CBE—Life Sciences Education
Vol. 12, 30–38, Spring 2013

Article

Stereotyped: Investigating Gender in Introductory Science Courses

Shanda Lauer,* Jennifer Momsen,* Erika Offerdahl,† Mila Kryjevskaia,‡ Warren Christensen,‡ and Lisa Montplaisir*

*Department of Biological Sciences, †Department of Chemistry and Biochemistry, and ‡Department of Physics, North Dakota State University, Fargo, ND 58108

Submitted September 1, 2012; Revised December 6, 2012; Accepted December 6, 2012
Monitoring Editor: Erin L. Dolan

Research in science education has documented achievement gaps between men and women in math and physics that may reflect, in part, a response to perceived stereotype threat. Research efforts to reduce achievement gaps by mediating the impact of stereotype threat have found success with a short values-affirmation writing exercise. In biology and biochemistry, however, little attention has been paid to the performance of women in comparison with men or perceptions of stereotype threat, despite documentation of leaky pipelines into professional and academic careers. We used methodologies developed in physics education research and cognitive psychology to 1) investigate and compare the performance of women and men across three introductory science sequences (biology, biochemistry, physics), 2) document endorsement of stereotype threat in these science courses, and 3) investigate the utility of a values-affirmation writing task in reducing achievement gaps. In our study, analysis of final grades and normalized learning gains on content-specific concept inventories reveals no achievement gap in the courses sampled, little stereotype threat endorsement, and no impact of the values-affirmation writing task on student performance. These results underscore the context-dependent nature of achievement gaps and stereotype threat and highlight calls to replicate education research across a range of student populations.

INTRODUCTION

Despite decades of active recruitment, women remain underrepresented in science, technology, engineering, and math (STEM) disciplines both in the United States and globally (Hewlett *et al.*, 2008; Simard *et al.*, 2008). Women leave STEM fields at all stages of their careers—as undergraduates, graduate students, professionals, and in the transitions between each stage, a phenomenon described as the leaky pipeline. In biology, for example, although women have reached parity with men when graduating from undergraduate and postgraduate schooling, women represent approximately one-third of the academic workforce (National Science Foundation [NSF], 2011). In contrast, the physics pipeline leak begins much earlier and is more substantial. Despite the fact that women and men are nearly equally represented in high school physics classes (44% vs. 56%), the pipeline turns into a "gaping hole" when they reach college (McCullough, 2002). Women comprise only 21% of physics undergraduate degrees, 22% of master's degrees, and 16% of PhDs (Mulvey and Nicholson, 2008). As these women move into academic and professional roles, they comprise 11% of the workforce (NSF, 2011).

The underlying causes of this disparity between men and women are numerous, complex, and pervasive. However, a recent meta-analysis of research on the gender gap in STEM (Hill *et al.*, 2010) found bias, stereotype threat, and social factors as prime driving forces contributing to the loss of women from STEM fields. In fact, recent work by Moss-Racusin *et al.* (2012) found science faculty across disciplines and regardless of gender exhibited an unconscious gender bias against undergraduate women, underscoring the pervasive and

DOI: 10.1187/cbe.12-08-0133
Address correspondence to: Jennifer Momsen (jennifer.momsen@ndsu.edu).

© 2013 S. Lauer *et al. CBE—Life Sciences Education* © 2013 The American Society for Cell Biology. This article is distributed by The American Society for Cell Biology under license from the author(s). It is available to the public under an Attribution–Noncommercial–Share Alike 3.0 Unported Creative Commons License (http://creativecommons.org/licenses/by-nc-sa/3.0).
"ASCB®" and "The American Society for Cell Biology®" are registered trademarks of The American Society for Cell Biology.

persistent nature of cultural stereotypes regarding women in science.

Gender and Achievement in Undergraduate Science Courses

The disparity between women and men in STEM disciplines may extend to achievement at the college level, resulting in a gender achievement gap—the persistent and pervasive underperformance of women as measured by exam scores, course grades, and learning gains on validated concept inventories.

Evidence for an achievement gap in biology and biochemistry at the undergraduate level is largely missing, in part because the fields are young. Women routinely underperform their male counterparts on the Medical College Admission Test, a pattern that can be traced back at least a decade (American Association of Medical Colleges, 2012). Further, a recent study by Willoughby and Metz (2009) found mixed evidence of a gender gap in an introductory biology course: women had significantly lower normalized learning gains as measured by a biological diagnostic test, but this result was not reproducible with any other measure, including alternative learning gain calculations, overall course grades, and individual exam scores. Many students from introductory biology go on to take introductory biochemistry. Yet there are few diagnostic tests for biochemistry (e.g., American Chemical Society Biochemistry Exam, Biochemistry and Cell Biology Graduate Record Examinations), and, to date, none have been used to explore the existence of a gender gap. Such limited results underscore the need for additional studies of how women and men perform in undergraduate life sciences courses, a need echoed by the recently released report on the status of discipline-based education research (DBER) by the National Academies of Science (2012).

