Developmental Psychology

An Examination of Stereotype Threat Effects on Girls' Mathematics Performance

Colleen M. Ganley, Leigh A. Mingle, Allison M. Ryan, Katherine Ryan, Marina Vasilyeva, and Michelle Perry

Online First Publication, January 28, 2013. doi: 10.1037/a0031412

Developmental Psychology

0012-1649/13/$12.00 DOI: 10.1037/a0031412

An Examination of Stereotype Threat Effects on Girls' Mathematics Performance

Colleen M. Ganley and Leigh A. Mingle
University of Illinois at Urbana-Champaign

Allison M. Ryan
University of Michigan at Ann Arbor

Katherine Ryan
University of Illinois at Urbana-Champaign

Marina Vasilyeva
Boston College

Michelle Perry
University of Illinois at Urbana-Champaign

Stereotype threat has been proposed as 1 potential explanation for the gender difference in standardized mathematics test performance among high-performing students. At present, it is not entirely clear how susceptibility to stereotype threat develops, as empirical evidence for stereotype threat effects across the school years is inconsistent. In a series of 3 studies, with a total sample of 931 students, we investigated stereotype threat effects during childhood and adolescence. Three activation methods were used, ranging from implicit to explicit. Across studies, we found no evidence that the mathematics performance of school-age girls was impacted by stereotype threat. In 2 of the studies, there were gender differences on the mathematics assessment regardless of whether stereotype threat was activated. Potential reasons for these findings are discussed, including the possibility that stereotype threat effects only occur in very specific circumstances or that they are in fact occurring all the time. We also address the possibility that the literature regarding stereotype threat in children is subject to publication bias.

Keywords: gender differences, stereotype threat, mathematics performance

There is currently a debate as to whether or not the gender gap in mathematics achievement has closed (Corbett, Hill, & St. Rose, 2008; Entwisle, Alexander, & Olson, 1994; Hyde, 2005; Lindberg, Hyde, Peterson, & Linn, 2010; Robinson & Lubienski, 2011). With regard to research on mathematics test performance, some studies have found no gender differences (Hyde, 2005; Hyde & Linn, 2006; Hyde, Lindberg, Linn, Ellis, & Williams, 2008; Spelke, 2005), whereas others have found small gender differences (College Board, 2009, 2010; Gibbs, 2010; Gonzales et al., 2008; McGraw, Lubienski, & Strutchens, 2006; Robinson & Lubienski, 2011). A particular concern is that a larger gap exists at the top end of the distribution. That is, the highest performing boys significantly outperform the highest performing girls (Hedges & Friedman, 1993; Lindberg et al., 2010; Strand, Deary, & Smith, 2006).

In a recent meta-analysis, Lindberg et al. (2010) concluded that there is no overall gender difference in mathematics performance. However, their data suggest that a small to medium-sized difference exists in high school (d = 0.23) and with high-performing students (d = 0.40). In other studies, using a nationally representative data set, researchers have shown that although there is no gender difference at kindergarten, a gender difference develops by the third grade (d = 0.24; Fryer & Levitt, 2010; Penner & Paret, 2008; Robinson & Lubienski, 2011). In addition, these researchers found that the gender

Colleen M. Ganley, Departments of Curriculum and Instruction and Educational Psychology, University of Illinois at Urbana-Champaign; Leigh A. Mingle, Department of Educational Psychology, University of Illinois at Urbana-Champaign; Allison M. Ryan, Combined Program in Education and Psychology, University of Michigan at Ann Arbor; Katherine Ryan, Department of Educational Psychology, University of Illinois at Urbana-Champaign; Marina Vasilyeva, Department of Counseling, Developmental and Educational Psychology, Boston College; Michelle Perry, Department of Educational Psychology, University of Illinois at Urbana-Champaign.

This research was supported by Boston College through a Dissertation Fellowship from the Lynch School of Education and an Individual Research Grant from the Graduate Student Association; by the Department of Educational Psychology at the University of Illinois at Urbana-Champaign; by the Spencer Foundation; and by the Institute of Education Sciences, U.S. Department of Education, through Grant R305B100017 to the University of Illinois at Urbana-Champaign. The opinions expressed are those of the authors and do not represent views of the institute or the U.S. Department of Education.

We would like to thank the principals, teachers, and students involved in this research. We would also like to thank Alana Dulaney, Genevieve Henricks-Lepp, Sarah Keifer, and Katie-Beth Price for assistance with data collection, as well as Beth Casey, Lillie Albert, and Elida Laski for helpful feedback on this research.

Correspondence concerning this article should be addressed to Colleen M. Ganley, College of Education, University of Illinois at Urbana-Champaign, 1310 South Sixth Street, Champaign, IL 61820. E-mail: cganley@illinois.edu

difference emerges earlier, and is often larger, in the upper tail of the distribution.

In contrast to the research on mathematics test performance, research on mathematics classroom grades shows that girls perform similarly or better than boys across all years of schooling (Corbett et al., 2008; Ding, Song, & Richardson, 2006; Pomerantz, Altermatt, & Saxon, 2002). Classroom grades provide a somewhat different measure of mathematics performance when compared to test scores. Grades measure mastery of material explicitly taught in school and teachers are likely to take into account nonacademic factors that favor girls (e.g., conscientiousness; Friedman & Frisbie, 1995; Ornstein, 1994).

Taken together, this body of research reveals complex and sometimes conflicting results concerning gender differences in mathematics. Yet, despite the diversity of findings across ages and types of measures, most evidence points to a small gender difference in mathematics test performance that may increase as students get older, and is larger in higher performing students. Because significant gender differences do not exist early on but develop over time, it is important to understand the factors related to the development of this gender gap.

Researchers have posited a number of potential explanations for gender differences in mathematics performance and for the underrepresentation of women in mathematics- and science-related careers, including biological factors (e.g., Benbow & Stanley, 1980; Geary, 1996; Scarr & Carter-Saltzman, 1982), social factors (e.g., Eccles, 1987; Eccles & Jacobs, 1986; Heller & Ziegler, 1996), and the interaction between biological and social factors (e.g., Halpern & Tan, 2001; Nuttall, Casey, & Pezaris, 2005). One of the key social factors that has been suggested as contributing to high-achieving women's underperformance on mathematics tests is stereotype threat (Steele, 1997; Steele & Aronson, 1995).

Stereotype Threat Theory and Gender Differences in Mathematics Performance

Stereotype threat is a phenomenon whereby certain groups of people are affected by an unconscious fear of confirming a negative stereotype concerning their performance in a particular domain (e.g., that men are better than women in mathematics). The idea is that women, when the stereotype is primed prior to taking a mathematics test, perform worse on the test than women in a situation in which the stereotype is not primed, whereas men perform equally in both conditions (Spencer, Steele, & Quinn, 1999; Steele, 1997). Consistent with the notion of stereotype threat, women in a stereotype nullification condition, in which they are presented with information that is inconsistent with the stereotype (e.g., about girls doing as well as boys in mathematics), could be expected to perform better than women in normal or stereotype threat conditions (e.g., Smith & White, 2002).

Conditions Under Which Stereotype Threat Effects Occur

In discussing stereotype threat effects, it is important to address for whom and under what conditions these effects occur. Past research with college students has suggested that to be impacted by stereotype threat, women must be identified with mathematics and take a difficult mathematics test in an evaluative situation in which their gender is made salient.

