Journal of Educational Psychology

0022-0663/14/$12.00 http://dx.doi.org/10.1037/edu0000005

Women's Representation in Science Predicts National Gender-Science Stereotypes: Evidence From 66 Nations

David I. Miller and Alice H. Eagly
Northwestern University

Marcia C. Linn
University of California, Berkeley

In the past 40 years, the proportion of women in science courses and careers has dramatically increased in some nations but not in others. Our research investigated how national differences in women's science participation related to gender-science stereotypes that associate science with men more than women. Data from ~350,000 participants in 66 nations indicated that higher female enrollment in tertiary science education (community college or above) related to weaker explicit and implicit national gender-science stereotypes. Higher female employment in the researcher workforce related to weaker explicit, but not implicit, gender-science stereotypes. These relationships remained after controlling for many theoretically relevant covariates. Even nations with high overall gender equity (e.g., the Netherlands) had strong gender-science stereotypes if men dominated science fields specifically. In addition, the relationship between women's educational enrollment in science and implicit gender-science stereotypes was stronger for college-educated participants than participants without college education. Implications for instructional practices and educational policies are discussed.

Keywords: diversity, gender, science education, science workforce, stereotypes

Supplemental materials: http://dx.doi.org/10.1037/edu0000005.supp

Pervasive stereotypes associating science with men emerge early in development (Chambers, 1983; Steffens, Jelenec, & Noack, 2010) and exist across cultures (Nosek et al., 2009). Over 40 years ago, Chambers asked nearly 5,000 American and Canadian children to a draw a picture of a scientist, and only 28 children (0.6%) depicted a woman scientist. Although most children still associate science with men, these associations may have weakened over time at least in the United States (Fralick, Kearn, Thompson, & Lyons, 2009; Milford & Tippett, 2013). For example, in one recent study (Farland-Smith, 2009), 35% of American children depicted a woman scientist. These changes in stereotypes mirror women's increasing participation in science in the United States (Hill, Corbett, & St. Rose, 2010). For instance, women earned 19% of the U.S.'s chemistry bachelor's degrees in 1966 but now earn 49% of such degrees (National Science Board, 2014). To investigate how women's national participation in science relates to such associations, our analyses used cross-sectional data from ~350,000 participants in 66 nations. These individuals completed measures of *gender-science stereotypes*, defined as associations that connect science with men more than women. Comparing these stereotypes across nations could help identify how they are shaped by several interacting sociocultural factors. Such sociocultural factors could include messages in mass media; opinions of teachers and peers; participation of family members in science, technology, engineering, and mathematics (STEM) fields; and/or experiences learning STEM topics in male-dominated courses.

Eagly and colleagues' social role theory (Eagly & Wood, 2012; Wood & Eagly, 2012) provides a framework for understanding how gender stereotypes form and change in response to observing women and men in differing social roles within a culture. Both direct (e.g., through social interactions) and indirect (e.g., through mass media) observations associate social groups such as women and men with their typical role-linked activities and thus form the basis for cultural stereotypes (Koenig & Eagly, in press). These observations begin at early ages. For instance, kindergarten girls endorsed gender-mathematics stereotypes if their female teacher was anxious about mathematics (Beilock, Gunderson, Ramirez, & Levine, 2010; Gunderson, Ramirez, Levine, & Beilock, 2012). In contrast, exposure to successful women scientists and mathematicians can weaken gender-STEM stereotypes among young girls (Galdi, Cadinu, & Tomasetto, 2014), high school students taking biology (Mason, Kahle, & Gardner, 1991), or undergraduate female STEM majors who identify with their professor (Young,

David I. Miller and Alice H. Eagly, Department of Psychology, Northwestern University; Marcia C. Linn, Department of Psychology, University of California, Berkeley.

This material is based upon work supported by National Science Foundation Graduate Research Fellowship Grant No. DGE-0824162, awarded to David I. Miller, and by DRL-08222388: Cumulative Learning using Embedded Assessment Results (CLEAR), awarded to Marcia C. Linn. Figure 1 was created using the software StatWorld. We especially thank Frederick Smyth and Brian Nosek for generously providing this study's stereotype data and answering our questions about their previous analysis. We also thank Douglas Medin, Erin Maloney, Galen Bodenhausen, Jennifer Richeson, Mesmin Destin, and Sian Beilock for comments on earlier drafts of this article.