In contrast, gender achievement gaps are well documented in physics at the undergraduate level (Lorenzo *et al.*, 2006; Pollock *et al.*, 2007; Kost *et al.*, 2009; Brewe *et al.*, 2010; Kost-Smith *et al.*, 2010). The calculus-based introductory physics sequence, a gateway to majors in physics and many other STEM disciplines, is the most frequently studied in physics education research (PER). A distinct gender gap exists on conceptual surveys among students before instruction (Lorenzo *et al.*, 2006; Pollock *et al.*, 2007; Brewe *et al.*, 2010), but some of this disparity may be due to gender bias in the instruments themselves (McCullough and Meltzer, 2001; Docktor and Heller, 2008; Willoughby and Metz, 2009; Dietz *et al.*, 2012). In courses with traditional instructional methods, this gap appears to persist; however, when instruction consists of highly interactive, research-validated instruction, the prevalence of an achievement gap is less consistent. Although learning gains are significant regardless of gender, some research finds the achievement gap reduced (Lorenzo *et al.*, 2006), while other research finds the gap persists (Pollock *et al.*, 2007; Brewe *et al.*, 2010). As noted previously, the presence of an achievement gap may be an artifact of overreliance on potentially biased conceptual surveys, especially when associated course grades and final exams do not reveal such a significant gap (Docktor and Heller, 2008; Willoughby and Metz, 2009).

In many instances, the gender gap in physics is attributed to disparities in mathematical preparation and ability. While a strong and persistent belief in a gender achievement gap in mathematics has prevailed for decades (e.g., Kane and Mertz, 2012), evidence for its existence is less conclusive (e.g., Hyde, 2005; Guiso *et al.*, 2008). In a meta-analysis of six large survey studies, Hedges and Nowell (1995) documented a small mean difference in mathematics achievement between men and women and modest differences in variance. More recent data in the United States refute a mathematics gender achievement gap, at least in the general populace grades 2 through 11 (Hyde *et al.*, 2008). Analyses of international data collected through studies such as the 2003 Trends in International Mathematics and Science Study (TIMMS) and 2003 Program for International Student Assessment (PISA) reveal significant variability between nations in the presence and effect size of a gap (Guiso *et al.*, 2008; Nosek *et al.*, 2009). While there seems to be some agreement that, in some contexts, the gender achievement gap is narrowing or may no longer exist, the implications for such a gap, no matter how small, are still of import. Hedges and Friedman (1993) predict that even a difference as small as 0.3 SD coupled with modest variance can account for as much as 2.5 times as many men in the top scoring percentiles than women.

In instances in which an achievement gap has been documented, the underlying causes of these differences in math performance are likely multiple and the relationships between them complex. Contextual factors play a key role in predicting differences in achievement. Analyses of TIMMS and PISA data identified sociocultural indicators of gender equality within a nation as a strong predictor of differences in achievement (Guiso *et al.*, 2008; Nosek *et al.*, 2009). Niederle and Vesterlund (2011) provide evidence that women perform differently than men on mathematics-related tasks when the situation is perceived to be highly competitive.

Stereotype Threat

Stereotype threat, described as a "risk of confirming ... a negative stereotype about one's group" (Steele and Aronson, 1995), may undermine achievement in the STEM classroom. Stereotype threat is not limited to gender and can apply to many intrinsic characteristics, including race, ethnicity, income level, and academic ability (Allport, 1954; Steele, 1997); however, we focus here on the impact of stereotype threat on the performance of women in undergraduate STEM courses.

Stereotype threat may be highly contextual, triggered by a survey item (Steele and Aronson, 1995), the gender of the instructor (Delisle *et al.*, 2009), or instructional practices (Kreutzer and Boudreaux, 2012), and can undermine academic success in several ways. First, stereotype threat can produce stress and induce anxiety, causing a student to become more self-conscious about his or her performance and to actively try to suppress those emotions, which may tax working memory and lead to decreased performance (Steele and Aronson, 1995; Schmader *et al.*, 2008; Delisle *et al.*, 2009). Second, prolonged exposure to stereotype threat can result in disidentification, wherein a student stops associating with a given stereotyped group and avoids situations likely to be perceived as threatening (Aronson *et al.*, 2002; Steele *et al.*, 2002). In science, stereotype threat may contribute to the leaky pipeline, causing the attrition of women from science-related majors.

While stereotype threat has become a popular explanation for differences in performance between men and women in STEM disciplines, recent work by Stoet and Geary (2012) calls into question the strength of empirical evidence supporting this hypothesis. They reviewed the research on gender differences in mathematics and performance and achievement to determine the strength of evidence supporting results from the original, critical study documenting activation of stereotype threat in mathematics (Spencer *et al.*, 1999). Stoet and Geary (2012) concluded that the evidence for activation of stereotype threat as the mediating factor of a gender achievement gap is far from robust. Although they identified 141 articles related to stereotype threat in mathematics, 20 of these were replication studies. Of these, just 11 (55%) were able to replicate the activation of stereotype threat as presented in the original paper. While they do not dismiss stereotype threat as a valid hypothesis, they do call into question the strength of the effect on achievement and performance, and they caution researchers and policy makers alike to consider the vast array of other possible contributing factors to the gender achievement gap.

Reducing the Impact of Stereotype Threat

Empirical work focused on ways to reduce or eliminate the effects of stereotype threat has revealed a number of simple yet effective measures, including educating at-risk populations (Johns *et al.*, 2005) and manipulating test-taking instructions (Steele and Aronson, 1995; Spencer *et al.*, 1999; Johns *et al.*, 2005). Social psychologists have also reduced the effects through mediation of contextual and societal factors related to stereotypes. Individuation has proved effective by explicitly distinguishing between the stereotyped individual and the stereotype to minimize stereotype usage (Locksley *et al.*, 1980; Langer *et al.*, 1985) and allowing stereotyped students to distance themselves from the stereotype in question, while remaining engaged in the task or course (Ambady *et al.*, 2004). Finally, because women are more likely to endorse the stereotype that science is for men when suitable female role models are largely absent (i.e., few female faculty; Delisle *et al.*, 2009), simply increasing the visibility of and engagement with positive female role models has proven efficacious (McIntyre *et al.*, 2004). In fact, simply having a competent woman administer a mathematics exam was sufficient to reduce the achievement gap in one study (Marx and Roman, 2002).