Identification with mathematics. Steele (1997) proposed that stereotype threat affects people who identify with the domain in question (in this case, women who are identified with mathematics; Forbes, Schmader, & Allen, 2008; Jamieson & Harkins, 2007; Smith & White, 2001). Mathematics identification involves two components: feeling that you are good at mathematics and feeling that it is important to you to be good at mathematics (Smith & White, 2001). Research with high school and college students shows that women who are at least moderately identified with mathematics are more susceptible to stereotype threat effects than those who are not mathematics identified (Keller, 2007; Nguyen & Ryan, 2008; Smith & White, 2001).

Testing conditions. Tests that are introduced as evaluative, or indicative of one's ability, lead to the feeling that poor performance on the test indicates low ability (Aronson & Steele, 2005; Good & Aronson, 2008; Steele, 1997). This, combined with one's gender being made salient, leads women to believe that if they perform poorly on the test, they are at risk of confirming the negative stereotype about women and mathematics (Good & Aronson, 2008; Steele, 1997). Note that gender can be made salient in a number of ways, such as mentioning gender differences, marking ones gender, or taking the test in a mixed-gender group.

The effects of these feelings are more salient when women are taking difficult tests for at least two reasons. First, women are more likely to perform poorly on these assessments, making their fear of confirming the stereotype more plausible (Neuville & Croizet, 2007; O'Brien & Crandall, 2003; Spencer et al., 1999; Steele, 1997). Second, more difficult tests contain items that require more processing in working memory, and because working memory appears to be compromised when students are under stereotype threat (Schmader, Johns, & Forbes, 2008), performance on items that require more working memory resources would suffer more than performance on those items requiring fewer working memory resources (Quinn & Spencer, 2001; Schmader & Johns, 2003). These characteristics of the participants and testing situation are critical when examining stereotype threat.

Stereotype Threat Effects in Childhood and Adolescence

Much research has been conducted investigating stereotype threat in samples of college women (see Nguyen & Ryan, 2008). Although researchers often consider stereotype threat to be a well-established phenomenon in college women, a recent review and meta-analysis calls the strength of this phenomenon into question, suggesting that claims that stereotype threat is a robust phenomenon are exaggerated (Stoet & Geary, 2012). We have even less of an understanding of the nature of stereotype threat effects in childhood and adolescence (Good & Aronson, 2008). Knowing at what ages students are susceptible to stereotype threat effects can help to identify the most appropriate ages at which to target interventions designed to alleviate the effects of stereotype threat.

Developmental Requirements

It has been posited that for stereotype threat to impact girls' mathematics performance, several cognitive and social-cognitive abilities are needed. Specifically, Aronson and Good (2003) suggested four necessary developmental conditions. Girls must (a) be aware of gender stereotypes, (b) understand the societal and personal implications of these stereotypes, (c) have a sufficiently developed gender identity, and (d) have a well-formed conception of academic ability. They argued that these developmental requirements are all met by the time students begin middle school (ages 11–12). However, many of these factors likely emerge earlier, and gradually become more stable over the course of development (e.g., Cvencek, Meltzoff, & Greenwald, 2011; Levy & Carter, 1989; Nicholls, 1978).

Empirical Findings

Evidence for stereotype threat effects in mathematics for elementary, middle, and high school girls is inconsistent. Some studies report evidence of stereotype threat effects with girls as young as kindergarten age (Ambady, Shih, Kim, & Pittinsky, 2001; Tomasetto, Alparone, & Cadinu, 2011), whereas others have not found these effects even for high school girls (e.g., Cruz-Duran, 2009; Stricker & Ward, 2004). In Table 1 we have summarized the findings from the extant research on stereotype threat effects on girls' mathematics performance, including the studies reported in the present article and unpublished dissertations. Stereotype threat activation methods for each study can be found in Table 2.

Before discussing the findings summarized in Table 1, we would like to make several points concerning our approach to interpreting the analyses reported in prior studies. First, some studies have reported marginal findings ($.08 < p$s $< .12$) as indicative of stereotype threat effects; however, we consistently used a significance level of .05 across studies in our interpretations of their findings. Second, we included in Table 1 a separate column indicating whether the study involved both girls and boys because we view this as an important issue in understanding stereotype threat effects. Several of the reviewed studies included only girls. This leaves open the possibility that boys would show the same pattern of performance across the two conditions, preventing us from concluding that the observed difference in performance as a function of stereotype threat is unique to girls (Stoet & Geary, 2012). Thus, we obtain stronger evidence from studies with both genders because we can deduce both that there is a stereotype threat effect for girls and that there is not one for boys.

From a statistical perspective, we interpret findings as showing stereotype threat effects only if certain conditions are met. For studies involving both boys and girls, we require that there be a significant interaction between gender and stereotype threat condition. Further, for all studies, we require there to be a significant difference between girls in the stereotype threat condition and girls in the no-threat condition. If a study involving both gender groups finds a significant interaction but not a difference between girls' performance in the two conditions, it may mean that the interaction is pulled by an opposite performance pattern in boys, preventing one from making strong conclusions about stereotype threat effects on girls' performance.

Early elementary school. Among studies investigating stereotype threat effects in early elementary school students, one study found stereotype threat effects (Ambady et al., 2001), two studies reported mixed results (Neuville & Croizet, 2007; Tomasetto et al., 2011), and one study did not find effects (Muzzatti & Agnoli, 2007). Neuville and Croizet (2007) found that stereotype threat had a negative effect on performance on difficult items but had a positive effect on easy items. Tomasetto et al. (2011) found that girls whose mothers neither accepted nor rejected the gender stereotype about mathematics were susceptible to stereotype threat effects but girls whose mothers rejected the stereotype were not affected.

Upper elementary school. Two published studies and one unpublished dissertation have found no stereotype threat effect in upper elementary school students (Ambady et al., 2001; Good, 2001; Muzzatti & Agnoli, 2007). In fact, Ambady et al. (2001) found that girls in the stereotype threat condition performed better than girls in the no-threat condition.

Middle school. Three published studies found stereotype threat effects during middle school (Ambady et al., 2001; Huguet & Regner, 2009; Muzzatti & Agnoli, 2007, Experiment 2), one published study showed mixed results (Huguet & Regner, 2007, Study 2), and one published study (Huguet & Regner, 2007, Study 1) and one unpublished dissertation (Good, 2001) showed no evidence of a stereotype threat effect. Huguet and Regner (2007, Study 2) found that girls who completed the task in a mixed-gender setting were impacted by stereotype threat but those taking it in a same-gender setting were not.

High school. For high school students, two published studies found mixed results (Keller, 2007; Picho & Stephens, 2012), and two published and two unpublished dissertations found no effect (Cruz-Duran, 2009; Dinella, 2004; Keller & Dauenheimer, 2003; Stricker & Ward, 2004). Keller (2007) found that stereotype threat led to poorer mathematics performance only on difficult items for highly mathematics-identified girls. Picho and Stephens (2012) found stereotype threat effects among girls attending coeducational schools but not among girls attending single-sex schools.

Purpose of the Present Research

Given the inconsistent evidence in the existing research, our aim was to further investigate stereotype threat effects on girls' mathematics performance in childhood and adolescence. We conducted three studies, two with young adolescents and a third with children, younger adolescents, and older adolescents.