Correspondence concerning this article should be addressed to David I. Miller, Department of Psychology, Northwestern University, Swift Hall 102, 2029 Sheridan Road, Evanston, IL 60208. E-mail: david.isaac.miller@gmail.com

Rudman, Buettner, & McLean, 2013). Hence, stereotypes are formed and changed, in part, by repeatedly observing members of different social groups in role-linked activities. This theoretical framework can also help to explain why stereotypes about other social groups vary across nations. For instance, consistent with social role theory, stereotypes about older adults' incompetence were weaker in nations where more older adults participated in paid and volunteer work; this cross-national relationship remained even after controlling for national differences in older adults' cognitive abilities (Bowen & Skirbekk, 2013).

Multiple observations of counterstereotypic women across diverse contexts, such as directly in science courses and indirectly in televisions shows, are critical to changing stereotypes (Eagly & Wood, 2012; Koenig & Eagly, in press; Wood & Eagly, 2012). People need multiple, mutually reinforcing examples to see counterstereotypic individuals as evidence of trends. Otherwise, sparse counterstereotypic examples can be dismissed as atypical through a process called *subtyping* (Bigler & Liben, 2006; Richards & Hewstone, 2001). For instance, individual women scientists could be perceived as having followed unusual paths to science and exerted exceptional effort to succeed (Smith, Lewis, Hawthorne, & Hodges, 2013). These stereotyping processes may explain why experimental studies have revealed that exposure to successful women engineers and mathematicians have not consistently weakened gender-STEM stereotypes (Ramsey, Betz, & Sekaquaptewa, 2013; Steinke et al., 2007; Stout, Dasgupta, Hunsinger, & McManus, 2011; Young et al., 2013). For instance, in Stout et al.'s Study 3, intended STEM majors ($n = 100$, 47% women) took a 3-month calculus class from a professor and teaching assistant who were either both male or both female. Although taking the calculus course from female instructors increased female students' implicit identification with mathematics, the gender of the course instructors had no observable effect on gender-math stereotypes. Such short-term interventions may be insufficient to override pervasive, everyday experiences linking math-intensive science fields with men. For instance, male students outnumbered female students by three to one in the calculus course taken by Stout et al.'s participants. In such contexts, sparse examples of female math professors may have been subtyped and seen as atypical. Moreover, taking a STEM course from a female rather than male professor can even strengthen gender-science stereotypes if students do not view the professor as similar to themselves (Young et al., 2013).

Even students in female-dominated science majors could still strongly associate science with men. For instance, although women currently earn 60% of biology bachelor's degrees in the United States (National Science Board, 2014), biology majors would likely encounter other stereotype-consistent evidence. This evidence could include the preponderance of men among biology faculty (Ceci, Williams, & Barnett, 2009) or students in required STEM courses in other fields such as physics (Barone, 2011). Moreover, students could form separate stereotypes about biologists while maintaining their belief that science is generally associated with men (Richards & Hewstone, 2001). Such conflicting experiences suggest that gender-science stereotypes would likely vary in nuanced ways across students' field of study. For instance, one large correlational study ($n \sim 100{,}000$) revealed that, compared with physical science majors, biological science majors reported weaker explicit gender-science stereotypes but still implicitly associated science with men to the same extent (Smyth & Nosek, 2013). Furthermore, pervasive cultural images associating science with men fuel stereotyping processes for students in all academic disciplines. Archetypes of White male scientists are present in diverse cultural artifacts such as television shows (Long et al., 2010), movies (Flicker, 2003), national news reports (Chimba & Kitzinger, 2010; Shachar, 2000), science textbooks (Bazler & Simonis, 1991; Brotman & Moore, 2008), and even advertisements in the journal *Science* (Barbercheck, 2001). Such shared cultural experiences likely disseminate and reinforce stereotypes about gender in general (Furnham & Paltzer, 2010; Kimball, 1986) and women in science specifically (Steinke, 2013).