Values-affirmation tasks have recently received a great deal of attention (e.g., Cohen *et al.*, 2006; Miyake *et al.*, 2010) for their ability to reduce or eliminate stereotype threat. In this type of intervention, individuals take 10–15 min to write about values that are personally important but unrelated to the course. Such writing tasks appear effective in reducing or eliminating stereotype threat for African Americans (Cohen *et al.*, 2006; Walton and Cohen, 2007) and women (Martens *et al.*, 2006; Miyake *et al.*, 2010), with effects that may persist over time (Cohen *et al.*, 2009; Walton and Cohen, 2011). Although short and simple, values-affirmation writing tasks draw directly on students' experiences to actively engage each student as an individual (Yeager and Walton, 2011) and may promote deep processing to effect powerful results (Schwartz and Martin, 2004; Chase *et al.*, 2009). Thus, although simple, values-affirmation writing tasks have the potential to profoundly impact students experiencing stereotype threat (Yeager and Walton, 2011).

Testing the Efficacy of Values-Affirmation Tasks in Introductory Science

The work of Miyake *et al.* (2010) and Cohen *et al.* (2006) is encouraging, but each study represents only a single course or cohort of students at one institution. Given the complex nature of the classroom and the myriad factors that contribute to learning, it is necessary to replicate the values-affirmation study across institutions, semesters, and courses; indeed, this lack of replication studies is a serious deficit of current DBER practices (Singer *et al.*, 2012).

This study addresses this deficiency and specifically investigates the gender achievement gap across introductory science courses and tests the efficacy of a values-affirmation task in improving student performance. Specifically, we 1) characterized and compared the performance of women and men across three introductory science sequences (biology, biochemistry, and physics) at a large, public, research-intensive university; 2) documented endorsement of stereotype threat in these science courses; and 3) determined the utility of a values-affirmation writing task in reducing achievement gaps that may exist.

METHODS

University and Course Context

This land-grant, research university serves more than 14,000 undergraduate and graduate students. Women comprise 42% of the undergraduate population and 50% of the graduate population. Across the university, incoming freshmen have an average composite ACT score of 23.8 and an average high school grade point average (GPA) of 3.37.

This study targeted four science courses considered introductory for majors in the discipline, including introductory calculus-based physics 1 and 2, introductory biology, and introductory biochemistry. Introductory physics 1 is a lecture-based course taught by a male faculty member, and introduces Newtonian mechanics of translational and rotational motion, energy, work, power, momentum, conservation of energy and momentum, periodic motion, waves, sound, and heat and thermodynamics. Enrollment is typically 90–100 students. Introductory physics 2, taught by a female faculty member, is also a lecture-based course, and focuses on conceptual understanding of topics including electric charge; electric field; potential and current; magnetic field; capacitance, resistance, and inductance; circuits; electromagnetic waves; and optics. Enrollment is typically around 200 students. Introductory biology is a very large (300–400 students), lecture-based course taught by a female faculty member, and introduces students to cellular and molecular biology, genetics, and evolution. Biochemistry is also a large, lecture-based course with average enrollments of 300 students taught by a female faculty member, and focuses on biomolecules, generation and use of metabolic energy, biosynthesis, metabolic regulation, storage, transmission, and expression of genetic information.

Gender Achievement Gap

To investigate the presence and persistence of a gender achievement gap, we collected data, specifically final course grades by gender, from iterations of these courses taught in

Table 1. Participants in the values-affirmation task, as distributed among treatment groups

	Males (T/C)[a]	Females (T/C)[a]	Total
Introductory biology	138 (74/64)	131 (85/46)	269
Biochemistry	97 (61/36)	122 (74/48)	218
Physics 1	52 (29/23)	13 (9/4)	65
Physics 2	111 (66/45)	15 (9/6)	126

[a]T/C, treatment group vs. control group.

the 2010–2011 academic year. We also collected these data from Fall 2011, the same semester in which the values-affirmation writing task was implemented.

Values-Affirmation Exercise

We followed the protocol described by Miyake *et al.* (2010) to implement the values-affirmation exercise in four different introductory science courses in the Fall 2011 semester. This exercise was unrelated to the content of any of the courses included in this study. The exercise was distributed in a double-blind manner within the lecture component of each course. Given the predicted benefits of the task, we randomly assigned ∼60% of students in each course to the values-affirmation treatment group and ∼40% to the control group (Table 1). The first writing exercise was distributed the second week of classes, following students' completion of a discipline-appropriate concept inventory (Figure 1). A research assistant unaffiliated with any of the courses included in this study implemented the writing task following a well-defined script. Students were given 15 min to complete the writing task.

In the week prior to the second exam, students were asked to again complete the values-affirmation writing exercise. This "booster shot" was intended to help students reaffirm their values. This time, the writing exercise was administered online through a class Web page as a regular homework assignment. Students were invited individually to follow a link to an online replica of the writing exercise done in class, and the treatment conditions were kept the same as the first implementation. The instructions were the same, suggesting that students spend ∼15 min on the exercise.

Table 2. Discipline-specific concept inventories

Course	Concept inventory
Physics 1	Force and Motion Conceptual Evaluation[a]
Physics 2	Brief Electricity and Magnetism Assessment[b]
Introductory biology	Concept Inventory of Natural Selection[c]
Introductory biochemistry	Introductory Molecular and Cell Biology Assessment[d]

[a]Thornton and Sokoloff, 1998.
[b]Ding *et al.*, 2006.
[c]Anderson *et al.*, 2002.
[d]Shi *et al.*, 2010.