We used evidence from stereotype threat research and theory to inform our choices with regard to sampling, activation methods, and mathematics assessments. First, we chose higher performing participants because they are more likely to be identified with mathematics and are thus more likely to be susceptible to stereotype threat effects (Forbes et al., 2008; Smith & White, 2001; Steele & Aronson, 1995). Second, the testing situations were designed both to make gender salient and to be evaluative (Aronson & Steele, 2005; Good & Aronson, 2008; Steele, 1997). Third, in each study we used fairly difficult mathematics assessments (Neuville & Croizet, 2007; Nguyen & Ryan, 2008; O'Brien & Crandall, 2003; Spencer et al., 1999; Steele, 1997).

Study 1

Study 1 was conducted with middle- and high-performing eighth-grade students. In this study, we assigned boys and girls to either a stereotype threat condition or a stereotype nullification condition using a randomized block design. Students in the ste-

Table 1
Summary of Research on Stereotype Threat in Childhood and Adolescence

Study	Population	Grade(s)	Age (years)	Stereotype threat effect found	*N* (girls)	*d*[a]	Male comparison
			Early elementary school				
Ambady et al. (2001)	United States (Asian American)	K−2	5–7	Yes	20[b]		No
Muzzatti & Agnoli (2007), Experiment 1	Italy	2	7–8	No	34	0.05	Yes
Neuville & Croizet (2007)	France	2	7–8	Yes, difficult items	45	−0.62	Yes
				No, easy items		0.62	
Tomasetto et al. (2011)	Italy	K−2	5–8	Yes, when mom has no stereotype	Total: 124	–0.74[c]	No
				No, when mom rejects stereotype		0.00[c]	
			Upper elementary school				
Ambady et al. (2001)	United States (Asian American)	3–5	8–10	No (opposite effect)	29[b]		No
Study 3	United States	4	9–10	No	29	0.17	Yes
Good (2001)[d]	United States	4, 5	9–11	No	4th: 22		Yes
				No	5th: ≈17		
Muzzatti & Agnoli (2007), Experiment 1	Italy	3, 4, 5	8–11	No	3rd: 68	0.23	Yes
				No	4th: 64	0.13	
				No	5th: 42	−0.43	
Muzzatti & Agnoli (2007), Experiment 2	Italy	3, 5	8–9,	No	3rd: 44	0.29	Yes
			10–11	No	5th: 48	–0.28	
			Middle school				
Ambady et al. (2001)	United States (Asian American)	6–8	11–13	Yes	28[b]		No
Study 1	United States	8	13–14	No	110	0.14	Yes
Study 2	United States	7, 8	12–14	No	7th: 119	0.28	Yes
					8th: 95	−0.16	
Study 3	United States	8	13–14	No	65	0.14	Yes
Good (2001)[d]	United States	6	11–12	No	26		Yes
Huguet & Regner (2007), Study 1	France	6, 7	11–13	No	20	–0.78	Yes
Huguet & Regner (2007), Study 2	France	6, 7	11–13	Yes, in a mixed-gender setting	223	–0.82	Yes
				No, in a same-gender setting		0.01	
Huguet & Regner (2009)	France	6, 7	11–13	Yes	92	−0.78	Yes
Muzzatti & Agnoli (2007), Experiment 2	Italy	8	13–14	Yes	33	−0.37	Yes
			High school				
Cruz-Duran (2009)[d,e]	United States	10–12	14–18	No	415	−0.18	No
Dinella (2004)[d]	United States	9–12	13–18	No	133	0.36	Yes
Study 3	United States	12	17–18	No	76	−0.27	Yes
Keller (2007)[e]	Germany	10	15–16	Yes, high math ID, difficult items	High ID: 23	−0.82	Yes
				No (opposite effect), low math ID, difficult items	Low ID: 32	0.80	
				No, high math ID, easy items			
				No, low math ID, easy items			
Keller & Dauenheimer (2003)[e]	Germany	10	15–16	No	35	−0.47	Yes
Picho & Stephens (2012)	Uganda	10	15–16	Yes, coed school	Coed: 38	−0.76	No
				No, single sex school	Single sex: 51	−0.14	
Stricker & Ward (2004), Study 1	United States	11, 12	16–18	No	694	−0.16	Yes

Note. K = kindergarten; ID = identification.
[a] A negative effect size indicates that girls in the stereotype threat condition performed worse than girls in the no-threat condition (i.e., a stereotype threat effect for girls). Also note that a number of studies did not include enough information with which to compute effect sizes. Some effect sizes reported here were calculated from information available in the articles, whereas others were calculated from information obtained from the authors. [b] The sample size reflects the number of girls in three conditions, one of which is an Asian identity condition not discussed in the current study; sample sizes were not disaggregated by condition. [c] These effect sizes are based on mother's stereotyping one standard deviation below the mean and one standard deviation above the mean. [d] Unpublished dissertation. [e] This study compared a stereotype threat condition with a stereotype nullification condition.

Table 2
Stereotype Threat Activation Methods Used With Children and Adolescents

Study	Stereotype threat condition	Comparison condition
Ambady et al. (2001), Grades K–2; Neuville & Croizet (2007); Tomasetto et al. (2011)	Drew a picture of a girl holding doll	Drew a picture of a landscape
Ambady et al. (2001), Grades 3–8	Answered gender-related questions	Answered neutral questions
Muzzatii & Agnoli (2007)	Saw a picture of nine male and one female mathematicians	Saw a picture of nine flowers and one fruit
Huguet & Regner (2007), Study 1	Told that the task was a "geometry test"	Told that the task was a "memory game"
Huguet & Regner (2007), Study 2; Huguet & Regner (2009)	Told that the task measured ability in geometry	Told that the task measured ability in drawing
Stricker & Ward (2004)	Checked gender before test	Checked gender after test
Keller (2007); Keller & Dauenheimer (2003)	Read that the test showed gender differences	Read that the test showed no gender differences[a]
Cruz-Duran (2009)	Shown research evidence that men do better than women on math tasks	Shown research evidence that there are no gender differences on math tasks[a]
Dinella (2004)	Indicated gender at beginning of tests and told that it is important to do because gender differences are sometimes present on tests	Indicated gender at beginning of test
Good (2001)	Told that the test will show how smart they are in math	Told that the problems are to see how students think about math and reading
Picho & Stephens (2012)	Told that the test assesses students' ability in math and that there are gender differences on the test	Given only basic test instructions

Note. K = kindergarten.
[a] This is a stereotype nullification condition.

reotype threat condition were shown a video that presented fictitious scientific evidence showing that mathematics ability is fixed and that girls have lower levels of this ability. Students in the stereotype nullification condition were shown a video that presented evidence that the brain is malleable and that boys and girls have equal levels of mathematics ability.

Method

Participants. Participants were 212 (102 boys, 110 girls) middle- and high-performing eighth-grade students (13–14 years old) from three small urban schools in the Midwest with about 15%–45% of the student body eligible for free or reduced price lunch. These middle- and high-performing students were identified based on course enrollment, standardized test scores, and classroom grades. Almost all students (98.6%) scored above the midpoint of a 10-item mathematics identification scale (α = .87). The ethnic makeup of the sample was 85.4% Caucasian, 6.1% African American, 1.4% Hispanic, 0.9% Asian American, and 3.8% other.

Stereotype threat manipulation. Before taking the test, students watched a video that adopted imagery similar to that used by Dweck and her colleagues in their Brainology program (Dweck, 2008). The video shown to the stereotype threat group depicted a scientist telling students that recent research "shows that math intelligence levels among students do not change as students get older. Students are born with a certain amount of natural math ability which does not change." Students were then shown brain imagery and were given a detailed explanation regarding how some students are born with better mathematics skills (as indicated by more brain activity). In this condition, the students were also told that "females have lower levels of this kind of brain activity than males. This makes sense because girls often get lower scores on standardized tests compared to boys." These students were also told that the "test that you will take today is a very good measure of your natural math ability."

