *Sex Roles, 47*(3–4), 179–191. https://doi.org/10.1023/A:1021051223441.

Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology, 35*(1), 4–28. https://doi.org/10.1006/jesp.1998.1373.

Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist, 52*(6), 613–629. https://doi.org/10.1037/0003-066X.52.6.613.

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology, 69*(5), 797–811. https://doi.org/10.1037/0022-3514.69.5.797.

Steinberg, J. R., Okun, M. A., & Aiken, L. S. (2012). Calculus GPA and math identification as moderators of stereotype threat in highly persistent women. *Basic and Applied Social Psychology, 34*(6), 534–543. https://doi.org/10.1080/01973533.2012.727319.

Sunny, C. E., Taasoobshirazi, G., Clark, L., & Marchand, G. (2017). Stereotype threat and gender differences in chemistry. *Instructional Science, 45*(2), 157–175.

Taasoobshirazi, G., & Sinatra, G. M. (2011). A structural equation model of conceptual change in physics. *Journal of Research in Science Teaching, 48*(8), 901–991.

Tenenbaum, H. R., & Leaper, C. (2003). Parent–child conversations about science: The socialization of gender inequities? *Developmental Psychology, 39*(1), 34–47. https://doi.org/10.1037/0012-1649.39.1.34.

Wang, M. T., & Degol, J. L. (2017). Gender gap in science, technology, engineering, and mathematics (STEM): Current knowledge, implications for practice, policy, and future directions. *Educational Psychology Review, 29*(1), 119–140.

Wang, X. (2013). Modeling entrance into STEM fields of study among students beginning at community colleges and four-year institutions. *Research in Higher Education, 54*(6), 664–692. https://doi.org/10.1007/s11162-013-9291-x.

Wout, D., Danso, H., Jackson, J., & Spencer, S. (2008). The many faces of stereotype threat: Group- and self-threat. *Journal of Experimental Social Psychology, 44*(3), 792–799.

Xue, Y., & Larson, R. (2015). STEM crisis or STEM surplus? Yes and yes. *Monthly Labor Review*. U.S. Bureau of Labor Statistics. https://doi.org/10.21916/mlr.2015.14.