Stereotype Endorsement Measures

Again, following the protocol of Miyake *et al.* (2010), we also distributed a survey to measure students' endorsement of gendered stereotype threats, namely that men are generally better at a particular science (e.g., physics, biochemistry, or biology). Within the 45-item survey, we distributed two stereotype endorsement prompts, customized to each course: 1) according to my own personal beliefs, I expect men to generally do better than women in physics (or biochemistry or biology), and 2) according to my own personal beliefs, I expect women to generally do better than men in physics (or biology or biochemistry). The participants were asked to indicate their agreement on a 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). This approach does not specifically prime students' stereotype threat (e.g., by asking them to identify as female); rather, stereotype threat is activated by situational pressure, that is, being aware of the stereotype threat and being a member of the threatened group (e.g., women perform more poorly than men in science and I am a woman; e.g., Marx and Stapel, 2006).

Outcome Measures

The main outcome measures for this study included final course grades and learning gains (Hake, 1998), the latter measured by student performance on a discipline-appropriate concept inventory (Table 2). To test for differences between the performance of men and women, we used a chi-square analysis, with Fisher's exact test when sample sizes were too

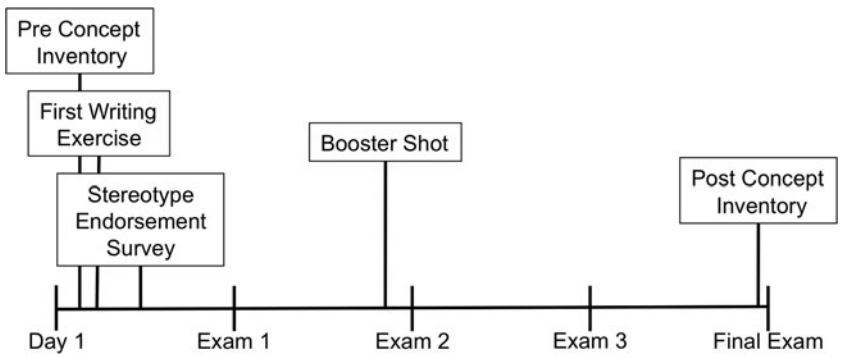

Figure 1. General timeline of the intervention and data collection.

Table 3. Chi-square analysis of final course grade distributions by gender

Course	Year	df	*n*	χ^2	*p* value
Introductory biology	2010	4	323	1.83	0.78
	2011	4	269	5.06	0.28
Biochemistry	2010	4	264	2.26	0.69
	2011	4	219	10.05	0.04
Physics 1	2010	4	74	3.14	0.56[a]
	2011	4	65	5.41	0.27[a]
Physics 2	2010	4	188	2.52	0.71[a]
	2011	4	126	1.28	0.94[a]

[a]Fisher's exact test used when the data set violated the assumption that each expected cell count was greater than five.

small to meet the assumptions of the chi-square analysis. To compare learning gains of men and women in treatment and control groups, we used Student's *t* test. Where appropriate, we calculated effect sizes using Cohen's *V* or *d* and included confidence intervals. Analyses were conducted using SAS (Cary, NC) software.

RESULTS

Gender Achievement Gap

There was no significant relationship between the distribution of final course grades and gender in biology or physics for any semester or section (Table 3). For biochemistry, however, there was significance, which shows a relationship between gender and letter grade for Fall 2011; however, women seemed to outperform men in this class and semester, although the effect size was small ($V = 0.2$, 95% CI [0.14, 0.3]). Further, we found no significant differences between normalized learning gains of men and women for any course (Table 4).

Stereotype Threat Endorsement

In all courses, students overwhelmingly rejected the claim that men do better than women in biology, biochemistry, or physics, with more than two-thirds of students strongly disagreeing or disagreeing with the statement (Figure 2). The distribution of responses for men differed significantly from women only in biology ($\chi^2(4) = 23.29$, $p < 0.001$), with women more likely to disagree with this claim.

Table 4. Comparing normalized learning gains for men and women in Fall 2011

Course	Mean difference[a]	df	*t*	*p* value
Introductory biology	−0.01	171.18	−0.20	0.84
Biochemistry	0.01	183	0.14	0.89
Physics 1	−0.17	42	−1.27	0.21
Physics 2	−0.10	89	−1.46	0.15

[a]A negative mean difference value indicates higher learning gains for the treatment group.

Values-Affirmation Writing Task

In all courses but one, physics 2, learning gains were higher for the treatment group over the control group, significantly so for only physics 1 (Table 5), with a moderate effect size ($d = -0.7$, 95% CI [−1.3, −0.09]). Further, in all courses but physics 1, final course grades were higher for the control group over the treatment group, significantly so for only physics 2 (Table 6), although the effect size was small ($d = 0.4$, 95% CI [0.04, 0.8]). Further, there was no significant difference in the distribution of final grades between treatment and control groups for women or men in any course (Table 7).

DISCUSSION

The existence of an achievement gap is often an assumption of the undergraduate physics classroom, yet remains an unknown in introductory biology and biochemistry courses. However, across semesters and outcome measures, we found no substantial evidence of an achievement gap between men and women in either introductory calculus-based physics courses or introductory biology and biochemistry. Although these findings align with studies in astronomy (Hufnagel *et al.*, 2004; Willoughby and Metz, 2009) and biology (Willoughby and Metz, 2009), they contradict what is typically reported in physics (Lorenzo *et al.*, 2006; Pollock *et al.*, 2007; Miyake *et al.*, 2010). Such discrepancies may be attributable to biases in how learning gains are calculated; indeed, normalized learning gains are particularly susceptible to bias, because there is a strong relationship between pretest scores and normalized learning gains (Coletta and Phillips, 2005; Brogt *et al.*, 2007). For example, because men typically have higher pretest scores than women on common physics concept inventories (e.g., Force Concept Inventory or Force and Motion Conceptual Evaluation), the subsequent calculation of normalized learning gains is particularly likely to identify a gender achievement gap. Our results utilized normalized learning gains, further underscoring the lack of an achievement gap in the sampled science courses.