In contrast, the video shown to the stereotype nullification group depicted a scientist telling students that recent research

> shows that among students, such as yourselves, math ability levels can and do change as students get older. We are finding that math ability is not just something you are born with. Math ability grows with practice, just like exercise strengthens your muscles. Students are capable of learning and mastering new math concepts at any time in their lives.

Students were then shown brain imagery and were given a detailed explanation regarding how learning mathematics can change the brain and increase brain activity over time. In this condition, students were also told that "males and females have equal levels of this kind of brain activity. This makes sense because young men and women, like yourselves, score the same on standardized math tests."

Mathematics test. Each participant was given a mathematics test (α = .90; Cronbach, 1951) consisting of 30 retired items (mean percent correct = 65%) from the eighth-grade mathematics section of a state No Child Left Behind accountability assessment, the Large Midwestern Achievement Test.

Procedure. We randomly assigned participants to either the stereotype threat or stereotype nullification condition using a blocking design based on their gender, seventh-grade standardized test scores, and grades. Students in the two groups were brought to separate classrooms and watched the video that either activated stereotype threat or nullified the stereotype and then completed the mathematics test. The testing session took approximately 40 min.

Results and Discussion

To investigate the effects of stereotype threat, we ran a factorial analysis of variance (ANOVA) with two between-subject variables: stereotype threat condition (stereotype threat, stereotype nullification) and gender (girls, boys). The dependent variable was the percent of items answered correctly on the mathematics test.

The interaction between gender and stereotype threat was not significant, indicating that there was no stereotype threat effect, $F(1, 208) = 0.40$, $p = .53$ (see Table 3). The effect sizes for the differences between performance in the two conditions were 0.14 for girls and 0.00 for boys. Note that a positive effect size indicates that the stereotype threat group performed better and a negative effect size indicates that the stereotype threat group performed worse. The main effect of stereotype threat was also not significant, $F(1, 208) = 0.21$, $p = .65$. There was, however, a significant main effect of gender on the mathematics test, with boys outperforming girls, $F(1, 208) = 6.23$, $p = .01$, $d = 0.34$. A follow-up analysis focusing only on more difficult items (those with less than 50% correct; six items; 16.4% of participants incorrectly answered all six problems) revealed the same pattern of findings: The interaction between gender and stereotype threat was not statistically significant ($d_{\text{girls}} = 0.01$).

In summary, in Study 1, we found no evidence of a stereotype threat effect (even on the most difficult items) when a stereotype threat condition was compared to a stereotype nullification condition. However, there was a main effect of gender, indicating that the girls in this study underperformed compared to boys regardless of condition.

Study 2

The second study was conducted with high-achieving seventh- and eighth-grade students. Participants were randomly assigned to stereotype threat and no-threat conditions and given a very difficult mathematics test (to increase the likelihood that effects of stereotype threat on performance would be found). The stereotype threat activation method was explicit, but it was more similar to methods used in past research than the explicit activation method used in Study 1 (e.g., Keller, 2007; Keller & Dauenheimer, 2003).

Method

Participants. Participants were 224 (105 boys, 119 girls) seventh-grade students (12–13 years old) and 177 (82 boys, 95 girls) eighth-grade students (13–14 years old) in advanced mathematics classes. Almost all students (99.3%) scored above the midpoint of the 10-item mathematics identification scale used in Study 1 ($\alpha = .84$). Participants were recruited from five schools in a small urban midwestern community and the surrounding areas: four regular-education public middle schools and one selective-admission public laboratory high school. In these schools 17.7%–73.4% of the students were eligible for free or reduced-price lunch. The ethnic makeup of the sample was 65.0% White, 12.4% Asian, 13.6% African American, 4.4% Latino, and 1.7% other.

Stereotype threat manipulation. In the stereotype threat condition, participants read instructions just before they took the exam that stated, "Please be sure to put an 'X' on the line next to your gender on the cover of the test booklet. This is very important, as boys have done much better than girls on this test in the past." In the no-threat condition, the students were given only instructions pertaining to the test.

A manipulation check was included to ensure that students in the stereotype threat condition were aware of the statement about gender differences. Eighty-seven percent of students indicated that the instructions said that boys performed better on this test in the past, indicating that the manipulation was successful.

Mathematics test. Students were given a mathematics test containing 13 open-ended questions taken from a larger set of Scholastic Assessment Test (SAT) problems ($\alpha = .87$). The problems represented three difficulty levels (based on pilot testing): Four were medium, four were difficult, and five were challenge problems. The questions of each difficulty level were interspersed throughout the test. Students were instructed to show their work and not to use calculators. The problems covered subject areas such as basic algebra, geometry, and arithmetic.

Procedure. Participants were randomly assigned to the stereotype threat condition or the no-threat condition. Then students were told that the test measured their problem-solving ability and were asked to read the instructions on the front cover of the testing packet. Students were given 25 min to complete the mathematics test.

Results and Discussion

To investigate the effects of stereotype threat, we ran a factorial ANOVA. There were three between-subject variables: stereotype threat condition (stereotype threat, no threat), gender (girls, boys), and grade (seventh, eighth). The dependent variable was the percent of mathematics test items answered correctly. Both grades were included in the same analysis because the students took the same mathematics test, although the test was more difficult for the seventh-grade students ($M = 24\%$) than the eighth-grade students ($M = 34\%$).

The two-way interaction between gender and stereotype threat, $F(1, 393) = 2.93$, $p = .09$, $d_{\text{girls}} = 0.03$, $d_{\text{boys}} = -0.04$, and the three-way interaction between gender, stereotype threat, and grade, $F(1, 393) = 1.45$, $p = .23$, $d_{\text{seventhgirls}} = 0.28$, $d_{\text{seventhboys}} = -0.30$, $d_{\text{eighthgirls}} = -0.16$, $d_{\text{eighthboys}} = -0.21$, were not signif-

Table 3
Study 1: Mathematics Scores (Proportion) by Gender and Stereotype Threat Condition

	Girls						Boys					
	Stereotype threat			Stereotype nullification			Stereotype threat			Stereotype nullification		
Grade	*N*	*M*	*SD*	*N*	*M*	*SD*	*N*	*M*	*SD*	*N*	*M*	*SD*
Eighth	58	.64	.15	52	.62	.14	53	.68	.16	49	.68	.18

icant (see Table 4), indicating that there was no stereotype threat effect overall and that it did not differ by grade. Although the interaction between gender and stereotype threat was marginally significant, the pattern was not in the expected direction. We found a significant main effect of grade, $F(1, 393) = 30.04$, $p < .001$, such that eighth-grade students performed better than seventh-grade students, as would be expected. The main effects of gender, $F(1, 393) = 0.32$, $p = .57$, and stereotype threat, $F(1, 393) = 0.70$, $p = .40$, were not significant. The two-way interactions between stereotype threat and grade, $F(1, 393) = 0.92$, $p = .34$, and gender and grade, $F(1, 393) = 0.17$, $p = .68$, were also not significant, showing that gender and stereotype threat condition effects did not differ by grade.