Comparing gender-science stereotypes across nations could help reveal the impact of such varied cultural experiences. In one such effort, Nosek et al. (2009) found that nations with stronger implicit gender-science stereotypes also had larger national gender differences favoring boys in science and mathematics achievement. The authors suggested that this result reflected a bidirectional relationship in which stereotypes influence achievement and achievement influences stereotypes. We built on this prior research by investigating how women's participation in science relates to cross-national differences in gender-science stereotypes. Our focus on participation in science extends Nosek et al.'s study because women's participation in science does not necessarily reflect gender differences in science achievement (Riegle-Crumb, King, Grodsky, & Muller, 2012). When more women enter science, people can observe counterstereotypic women across diverse contexts such as in science classes and news articles, especially if these changes occur across multiple science fields. These diverse observations can then influence stereotypes, as predicted by social role theory (Eagly & Wood, 2012; Wood & Eagly, 2012). To test these predictions, our study analyzed two aspects of women's participation in science: percentage of women among (a) all science majors (community college or above) and (b) employed researchers. Many participants in our mostly college-educated sample likely had direct repeated exposure to women and men enrolled as science majors; direct exposure to employed researchers was perhaps more limited.

We investigated how women's participation in science related to both implicit and explicit measures of gender-science stereotypes. Consistent with contemporary theorizing about dual processes in social cognition (Sherman, Gawronski, & Trope, 2014), the implicit measure assessed aspects of stereotyping that are generally more automatic and less conscious, whereas the explicit measure assessed those aspects that emerge as conscious knowledge that is willingly reported (Nosek, Hawkins, & Frazier, 2011). Empirical findings have generally supported the interpretation that these measures assess related, but distinct, constructs. For instance, explicit and implicit attitude measures often significantly, but weakly, correlate with each other (Greenwald, Poehlman, Uhlmann, & Banaji, 2009; Nosek et al., 2007). Moreover, both measures often add incremental validity when predicting behavioral outcomes such as discrimination (Greenwald et al., 2009; but see Oswald, Mitchell, Blanton, Jaccard, & Tetlock, 2013).

Gawronski and Bodenhausen's (2006, 2011) associative-propositional model provides a theoretical account for why explicit and implicit measures should often differ. According to this model, implicit measures reflect the activations of associations in memory, whereas explicit measures reflect the outcomes of propositional processes. For instance, a person could automatically associate Black people with negative attributes such as violent crime but reject the proposition that "I dislike Black people." That person would therefore show negative bias toward Black people on an implicit attitude measure, but not on an explicit measure. Also consistent with this theoretical model, different types of counterstereotypic exposure may be necessary to change implicit versus explicit stereotypes. Specifically, repeated counterstereotypic exposure would be critical to changing implicit stereotypes, which reflect associations learned from repeated pairings of stimuli representing two concepts (e.g., *science* and *male*). In contrast, brief exposure to propositional information (e.g., statistics about women's representation in science) could change explicit stereotypes. For instance, a person could learn that women earn half of the U.S.'s chemistry bachelor's degrees (National Science Board, 2014) and readily incorporate that information into explicit responses (e.g., answering a questionnaire item asking how much that person associates chemistry with men or women).

To explore these ideas, we analyzed four relationships between gender-science stereotypes and women's participation in science by crossing two types of women's participation (in educational enrollment and in the workforce) with two types of gender-science stereotypes (explicit and implicit). Our critical hypothesis was that a higher participation of women in science would relate to weaker national-level gender-science stereotypes, consistent with social role theory. The associative-propositional model would additionally predict that, compared with explicit stereotypes, implicit stereotypes should relate more strongly to repeated counterstereotypic exposure. As a proxy for this repeated exposure to women in STEM fields, we used participants' level of education (e.g., college-educated vs. some or no college). In nations with a high percentage of women among science majors, college-educated individuals would have frequently encountered examples of female science majors during college.

Method

Sample

The 66 nations included in our focal analyses (see Figure 1) represented ~350,000 participants who self-selected into our sample by completing stereotype measures on a widely distributed website called Project Implicit (see Nosek et al., 2009). These nations met the requirements of (a) a minimum sample size of $n > 50$ and (b) populations of more than 5% Internet users during the time of stereotype data collection (years 2000–2008). The Results section explains the rationale for these selection criteria and reports results across alternate criteria. In an average national sample, 50% of participants had a college degree or higher, and 79% had some college or higher. Therefore, most participants likely had direct, repeated exposure to the representation of women among college science majors. Also, in an average national sample, 60% of participants were women, and the average age was 27 years ($SD = 11$ years within nations).