Explaining gender achievement gaps, however, goes beyond statistical biases. Stereotype threat can play a role in student achievement, especially, as noted, on standardized tests and concept inventories in science and math. Women in science often ascribe to a negative stereotype regarding women's scientific competency. However, in this study, we found little to support the claim that women in the sampled population were endorsing a stereotype threat; rather, our evidence suggests that most women, and even men, reject this claim. We are cautious in our interpretation of these data for several reasons. In physics, these results may reflect the small sample size of women, although in such cases we might expect women would more readily self-identify as female and thus face an increased risk of experiencing stereotype threat. However, these results may reflect a stereotype reactance effect, wherein the stereotype is so blatant that women respond by overperforming (Kray *et al.*, 2001). Although our sample sizes for introductory biology and biochemistry are more robust, we believe this study is one of the first to explicitly explore gender achievement gaps and stereotype threat at the undergraduate level in either biology or biochemistry. As such, this research represents a single time

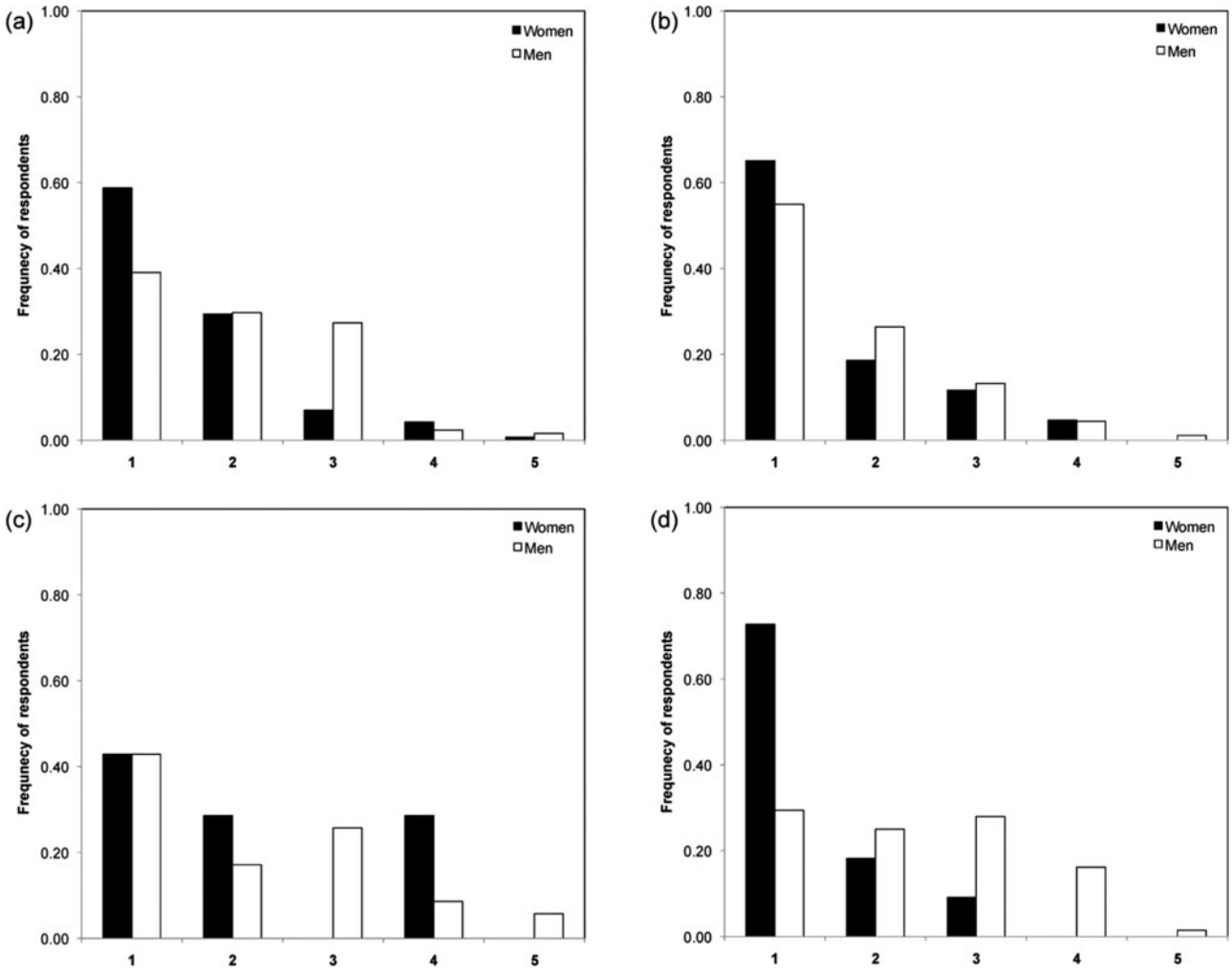

Figure 2. Frequency of student responses to the prompt: I expect men to generally do better than women in (a) biology (n = 227), (b) biochemistry (n = 243), (c) physics 1 (n = 44), or (d) physics 2 (n = 91). 1 = strongly disagree to 5 = strongly agree.

point and institution and is hardly representative of national trends.