To address the possibility that these very difficult test items did not reveal stereotype threat effects due to a floor effect, we ran the analysis with only the easiest items (50% or more correct; three items; 3.2% of participants incorrectly answered all three problems). This analysis revealed the same pattern: There was no stereotype threat effect ($d_{\text{girls}} = 0.04$).

Overall, these results are similar to those of Study 1: There was also no evidence of a stereotype threat effect in Study 2. However, unlike in Study 1, there was no main effect of gender on mathematics performance.

Study 3

In the third study, a larger range of ages was investigated (fourth-, eighth-, and 12th-grade students). This larger age span can help us understand how stereotype threat effects might vary across ages when students are presented with the same stereotype threat activation. In this study, boys and girls were randomly assigned to stereotype threat and no-threat conditions, and given a mathematics test. An implicit stereotype threat activation method was used (modeled after Muzzatti & Agnoli, 2007).

Method

Participants. Participants were 68 (39 boys, 29 girls) fourth-grade students (9–10 years old), 105 (40 boys, 65 girls) eighth-grade students (13–14 years old), and 145 (69 boys, 76 girls) 12th-grade students (17–18 years old). Participants were recruited from five high-performing (based on state standardized test scores) suburban schools in the Northeast. In these schools, 3.5%–18.6% of the students were eligible for free or reduced-price lunch.

Stereotype threat manipulation. The stereotype threat manipulation was done as part of the introduction to the testing session in the form of a sample mathematics problem. In the stereotype threat condition, the sample problem portrayed a situation in which a much larger number of boys than girls received a mathematics award or were chosen for the mathematics team based on their performance on a mathematics test. For example, eighth graders read the sample word problem stating:

> At the Miller Middle School, the boys were much better at math than the girls. The math teachers chose the 20 students with the highest math test scores for the math team to represent the school at the statewide math competition. Eighteen of the students were boys and two were girls. What proportion of the students on the math team were boys?

In the no-threat condition, students were presented with a sample problem about a topic unrelated to gender or mathematics (i.e., groups of students attending a field trip). For example, eighth graders read:

> At the Miller Middle School, students were invited to participate in a special field trip, but there were only 20 spots available. The teachers chose 18 students from Ms. Fletcher's homeroom and two other students from Ms. Johnson's homeroom. What proportion of the students going on the field trip were from Ms. Fletcher's homeroom?

Students then chose the correct answer to the mathematics problem from among five choices. In the no-threat condition, the students were told that they were going to do some mathematics problems, whereas in the stereotype threat condition students were told they would be taking a mathematics test. This was done to make the mathematics assessment seem more evaluative in the stereotype threat condition, which has been shown to increase stereotype threat effects (Aronson & Steele, 2005; Good & Aronson, 2008; Steele, 1997). The mathematical knowledge required for the sample problem was different for each grade level, to make it age appropriate. At each grade level, the computational task required to solve the sample problem was identical in the two conditions (Muzzatti & Agnoli, 2007).

Mathematics test. The mathematics test was made up of 12 multiple-choice items sampled from the National Assessment of Educational Progress, Trends in International Mathematics and Science Study, and Massachusetts Comprehensive Assessment System mathematics assessments for fourth, eighth, and 12th grades ($\alpha_{\text{fourth}} = .80$, $\alpha_{\text{eighth}} = .67$, $\alpha_{\text{twelfth}} = .64$). Students completed a block of six algebra problems and a block of six geometry and measurement problems. The order of the blocks was counterbalanced across students within each experimental condition. Fourth and eighth graders were given 5 min to complete each section, and 12th graders were given 6 min for each section.

Procedure. Students were randomly assigned to either the stereotype threat or no-threat condition. Each group was tested in

Table 4
Study 2: Mathematics Scores (Proportion) by Grade, Gender, and Stereotype Threat Condition

	Girls						Boys					
	Stereotype threat			No threat			Stereotype threat			No threat		
Grade	*N*	*M*	*SD*	*N*	*M*	*SD*	*N*	*M*	*SD*	*N*	*M*	*SD*
Seventh	55	.28	.19	60	.23	.17	56	.21	.15	52	.26	.18
Eighth	50	.33	.15	49	.36	.22	39	.32	.17	40	.36	.21

a separate room. Instructions, including the sample mathematics problems, were read aloud by the researcher as the students followed along. First, students read the sample word problem that either activated stereotype threat or did not. Next, they answered the mathematics question embedded in the word problem. Then students completed the mathematics test. The testing session took approximately 15–20 min.

Results and Discussion

To examine the effects of stereotype threat, we ran three factorial ANOVAs, one at each grade level. The analyses were run separately for each age group because students completed different mathematics tests at each grade ($M_{\text{fourth}} = 66\%$, $M_{\text{eighth}} = 61\%$, $M_{\text{twelfth}} = 51\%$). Stereotype threat condition (stereotype threat, no threat) and gender (girls, boys) were between-subject variables. The dependent variable was the percent of mathematics items answered correctly.

At all three grade levels, the interaction between gender and stereotype threat was not significant (see Table 5), indicating that there were no stereotype threat effects at any grade level: fourth grade, $F(1, 64) = 0.00$, $p = .99$, $d_{\text{girls}} = 0.17$, $d_{\text{boys}} = 0.28$; eighth grade, $F(1, 101) = 0.27$, $p = .61$, $d_{\text{girls}} = 0.14$, $d_{\text{boys}} = -0.05$; 12th grade, $F(1, 141) = 0.87$, $p = .35$, $d_{\text{girls}} = -0.27$, $d_{\text{boys}} = 0.00$. The main effect of stereotype threat was also not significant at any grade level: fourth grade, $F(1, 64) = 0.69$, $p = .41$; eighth grade, $F(1, 101) = 0.06$, $p = .80$; 12th grade, $F(1, 141) = 0.55$, $p = .46$. There were, however, significant main effects of gender at each grade, with boys outperforming girls in fourth grade, $F(1, 64) = 4.57$, $p = .04$, $d = 0.59$; eighth grade, $F(1, 101) = 6.13$, $p = .02$, $d = 0.51$; and 12th grade, $F(1, 141) = 10.63$, $p = .001$, $d = 0.54$).

A follow-up analysis focusing only on the most difficult items (with less than 50% correct; two items at fourth grade, five items at eighth grade, six items at 12th grade; 37% of fourth graders, 10% of eighth graders, and 7% of 12th graders incorrectly answered all difficult problems) revealed the same pattern of findings. In particular, there was no significant interaction between gender and stereotype threat ($d_{\text{fourthgirls}} = 0.31$, $d_{\text{eighthgirls}} = 0.11$, $d_{\text{twelfthgirls}} = -0.41$). We conducted another follow-up analysis with mathematics-identified students (92% of fourth graders, 80% of eighth graders, 82% of 12th graders). Students were mathematics identified if they scored greater than the midpoint on a five-item mathematics identification scale ($.76 < \alpha\text{s} < .82$). We found no evidence of a stereotype threat effect ($d_{\text{fourthgirls}} = 0.08$, $d_{\text{eighthgirls}} = 0.22$, $d_{\text{twelfthgirls}} = -0.43$) among mathematics-identified students.

Consistent with Studies 1 and 2, Study 3 found no evidence of a stereotype threat effect. As in Study 1, girls underperformed compared to boys in both conditions. One aspect of this study that is different from Studies 1 and 2 is the fact that the stereotype was implicitly activated. The word problem activates the stereotype in a much subtler way than explicitly stating that girls are not as good as boys in mathematics. Although this may make stereotype threat more difficult to induce, some studies with children have found stereotype threat with subtle activation methods (e.g., Ambady et al., 2001; Muzzatti & Agnoli, 2007).