Measures

Explicit gender-science stereotypes. For the explicit stereotype measure, participants rated "how much you associate science with males or females" on a 5-point or 7-point scale[1] ranging from *strongly male* to *strongly female*. This same question was repeated replacing "science" with "liberal arts" to serve as a comparison measure of stereotypes in an alternate academic domain. These questions were worded to correspond to the implicit measure (see below) and definition of gender-science stereotypes (i.e., associations connecting science with men more than women). These questions therefore did not ask about gender stereotypes regarding science-related abilities and interests (e.g., "Do you think males or females are more interested in science?"); such wording would have addressed gender stereotypes about science-related attributes rather than participants' more general associations between science and gender.

Single-item measures such as our study's explicit measure sometimes have lower reliabilities than multiple-item measures and therefore can underestimate relationships. Hence, to the extent that our explicit measure was unreliable, it would have provided conservative tests of hypotheses regarding explicit stereotypes. However, compared with multiple-item measures, single-item measures often have equal reliability and validity for assessing psychosocial constructs such as attitudes (Bergkvist & Rossiter, 2007; Fishbein & Ajzen, 1974), job satisfaction (Wanous, Reichers, & Hudy, 1997), and math anxiety (Núñez-Peña, Guilera, & Suárez-Pellicioni, in press).

Implicit gender-science stereotypes. For the implicit measure, participants completed a gender-science Implicit Association Test (IAT; for an overview of the IAT methodology, see Greenwald et al., 2009). As described by Nosek et al. (2009), this computerized task recorded how quickly participants associated science with males. Participants categorized words representing the categories of *male* (boy, father, grandpa, husband, male, man, son, uncle), *female* (aunt, daughter, female, girl, grandma, mother, wife, woman), *science* (astronomy, biology, chemistry, engineering, geology, math, physics, math), and *liberal arts* (arts, English, history, humanities, literature, music, philosophy). These 30 words were presented one at a time, and participants categorized them by pressing one of two keyboard keys; one response key was on the left side of the keyboard and the other was on the right. The response keys were paired stereotypically for some trials (e.g., participant presses the *e* key for *male* and *science* words, and *i* key for *female* and *liberal arts* words) and counterstereotypically for other trials (e.g., participant presses the *e* key for *female* and *science* words). Participants responded faster when the keys were paired stereotypically than counterstereotypically by an average of ~100–150 milliseconds (Nosek, Banaji, & Greenwald, 2002). This response time difference was interpreted as evidence of implicit gender-science stereotypes.

Participants were given unlimited time to make a response for each word, but were instructed to go as fast as possible. The precision of these reaction times was limited by the clock rates of

[1] The response categories for the 5-point scale (*strongly male*, *somewhat male*, *neither male nor female*, *somewhat female*, *strongly female*) and the 7-point scale (*strongly male*, *moderately male*, *somewhat male*, *neither male nor female*, *somewhat female*, *moderately male*, *strongly female*) were similar. These response categories were converted to a numeric scale by assigning *neither male nor female* to a value of 0 and assuming equal numeric spacing between the ordinal response categories. *Male* responses were given positive scores, and *female* responses were given negative scores. We standardized the variances of 5-point and 7-point scales to both be 1 before using the scales to compute national averages.

Figure 1. Nations analyzed (shown in black) by the criteria of $n > 50$ responses per nation and > 5% Internet user population.

participants' computers; this limitation introduced some random noise into the implicit measure, but no large systematic biases (Nosek, Greenwald, & Banaji, 2005). Each participant completed a block of 60 stereotype-consistent trials and a block of 60 stereotype-inconsistent trials. The ordering of stereotype-consistent and stereotype-inconsistent blocks can have weak to moderate effects on the magnitude of implicit bias (Nosek et al., 2005, Study 4). The ordering of these blocks was therefore counterbalanced across participants. Before completing these critical blocks, participants completed a practice block of 20 trials that involved categorizing only *male* and *female* words and then another practice block of 20 trials that involved categorizing only *science* and *liberal arts* words. These practice blocks helped participants become familiar with the IAT, consistent with standard practices for administering this task (Nosek et al., 2005).