Still, these results are perplexing in light of the broader research landscape, prompting us to question why these students may not ascribe to gender-based stereotype threats. One possible explanation emerges from self-efficacy literature, specifically, the role of vicarious experiences in shaping student's beliefs regarding self-efficacy. Vicarious experiences involve more than just a positive role model; they reflect repeated observations of "others perform[ing] threatening activities without adverse consequences" (Bandura, 1977). By extension, the observer can predict that her hard work and persistence can result in success. In the undergraduate setting, vicarious experiences for women include observing women in roles of authority and as experts, such as lab and recitation teaching assistants and course instructors. Given the institutional context of this study, vicarious

Table 5. Comparison of normalized learning gains between treatment and control groups

Course	Mean difference[a]	df	t	p value
Introductory biology	−0.07	130.13	−0.90	0.37
Biochemistry	−0.06	183	−1.36	0.18
Physics 1	−0.25	42	−2.32	0.03
Physics 2	0.04	87.36	0.83	0.41

[a]A negative mean difference value indicates higher learning gains for the treatment group.

Table 6. Comparison of final course grades between treatment and control groups

Course	Mean difference	df	t	p value
Introductory biology	1.70	257.94	0.96	0.34
Biochemistry	1.50	209.84	1.07	0.29
Physics 1	−3.96	63	−0.82	0.42
Physics 2	5.44	121.76	2.22	0.03

Table 7. Comparison of final course grades for treatment and control groups by gender and course

Course	Gender	Mean (± SD)	df	n	χ^2	p value[a]
Introductory biology	F	74.3 ± 15.1	4	131	7.67	0.11
	M	73.7 ± 14.4	4	138	3.03	0.57
Biochemistry	F	80.0 ± 8.8	4	122	2.21	0.80
	M	77.4 ± 12.7	4	97	6.55	0.17
Physics 1	F	83.2 ± 10.8	4	13	5.33	0.13
	M	78.8 ± 20.7	4	52	1.72	0.80
Physics 2	F	82.9 ± 11.1	4	15	2.85	0.60
	M	80.2 ± 15.5	4	111	3.28	0.55

[a]Fisher's exact test used when the data set violated the assumption that each expected cell count was greater than five.

experiences may play an important role in a student's perception of self-efficacy and stereotype threat. Introductory biology and biochemistry are both taught by female instructors, and female graduate students often lead the associated labs; thus, students are afforded multiple opportunities to observe women doing biology and biochemistry and may have greater self-efficacy when doing biology and biochemistry themselves. All women enrolled in biochemistry would have successfully completed at least one course in biology, and many would have also successfully completed a physics course. Prior success in biology and physics might serve to affirm women's beliefs in biochemistry that they "belong" in the field. Conversely, the physics department has only one female faculty member, and at the time of this study, no female graduate students. Thus, opportunities to observe women performing "threatening activities" were rare. However, we note the somewhat anomalous result of physics 2, in which 91% of women disagree or strongly disagree with the claim that men generally do better in physics. Taught by a female faculty member, instruction in this course regularly offers women an opportunity to observe a woman *doing* physics and may promote positive feelings of self-efficacy in female students. Further, women enrolled in physics 2 had successfully completed physics 1 (or equivalent), which is a prerequisite to physics 2, and therefore may have already identified themselves as capable of doing well in physics.

Just as vicarious experiences can influence endorsement of stereotype threat, other contextual elements might explain our inability to detect meaningful differences in achievement and stereotype threat endorsement. Schmader *et al.* (2008) presented a model postulating a link between stereotype threat and the activation of processes that tax otherwise available cognitive resources (e.g., physiological stress, suppression of negative emotions, and performance monitoring). When individuals endorse stereotypes, they are less likely to perform well, because they have fewer cognitive resources available. Alter *et al.* (2010) demonstrate that the way in which a task is presented can affect the degree to which an individual endorses or identifies with a given stereotype. They demonstrated differential performance in stereotyped groups dependent upon how a task was presented—either as a task or as a challenge. When groups susceptible to stereotype threat were presented a task couched as a threat (e.g., a measure of intelligence or academic ability), their performance was significantly poorer than when the task was presented as a challenge (e.g., a potentially difficult task from which much useful skills or knowledge could be learned). In our study, the concept inventories were introduced as neither a threat nor a challenge—rather the emphasis of the exercise was placed on completion of the task. As a result, we may have created an environment that reduced the activation of stereotype threat, which could explain the lack of achievement gap between groups of students.

Finally, the changing demographic of undergraduate students across the nation may impact the stereotypes students identify, the subsequent stereotype threats they are at risk of confirming, and ultimately, their performance and persistence in science. For example, we note that the student population sampled in this study differs substantially from the population studied in Miyake *et al.* (2010), with weaker academic preparation based on composite and subject area ACT scores and high school GPAs of entering freshmen. As a result, the aspirations, motivations, and self-efficacy of students in this study may differ markedly from those students attending a more competitive school, such as the one studied by Miyake *et al.* (2010).

IMPLICATIONS

Introductory science courses are diverse, complex systems with the potential to impact learning in multiple and sometimes unanticipated ways. Course context, including decisions about instructional practices, in concert with the changing demographic of our undergraduates, may reduce or enhance the prevalence of a gender achievement gap, as mediated by stereotype threat endorsement. As this research shows, gender achievement gaps are not a certainty in the science classroom, and stereotype threat endorsement may reflect factors of which we are currently unaware. We believe that this research supports recent calls from the DBER community (Singer *et al.*, 2012) for replication studies that investigate the role of gender in learning undergraduate science across a variety of course settings, time, and different outcome measures.