General Discussion

The present work adds to our understanding of stereotype threat effects in children and adolescents. The three studies put girls in a situation where, if stereotype threat effects occur at their age, they would be likely to experience it; however, stereotype threat effects were not found in any of the three studies. Below we discuss our findings in the context of previous research.

Summary of Findings: Past Research and Current Studies

An examination of Table 1, which summarizes results from published studies, unpublished dissertations, and the current three studies, shows how much inconsistency there is in the findings on stereotype threat in children and young adolescents. Note that we could not perform a meta-analysis because of the small number of available empirical investigations and because many studies did not report enough information for us to calculate effect sizes. We encourage researchers to include means, standard deviations, and sample sizes so that studies can be better utilized in future meta-analyses as the research base grows larger.

Instead of a formal meta-analysis, we summarized the results across the literature by examining the percentage of findings (within and across age groups) that revealed stereotype threat effects. We did this in two ways. First, we calculated the percentage of significant results for the individual tests (a total of 36) reported in the literature. In this analysis, if the results showed that stereotype threat effects interacted with another variable, we considered results separately for different levels of that variable. For example, when a study found that mathematics identification was an interaction variable, we examined the results for mathematics-identified and not-identified individuals as two tests. This analysis showed that for early elementary school, three out of six tests showed a stereotype threat effect (50%); for upper elementary school, none of the nine tests (0%) showed an effect; in middle

Table 5
Study 3: Mathematics Scores (Proportion) by Grade, Gender, and Stereotype Threat Condition

	Girls						Boys					
	Stereotype threat			No threat			Stereotype threat			No threat		
Grade	*N*	*M*	*SD*	*N*	*M*	*SD*	*N*	*M*	*SD*	*N*	*M*	*SD*
Fourth	14	.61	.30	15	.56	.29	18	.75	.18	21	.69	.25
Eighth	32	.58	.22	33	.55	.20	18	.67	.24	22	.68	.20
Twelfth	36	.42	.20	40	.48	.24	40	.57	.19	29	.57	.22

school, 4 of the 10 tests (40%) revealed an effect; and at the high school level, 2 out of 11 (18%) tests found stereotype threat effects. Across age groups, 25% of the 36 tests conducted found stereotype threat effects.

The second way we summarized the existing findings is by looking at them at the level of the article rather than at the individual tests included. As indicated earlier, we identified 14 published and unpublished articles that examined the issue of stereotype threat effects in children and adolescents. It should be noted that some of these studies investigated more than one age group, and in this analysis we examined the findings for each age group separately (thus articles in which stereotype threat effects were tested across multiple age groups are included more than once—once for each age group). For the studies examining stereotype threat in early elementary school, three out of four studies (75%) revealed a stereotype threat at least for some students or under certain manipulations. None of the four studies (0%) in upper elementary school, four out of six studies (67%) in middle school, and two out of seven studies (29%) in high school found stereotype threat effects. Overall, 43% of studies found stereotype threat effects (when counting each age group in a study as a separate study).

When comparing published and unpublished articles (excluding the current studies), we find that eight out of 10 (80%) published articles found at least one instance of a stereotype threat effect. Among the published articles, nonsignificant findings were almost always reported in an article along with some significant stereotype threat effects found either at another age (Ambady et al., 2001; Muzzatti & Agnoli, 2007), only with certain students (Keller, 2007), on certain items (Keller, 2007; Neuville & Croizet, 2007), or in certain contexts (Huguet & Regner, 2007, Study 2; Picho & Stephens, 2012; Tomasetto et al., 2011). Importantly, none of the three unpublished dissertations showed a stereotype threat effect. This observation suggests the possibility that publication bias is occurring. Publication bias refers to the fact that studies with null results are often not written up for publication or accepted for publication (Begg, 1994). This bias is a serious concern, especially if these results are being used to make recommendations for interventions.

Potential Explanations for Not Finding Stereotype Threat Effects and Future Directions

We offer two potential explanations for the inconsistency in finding stereotype threat effects. First, it is possible that stereotype threat has a limited effect on children and adolescents and that it takes some specific conditions to elicit this effect. Second, stereotype threat may be always present for school-age girls unless mitigated, and thus may affect performance in both conditions, regardless of the experimental manipulation. In discussing these potential reasons, we will also discuss directions for future research in each area.

Potential that stereotype threat has a limited effect. It is possible that stereotype threat manifests itself only under specific conditions. However, it is unclear exactly what those conditions are, because in many cases researchers have incorporated the factors currently known to induce stereotype threat and have not found an effect. In future research, it is critical to determine whether there are particular factors that could reliably produce stereotype threat effects in children and adolescents. Once these factors are identified, we can focus on them when designing interventions.

Activation methods. One important issue is the particular experimental manipulation employed. At present, studies with children and adolescents have reported a variety of methods (see Table 2) ranging from subtle, implicit manipulations aimed at activating students' gender awareness (e.g., by having them mark their gender or asking them gender-related questions before taking the test) to more explicit, almost blatant, ways of activating the stereotype (e.g., by telling students prior to taking the test that boys do better than girls on this test). The studies reported here used three stereotype activation methods, all of which were designed based on the findings of past research, and none of these produced stereotype threat effects.

To understand better the stereotype threat manipulations that may provoke stereotype threat effects in children and adolescents, future research could benefit from systematically testing different activation methods in the same sample. Large-scale studies that use multiple stereotype manipulations (and a no-threat condition), keeping all other experimental conditions equal, would help to tease apart which activation methods are most likely to lead to stereotype threat effects during childhood and adolescence.

Study populations. In addition to the particular stereotype threat activation method, it is important to consider the characteristics of the population being studied. Although past research with adult women has found that mathematics identification is an important factor, some studies with mathematics-identified children and adolescents, including the present studies, still do not find effects (e.g., Dinella, 2004). Perhaps other factors are also important. For example, Tomasetto et al. (2011) found that stereotype threat impacted the mathematics performance of girls whose mothers held neutral gender stereotypes about mathematics but not girls whose mothers rejected the gender stereotype about mathematics. It may be useful to follow up this finding to determine whether mothers' characteristics predict their daughters' behavior under stereotype threat (i.e., if we take a sample of children and collect information about their mothers' stereotyping, we should be able to predict how the children will respond to stereotype threat).

Potential that stereotype threat effects are always occurring. In two of the three studies reported here, we found gender differences on the mathematics tests regardless of condition. Thus, although we did not observe a difference in girls' performance between the stereotype threat and no-threat conditions, there is still a possibility, pointed out by some researchers (Smith & White, 2002; Steele, 1997), that stereotype threat impairs girls' performance in any mathematics testing situation. Perhaps the testing situation in general activates stereotype threat that occurs in everyday testing environments, regardless of the added manipulation.

If stereotype threat effects occur all the time, it should be the case that girls who are in a stereotype nullification condition do better than girls in both stereotype threat and no-threat conditions (see Smith & White, 2002). To date, only four studies, including Study 1 of the present article, have used a stereotype threat nullification condition with children or adolescents (Cruz-Duran, 2009; Keller, 2007; Keller & Dauenheimer, 2003). In the present Study 1, girls in the stereotype nullification condition did not perform better than girls in the stereotype threat condition, which potentially raises questions about the idea of whether stereotype threat occurs all the time. One of the prior studies with adolescents found a positive effect of the stereotype threat nullification condition compared to a stereotype threat condition (Keller, 2007), but two did not (Cruz-Duran, 2009; Keller & Dauenheimer, 2003). Due to these inconsistent findings, future research should include a stereotype nullification condition in addition to the stereotype threat and no-threat conditions to provide better understanding of the nature of this phenomenon.