We used the exact same data cleaning procedures used by Greenwald, Nosek, and Banaji (2003) and Nosek et al. (2009) to process the IAT data. Individual trial response times faster than 400 ms or slower than 10,000 ms were removed. Response times for trials with errors (i.e., participant presses the wrong response key for the presented word) was replaced with the mean of correct responses in that response block plus a 600-ms penalty. To help minimize the impact of careless responding, participants' IAT scores were disqualified if participants consistently made many errors (i.e., made errors on more than 30% of trials across all the critical blocks, 40% of trials in any one of the critical blocks, 40% of trials across all the practice blocks, and/or 50% of trials in any one of the practice blocks) or consistently responded too quickly (i.e., responded faster than 300 ms on more than 10% of the total test trials, 25% of trials in any one of the critical blocks, 35% of trials in any one of the practice blocks). These data quality standards disqualified 9% of IAT scores. The reaction time difference between stereotype-consistent and stereotype-inconsistent blocks was divided by each individual's standard deviation of reaction times to compute an *IAT D* score (Greenwald et al., 2003).

Scoring of stereotype measures. For both explicit and implicit stereotype measures, positive scores indicated male–science associations, negative scores indicated female–science associations, and scores of 0 indicated neutral gender–science associations (e.g., an explicit response of "neither male nor female"). To facilitate comparison across the two stereotype measures, each measure's raw scores were standardized by dividing by the standard deviation of all individual scores across the globe. These standardized scores are identical to *z*-scores if *z*-scores were computed without first subtracting the population mean. Hence, for both stereotype measures, a standardized score of 0.5 represented a response that differed 0.5 standard deviations in the male direction from neutral gender–science associations, with standard deviation representing variability across individuals. This approach has the advantage that the magnitude of stereotypes can be interpreted in Cohen's *d* effect size units (for an example meta-analytic application, see Koenig, Eagly, Mitchell, & Ristikari, 2011, masculinity-femininity paradigm). Hence, national averages exceeding 0.5 can be considered moderate to large.

Women's representation in science. Two indicators of women's representation in science were downloaded from UNESCO's website (stats.uis.unesco.org): the percentage of women among individuals (a) enrolled in tertiary science education and (b) employed as researchers. Both indicators were based on head counts. Statistics by field of science (e.g., life vs. physical sciences) were generally less available. The composite measure for women's representation in the researcher workforce combined statistics across sectors of employment: business enterprise, government, higher education, and private nonprofit. Although this measure aggregated researcher statistics across many fields, the composite measure correlated highly with the specific, but less available, measure for natural sciences ($r = .86$, $p < .0001$, $n = 28$). Our central results were similar when using the aggregated or disaggregated measure. Consistent with prior analyses (Else-Quest, Hyde, & Linn, 2010; Reilly, 2012), we therefore focused on the more available, aggregated statistics to maximize both statistical power and the diversity of nations in our analyses. We averaged all available statistics for the years of stereotype data collection (2000–2008), or if those data were not available, then for the 4 years before and after data collection.

Other national indicators. In addition to using women's representation in science to predict gender-science stereotypes, multiple regression analyses included 25 other national attributes as covariates. These covariates included broad and domain-specific indicators of gender equity, gender differences in science achievement, Hofstede's cultural dimensions, human development, prevalence of scientists, world region, and sample demographics (see the Appendix for a complete list). These covariates helped to eliminate alternate explanations of relationships between women's representation in science and gender-science stereotypes. For instance, women's representation in science might reflect broader gender equity across multiple societal domains such as employment opportunities and political agency. However, recent research also has demonstrated the multidimensional nature of gender equity (Else-Quest & Grabe, 2012). For instance, gender differences in STEM achievement and attitudes related more strongly to women's representation in the researcher workforce than in the overall workforce (Else-Quest et al., 2010; for a review, see Miller & Halpern, 2014). We similarly predicted that gender-science stereotypes should relate more strongly to domain-specific measures of sex segregation than to composite indices of national gender equity.

Procedure

Participants found the Project Implicit website mainly through links from other websites, media coverage, search engines, and word of mouth (Nosek et al., 2002). The website was available in 17 different languages and hosted on various web servers across the world. Participants choose the gender-science task from a list of five to 12 topics (e.g., implicit age attitudes, implicit racial attitudes). Participants therefore self-selected into the sample by having Internet access, learning about the Project Implicit website, visiting the website, and choosing the gender-science task. The Results and Limitations sections consider the influence of possible self-selection biases. The explicit stereotype measure, implicit stereotype measure, and a brief demographics questionnaire (e.g., about participants' gender, nationality) were completed in counterbalanced order.[2] The gender-science task required approximately 10 min to fully complete. We analyzed data from participants who had indicated their nationality and had usable data for at least one of the two gender-science stereotype measures (see Nosek et al., 2009, for description of the data cleaning procedures for the implicit measure).