ACKNOWLEDGMENTS

This research received approval from the local institutional review board (IRB protocol #SM12014) and was supported in part by the National Science Foundation under NSF-HRD 0811239 as part of the North Dakota State University Advance/FORWARD program.

REFERENCES

Allport GW (1954). The Nature of Prejudice, Reading, MA: Addison-Wesley.

Alter AL, Aronson J, Darley JM, Rodriguez C, Ruble DN (2010). Rising to the threat: reducing stereotype threat by reframing the threat as a challenge. J Exp Soc Psychol *46*, 166–171.

Ambady N, Paik SK, Steele J, Owen-Smith A, Mitchell JP (2004). Deflecting negative self-relevant stereotype activation: the effects of individuation. J Exp Soc Psychol *40*, 401–408.

American Association of Medical Colleges (2012). MCAT Scores and GPAs for Applicants to US Medical Schools, 2002–2011, Washington, DC.

Anderson DL, Fisher KM, Norman GJ (2002). Development and evaluation of the Conceptual Inventory of Natural Selection. J Res Sci Teach *39*, 952–978.

Aronson J, Fried CB, Good C (2002). Reducing the effects of stereotype threat on African American college students by shaping theories of intelligence. J Exp Soc Psychol *38*, 113–125.

Bandura A (1977). Self-efficacy: toward a unifying theory of behavioral change. Psychol Rev *84*, 191–215.

Brewe E, Sawtelle V, Kramer LH, O'Brien GE, Rodriguez I, Pamelá P (2010). Toward equity through participation in Modeling Instruction in introductory university physics. Phys Rev ST Phys Educ Res *6*, 010106.

Brogt E, Sabers D, Prather EE, Deming GL, Hufnagel B, Slater TF (2007). Analysis of the astronomy diagnostic test. Astro Educ Rev *6*, 25–42.

Chase C, Chin D, Oppezzo M, Schwartz D (2009). Teachable agents and the protégé effect: increasing the effort towards learning. J Sci Educ Technol *18*, 334–352.

Cohen GL, Garcia J, Apfel N, Master A (2006). Reducing the racial achievement gap: a social-psychological intervention. Science *313*, 1307–1310.

Cohen GL, Garcia J, Purdie-Vaughns V, Apfel N, Brzustoski P (2009). Recursive processes in self-affirmation: intervening to close the minority achievement gap. Science *324*, 400–403.

Coletta VP, Phillips JA (2005). Interpreting FCI scores: normalized gain, preinstruction scores, and scientific reasoning ability. Am J Phys *73*, 1172–1182.

Delisle M-N, Guay F, Senécal C, Larose S (2009). Predicting stereotype endorsement and academic motivation in women in science programs: a longitudinal model. Learn Individ Differ *19*, 468–475.

Dietz RD, Pearson RH, Semak MR, Willis CW, Rebello NS, Engelhardt PV, Singh C (2012). Gender bias in the force concept inventory? In: 2011 Physics Education Research Conference, ed. NS Rebello, PV Engelhardt, and C Singh, Melville, NY: American Institute of Physics, 171–174.

Ding L, Chabay R, Sherwood B, Beichner R (2006). Evaluating an electricity and magnetism assessment tool: Brief Electricity and Magnetism Assessment (BEMA). Phys Rev ST Phys Educ Res *2*, 010105.

Docktor J, Heller K (2008). Gender differences in both force concept inventory and introductory physics performance. In: 2008 Physics Education Research Conference, ed. C Henderson, M Sabella, and L Hsu, Melville, NY: American Institute of Physics, 15–18.

Guiso L, Monte F, Sapienza P, Zingales L (2008). Culture, gender, and math. Science *320*, 1164–1165.

Hake RR (1998). Interactive-engagement versus traditional methods: a six-thousand-student survey of mechanics test data for introductory physics courses. Am J Phys *66*, 64–74.

Hedges LV, Friedman L (1993). Gender differences in variability in intellectual abilities: a reanalysis of Feingold's results. Rev Educ Res *63*, 94–105.

Hedges LV, Nowell A (1995). Sex differences in mental test scores, variability, and numbers of high-scoring individuals. Science *269*, 41–45.

Hewlett SA, Buck Luce C, Servon L, Sherbin L, Shiller P, Sosnovich E, Sumberg K (2008). The Athena Factor: Reversing the Brain Drain in Science, Engineering, and Technology, Boston, MA: Harvard Business School.

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.

Hufnagel B, Deming GL, Landato JM, Hodari AK (2004). The effect of stereotype threat on undergraduates in an introductory astronomy class. J Women Minor Sci Eng *10*, 89–98.

Hyde JS (2005). The gender similarities hypothesis. Am Psychol *60*, 581–592.

Hyde JS, Lindberg SM, Linn MC, Ellis AB, Williams CC (2008). Gender similarities characterize math performance. Science *321*, 494–495.

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

Kane JM, Mertz JE (2012). Debunking myths about gender and mathematics performance. Notice AMS *59*, 10–21.

Kost LE, Pollock SJ, Finkelstein ND (2009). Characterizing the gender gap in introductory physics. Phys Rev ST Phys Educ Res *5*, 010101.

Kost-Smith LE, Pollock SJ, Finkelstein ND (2010). Gender disparities in second-semester college physics: the incremental effects of a "smog of bias." Phys Rev ST Phys Educ Res *6*, 020112.

Kray LJ, Thompson L, Galinsky A (2001). Battle of the sexes: gender stereotype confirmation and reactance in negotiations. J Pers Soc Psychol *80*, 942–958.