This explanation is somewhat at odds with the fact that girls earn similar or better mathematics classroom grades, which are made up at least in part by test scores (Corbett et al., 2008). If girls are earning similar or better grades, could they be suffering from stereotype threat in every testing situation? There are two possible ways that these ideas could coexist. First, because stereotype threat only affects performance on difficult items, and classroom tests may not contain a large portion of difficult items, there may be no stereotype threat effects on these tests. Second, it is possible that girls perform more poorly on class mathematics tests than boys, however teachers take other factors into account when assigning grades (e.g., homework completion, effort), which could lead to girls obtaining similar or better grades despite poorer test performance.

Conclusion

Taken together, the findings from published research, unpublished articles, and the present studies reveal inconsistency in the effects of stereotype threat on girls' mathematics performance. The discrepancy in results from published and unpublished studies suggests publication bias, which may create an inaccurate picture of the phenomenon. A recent review suggests that this publication bias may also be an issue in the literature on stereotype threat in adult women (Stoet & Geary, 2012). Overall, these results raise the possibility that stereotype threat may not be the cause of gender differences in mathematics performance prior to college. Although we feel that more nuanced research needs to be done to truly understand whether stereotype threat impacts girls' mathematics performance, we also believe that too much focus on this one explanation may deter researchers from investigating other key factors that may be involved in gender differences in mathematics performance. For example, there are a number of factors (e.g., mathematics anxiety, mathematics interest, spatial skills; see Ceci & Williams, 2010) that have been shown to be consistently related to mathematics performance and mathematics- and science-related career choices and may warrant more research attention than does stereotype threat.

References

Ambady, N., Shih, M., Kim, A., & Pittinsky, T. L. (2001). Stereotype susceptibility in children: Effects of identity activation on quantitative performance. *Psychological Science, 12,* 385–390. doi:10.1111/1467-9280.00371

Aronson, J., & Good, C. (2003). The development and consequences of stereotype vulnerability in adolescents. In F. Pajares & T. Urdan (Eds.), *Adolescence and education: Vol. 2. Academic motivation of adolescents* (pp. 299–330). Greenwich, CT: Information Age.

Aronson, J., & Steele, C. M. (2005). Stereotypes and the fragility of human competence, motivation, and self-concept. In A. J. Elliot & C. S. Dweck (Eds.), *Handbook of competence and motivation* (pp. 436–456). New York, NY: Guilford Press.

Begg, C. B. (1994). Publication bias. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 399–409). New York, NY: Russell Sage Foundation.

Benbow, C. P., & Stanley, J. C. (1980). Sex differences in mathematical ability: Fact or artifact? *Science, 210,* 1262–1264. doi:10.1126/science.7434028

Ceci, S. J., & Williams, W. M. (2010). Sex differences in math-intensive fields. *Current Directions in Psychological Science, 19,* 275–279. doi:10.1177/0963721410383241

College Board. (2009). *SAT percentile ranks for males, females and total group*. Retrieved from http://professionals.collegeboard.com/data-reports-research/sat/data-tables

College Board. (2010). *AP summary reports: National report*. Retrieved from http://www.collegeboard.com/student/testing/ap/exgrd_sum/2010.html

Corbett, C., Hill, C., & St. Rose, A. (2008). *Where the girls Are: The facts about gender equity in education. Executive summary*. Washington, DC: American Association of University Women Educational Foundation.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16,* 297–334. doi:10.1007/BF02310555

Cruz-Duran, E. (2009). *Stereotype threat in mathematics: Female high school students in all-girl and coeducational schools*. Available from ProQuest Dissertations and Theses. (UMI No. 3365692)

Cvencek, D., Meltzoff, A. N., & Greenwald, A. G. (2011). Math–gender stereotypes in elementary school children. *Child Development, 82,* 766–779. doi:10.1111/j.1467-8624.2010.01529.x

Dinella, L. M. (2004). *A developmental perspective on stereotype threat and high school mathematics* (Doctoral dissertation). Available from ProQuest Dissertations and Theses. (UMI No. 3135245)

Ding, C. S., Song, K., & Richardson, L. J. (2006). Do mathematical gender differences continue? A longitudinal study of gender difference and excellence in mathematics performance in the U.S. *Educational Studies, 40,* 279–295. doi:10.1080/00131940701301952

Dweck, C. (2008). The Brainology program: Cultivate a growth mindset. Retrieved from http://www.mindsetworks.com

Eccles, J. S. (1987). Gender roles and women's achievement-related decisions. *Psychology of Women Quarterly, 11,* 135–172. doi:10.1111/j.1471-6402.1987.tb00781.x

Eccles, J. S., & Jacobs, J. E. (1986). Social forces shape math attitudes and performance. *Signs, 11,* 367–380. doi:10.1086/494229

Entwisle, D., Alexander, K., & Olson, L. (1994). The gender gap in math: Its possible origins in neighborhood effects. *American Sociological Review, 59,* 822–838. doi:10.2307/2096370

Forbes, C. E., Schmader, T., & Allen, J. J. B. (2008). The role of devaluing and discounting in performance monitoring: A neurophysiological study of minorities under threat. *Social Cognitive and Affective Neuroscience, 3,* 253–261. doi:10.1093/scan/nsn012

Friedman, S. J., & Frisbie. D. A. (1995). The influence of report cards on the validity of grades reported to parents. *Educational and Psychological Measurement, 55,* 5–26. doi:10.1177/0013164495055001001

Fryer, R. G., & Levitt, S. D. (2010). An empirical analysis of the gender gap in mathematics. *American Economic Journal: Applied Economics, 2,* 210–240. doi:10.1257/app.2.2.210

Geary, D. C. (1996). Sexual selection and sex differences in mathematical abilities. *Behavioral and Brain Sciences, 19,* 229–284. doi:10.1017/S0140525X00042400

Gibbs, B. G. (2010). Reversing fortunes or content change? Gender gaps in math-related skill throughout childhood. *Social Science Research, 39,* 540–569. doi:10.1016/j.ssresearch.2010.02.005

Gonzales, P., Williams, T., Jocelyn, L., Roey, S., Kastberg, D., & Brenwald, S. (2008). *Highlights from TIMSS 2007: Mathematics and science* *achievement of U.S. fourth- and eighth-grade students in an international context* (NCES 2009-001 rev.). Washington, DC: National Center for Education Statistics.

Good, C. D. (2001). *The development of stereotype threat and its relation to theories of intelligence: Effects on elementary school girls' mathematics achievement and task choices*. Available from ProQuest Dissertations and Theses. (UMI No. 3008343)

Good, C., & Aronson, J. (2008). The development of stereotype threat: Consequences for educational and social equality. In C. Wainryb, J. G. Smetana, & E. Turiel (Eds.), *Social development, social inequalities, and social justice* (pp. 155–183). New York, NY: Erlbaum.