Kreutzer K, Boudreaux A (2012). Preliminary investigation of instructor effects on gender gap in introductory physics. Phys Rev ST Phys Educ Res *8*, 010120.

Langer EJ, Bashner RS, Chanowitz B (1985). Decreasing prejudice by increasing discrimination. J Pers Soc Psychol *49*, 113–120.

Locksley A, Borgida E, Brekke N, Hepburn C (1980). Sex stereotypes and social judgment. J Pers Soc Psychol *39*, 821–831.

Lorenzo M, Crouch CH, Mazur E (2006). Reducing the gender gap in the physics classroom. Am J Phys *74*, 118.

Martens A, Johns M, Greenberg J, Schimel J (2006). Combating stereotype threat: the effect of self-affirmation on women's intellectual performance. J Exp Soc Psychol *42*, 236–243.

Marx DM, Roman JS (2002). Female role models: protecting women's math test performance. Pers Soc Psychol Bull *28*, 1183–1193.

Marx DM, Stapel DA (2006). Distinguishing stereotype threat from priming effects: on the role of the social self and threat-based concerns. J Pers Soc Psychol *91*, 243–254.

McCullough L (2002). Women in physics: a review. Phys Teach *40*, 86–91.

McCullough L, Meltzer DE (2001). Differences in male/female response patterns on alternative-format versions of FCI items. In: 2011 Physics Education Research Conference, ed. NS Rebello, PV Engelhardt, and C Singh, Melville, NY: American Institute of Physics, 103–106.

McIntyre RB, Lord CG, Gresky DM, Frye GDJ, Bond CF, Jr. (2004). A social impact trend in the effects of role models on alleviating women's mathematics stereotype threat. Curr Res Soc Psychol *10*, 116–136.

Miyake A, Kost-Smith LE, Finkelstein ND, Pollock SJ, Cohen GL, Ito TA (2010). Reducing the gender achievement gap in college science: a classroom study of values affirmation. Science *330*, 1234–1237.

Moss-Racusin CA, Dovidio JF, Brescoll VL, Graham MJ, Handelsman J (2012). Science faculty's subtle gender biases favor male students. Proc Natl Acad Sci USA *109*, 16474–16479.

Mulvey PJ, Nicholson S (2008). Enrollments and Degree Report, 2006 (No. R-151.43), College Park, MD: American Institute of Physics Statistical Research Center.

National Science Foundation (2011). Women, Minorities, and Persons with Disabilities in Science and Engineering: 2011 (No. Special Report NSF 11-309), Arlington, VA.

Niederle M, Vesterlund L (2011). Gender and competition. Annu Rev Econ *3*, 601–630.

Nosek BA *et al.* (2009). National differences in gender–science stereotypes predict national sex differences in science and math achievement. Proc Natl Acad Sci USA *106*, 10593–10597.

Pollock SJ, Finkelstein ND, Kost LE (2007). Reducing the gender gap in the physics classroom: how sufficient is interactive engagement? Phys Rev ST Phys Educ Res *3*, 010107.

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

Schwartz DL, Martin T (2004). Inventing to prepare for future learning: the hidden efficiency of encouraging original student production in statistics instruction. Cogn Instr *22*, 129–184.

Shi J, Wood WB, Martin JM, Guild NA, Vicens Q, Knight JK (2010). A diagnostic assessment for introductory molecular and cell biology. CBE Life Sci Educ *9*, 453–461.

Simard E, Henderson AD, Gilmartin SK, Schiebinger L, Whitney T (2008). Climbing the Technical Ladder: Obstacles and Solutions for Mid-Level Women in Technology, Stanford, CA: Michelle R. Clayman Institute for Gender Research, Stanford University, and Anita Borg Institute for Women and Technology.

Singer SR, Nielsen NR, Schweingruber HA (2012). Discipline-Based Education Research: Understanding and Improving Learning in Undergraduate Science and Engineering, Washington, DC: National Academies Press.

Spencer SJ, Steele M, Quinn DM (1999). Stereotype threat and women's math performance. J Exp Soc Psychol *35*, 4–28.

Steele CM (1997). A threat in the air: how stereotypes shape intellectual identity and performance. Am Psychol *52*, 613–629.

Steele CM, Aronson J (1995). Stereotype threat and the intellectual test performance of African Americans. J Pers Soc Psychol *69*, 797–811.

Steele CM, Spencer SJ, Aronson J (2002). Contending with group image: the psychology of stereotype and social identity threat. In: Advances in Experimental Social Psychology, vol. 34, ed. MP Zanna, San Diego, CA: Academic Press, 379–440.

Stoet G, Geary DC (2012). Can stereotype threat explain the gender gap in mathematics performance and achievement? Rev Gen Psychol *16*, 93–102.

Thornton RK, Sokoloff DR (1998). Assessing student learning of Newton's laws: the Force and Motion Conceptual Evaluation and the Evaluation of Active Learning Laboratory and Lecture Curricula. Am J Phys *66*, 338–352.

Walton GM, Cohen GL (2007). A question of belonging: race, social fit, and achievement. J Pers Soc Psychol *92*, 82–96.

Walton GM, Cohen GL (2011). A brief social-belonging intervention improves academic and health outcomes of minority students. Science *331*, 1447–1451.

Willoughby SD, Metz A (2009). Exploring gender differences with different gain calculations in astronomy and biology. Am J Phys *77*, 651.

Yeager DS, Walton GM (2011). Social-psychological interventions in education: they're not magic. Rev Educ Res *81*, 267–301.