Halpern, D. F., & Tan, U. (2001). Stereotypes and steroids: Using a psychobiosocial model to understand cognitive sex differences. *Brain and Cognition, 45,* 392–414. doi:10.1006/brcg.2001.1287

Hedges, L. V., & Friedman, L. (1993). Gender differences in variability in intellectual abilities: A reanalysis of Feingold's results. *Review of Educational Research, 63,* 94–105. doi:10.3102/00346543063001094

Heller, K. A., & Ziegler, A. (1996). Gender differences in mathematics and the sciences: Can attributional retraining improve the performance of gifted females? *Gifted Child Quarterly, 40,* 200–210. doi:10.1177/001698629604000405

Huguet, P., & Regner, I. (2007). Stereotype threat among schoolgirls in quasi-ordinary classroom circumstances. *Journal of Educational Psychology, 99,* 545–560. doi:10.1037/0022-0663.99.3.545

Huguet, P., & Regner, I. (2009). Counter-stereotypic beliefs in math do not protect school girls from stereotype threat. *Journal of Experimental Social Psychology, 45,* 1024–1027. doi:10.1016/j.jesp.2009.04.029

Hyde, J. S. (2005). The gender similarities hypothesis. *American Psychologist, 60,* 581–592. doi:10.1037/0003-066X.60.6.581

Hyde, J. S., Lindberg, S. M., Linn, M. C., Ellis, A. B., & Williams, C. C. (2008). Gender similarities characterize math performance. *Science, 321,* 494–495. doi:10.1126/science.1160364

Hyde, J. S., & Linn, M. C. (2006). Gender similarities in mathematics and science. *Science, 314,* 599–600. doi:10.1126/science.1132154

Jamieson, J. P., & Harkins, S. G. (2007). Mere effort and stereotype threat performance effects. *Journal of Personality and Social Psychology, 93,* 544–564. doi:10.1037/0022-3514.93.4.544

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,* 323–338. doi:10.1348/000709906X113662

Keller, J., & Dauenheimer, D. (2003). Stereotype threat in the classroom: Dejection mediates the disrupting threat effect on women's math performance. *Personality and Social Psychology Bulletin, 29,* 371–381. doi:10.1177/0146167202250218

Levy, G. D., & Carter, B. D. (1989). Gender schema, gender constancy, and gender-role knowledge: The roles of cognitive factors in preschoolers' gender-role stereotype attributions. *Developmental Psychology, 25,* 444–449. doi:10.1037/0012-1649.25.3.444

Lindberg, S. M., Hyde, J. S., Petersen, J. L., & Linn, M. C. (2010). New trends in gender and mathematics performance: A meta-analysis. *Psychological Bulletin, 136,* 1123–1135. doi:10.1037/a0021276

McGraw, R., Lubienski, S. T., & Strutchens, M. E. (2006). A closer look at gender in NAEP mathematics achievement and affect data: Intersection with achievement, race/ethnicity, and socioeconomic status. *Journal for Research in Mathematics Education, 37,* 129–150.

Muzzatti, B., & Agnoli, F. (2007). Gender and mathematics: Attitudes and stereotype threat susceptibility in Italian children. *Developmental Psychology, 43,* 747–759. doi:10.1037/0012-1649.43.3.747

Neuville, E., & Croizet, J. (2007). Can salience of gender identity impair math performance among 7–8 years old girls? The moderating role of task difficulty. *European Journal of Psychology of Education, 22,* 307–316. doi:10.1007/BF03173428

Nguyen, H. H., & 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,* 1314–1334. doi:10.1037/a0012702

Nicholls, J. G. (1978). The development of the concepts of effort and ability, perception of academic attainment, and the understanding that difficult tasks require more ability. *Child Development, 49,* 800–814. doi:10.2307/1128250

Nuttall, R. L., Casey, M. B., & Pezaris, E. (2005). Spatial ability as a mediator of gender differences on mathematics tests: A biological–environmental framework. In A. M. Gallagher & J. C. Kaufman (Eds.), *Gender differences in mathematics: An integrative psychological approach* (pp. 121–142). New York, NY: Cambridge University Press.

O'Brien, L. T., & Crandall, C. S. (2003). Stereotype threat and arousal: Effects on women's math performance. *Personality and Social Psychology Bulletin, 29,* 782–789. doi:10.1177/0146167203029006010

Ornstein, A. C. (1994). Grading practices and policies: An overview and some suggestions. *NASSP Bulletin, 78,* 55–64. doi:10.1177/019263659407856111

Penner, A. M., & Paret, M. (2008). Gender differences in mathematics achievement: Exploring the early grades and the extremes. *Social Science Research, 37,* 239–253. doi:10.1016/j.ssresearch.2007.06.012

Picho, K., & Stephens, J. M. (2012). Culture, context and stereotype threat: A comparative analysis of young Ugandan women in coed and single-sex schools. *Journal of Educational Research, 105,* 52–63. doi:10.1080/00220671.2010.517576

Pomerantz, E. M., Altermatt, E. R., & Saxon, J. L. (2002). Making the grade but feeling distressed: Gender differences in academic performance. *Journal of Educational Psychology, 94,* 396–404. doi:10.1037/0022-0663.94.2.396

Quinn, D. N., & Spencer, S. J. (2001). The interference of stereotype threat with women's generation of mathematical problem-solving strategies. *Journal of Social Issues, 57,* 55–71. doi:10.1111/0022-4537.00201

Robinson, J. P., & Lubienski, S. T. (2011). The development of gender achievement gaps in mathematics and reading during elementary and middle school: Examining direct cognitive assessments and teacher ratings. *American Educational Research Journal, 48,* 268–302. doi:10.3102/0002831210372249

Scarr, S., & Carter-Saltzman, L. (1982). Genetics and intelligence. In R. J. Sternberg (Ed.), *Handbook of human intelligence* (pp. 792–896). Cambridge, England: Cambridge University Press.

Schmader, T., & Johns, M. (2003). Converging evidence that stereotype threat reduces working memory capacity. *Journal of Personality and Social Psychology, 85,* 440–452. doi:10.1037/0022-3514.85.3.440

Schmader, T., Johns, M., & Forbes, C. (2008). An integrated process model of stereotype threat effects on performance. *Psychological Review, 115,* 336–356. doi:10.1037/0033-295X.115.2.336

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,* 1040–1057. doi:10.1177/00131640121971635

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,* 179–191. doi:10.1023/A:1021051223441

Spelke, E. S. (2005). Sex differences in intrinsic aptitude for mathematics and science? A critical review. *American Psychologist, 60,* 950–958. doi:10.1037/0003-066X.60.9.950

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

Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist, 52,* 613–629. doi: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,* 797–811. doi:10.1037/0022-3514.69.5.797

Stoet, G., & Geary, D. C. (2012). Can stereotype threat explain the gender gap in mathematics performance and achievement? *Review of General Psychology, 16,* 93–102. doi:10.1037/a0026617

Strand, S., Deary, I. J., & Smith, P. (2006). Sex differences in Cognitive Abilities Test scores: A UK national picture. *British Journal of Educational Psychology, 76,* 463–480. doi:10.1348/000709905X50906

Stricker, L. J., & Ward, W. C. (2004). Stereotype threat, inquiring about test takers' ethnicity and gender, and standardized test performance. *Journal of Applied Social Psychology, 34,* 665–693. doi:10.1111/j.1559-1816.2004.tb02564.x

Tomasetto, C., Alparone, F. R., & Cadinu, M. (2011). Girls' math performance under stereotype threat: The moderating role of mothers' gender stereotypes. *Developmental Psychology, 47,* 943–949. doi:10.1037/a0024047

Received April 30, 2012
Revision received October 22, 2012
Accepted October 25, 2012 ■