Data Analysis

Our analysis addressed three questions: (a) Does women's participation in science predict national explicit and implicit gender-science stereotypes? If so, how robust are these relationships across criteria for including nations? (b) Can other variables alternatively explain these relationships? (c) Are gender-science stereotypes better predicted by women's representation in science or gender differences in science achievement? Unless otherwise noted, all analyses used mixed-effects meta-regression models, which assumed that national averages were combinations of fixed effects of predictor variables (e.g., women's representation in science), between-nation heterogeneity, and within-nation sampling variance (Borenstein, Hedges, Higgins, & Rothstein, 2009). The *metafor* package in the statistical software R (Viechtbauer & Cheung, 2010) identified potential outliers using a diagnostic (DFFITS) of a nation's influence on the overall regression model. Nations were considered outliers if their $|\text{DFFITS}| > 1$, a rule of thumb useful to previous researchers (e.g., Cohen, Cohen, West, & Aiken, 2003; Nosek et al., 2009). Our raw data and analysis scripts are available from the first author.

Results

Averaged across the nations, explicit and implicit measures indicated strong associations of science with men (Ms = 0.99 and 0.98, respectively, based on random-effects weighting). The magnitude of these stereotypes was large in all nations. For instance, 90% of national averages for explicit and implicit measures fell within the ranges 0.78–1.20 and 0.76–1.20, respectively, which were estimated using the between-nation heterogeneity (both τs = 0.13) that adjusts for within-nation sampling variance. As shown in Figure 2, stereotypes were large even in nations such as Argentina and Bulgaria where women were approximately half of the nation's science majors and employed researchers. However, the between-nation heterogeneity was significant (both ps $< .0001$) and substantial relative to sampling error (only 3%–4% of observed heterogeneity could be attributed to within-nation sampling variance). This heterogeneity suggests that national attributes (e.g., women's representation in science) may explain differences in observed national averages. In addition, explicit and implicit measures correlated weakly among individuals within nations ($r = .19$, $p < .0001$, based on random-effects weighting) and across nations (based on national averages, $r = .35$, $p = .004$, $N = 66$ nations), suggesting that some national attributes may differently predict explicit versus implicit stereotypes.

Does Women's Representation in Science Predict National Gender-Science Stereotypes?

As shown in Figure 2, higher female enrollment in tertiary science education predicted weaker national averages of explicit (Panel a, $p = .0006$) and implicit (Panel c, $p = .0002$) gender-science stereotypes. Higher female employment in the researcher workforce predicted weaker explicit (Panel b, $p = .0004$) but not implicit (Panel d, $p = .88$) stereotypes.[3] Additionally, the difference between women's representation in science education versus researcher workforce predicted implicit stereotypes ($p = .006$), but not explicit stereotypes ($p = .55$). This last result established that Panel c's regression coefficient significantly differed from Panel d's and that Panel a's and Panel b's were both significant but did not differ from each other.

What might explain the exception in which women's employment in the researcher workforce did not predict implicit stereotypes (Panel d)? As suggested earlier, repeated counterstereotypic exposure is critical to changing implicit associations between

[2] Prior research has generally revealed that the order of administration (i.e., explicit or implicit measure first) does not substantially affect measurement of stereotypes at least for Project Implicit samples (Nosek et al., 2005, Study 3). Moreover, we found similar results when separating analyses by order of administration. For instance, the relationships reported in Figure 2 never differed by order of administration (all ps $> .46$).

[3] We also reanalyzed explicit stereotypes using difference scores that resembled those for the implicit measure: individuals' male–female associations for science minus for liberal arts. National averages of these difference scores marginally related to female science enrollment ($p = .06$) and significantly related to female researcher employment ($p = .01$). These p values, which were higher compared with Panels a's and b's values, suggested that including the contrast category of liberal arts introduced some construct-irrelevant variance. However, because these relationships were still significant or marginally so, these results cannot explain why Panel d's relationship with the implicit measure (which included a contrast category by design of the implicit measure) was not significant.

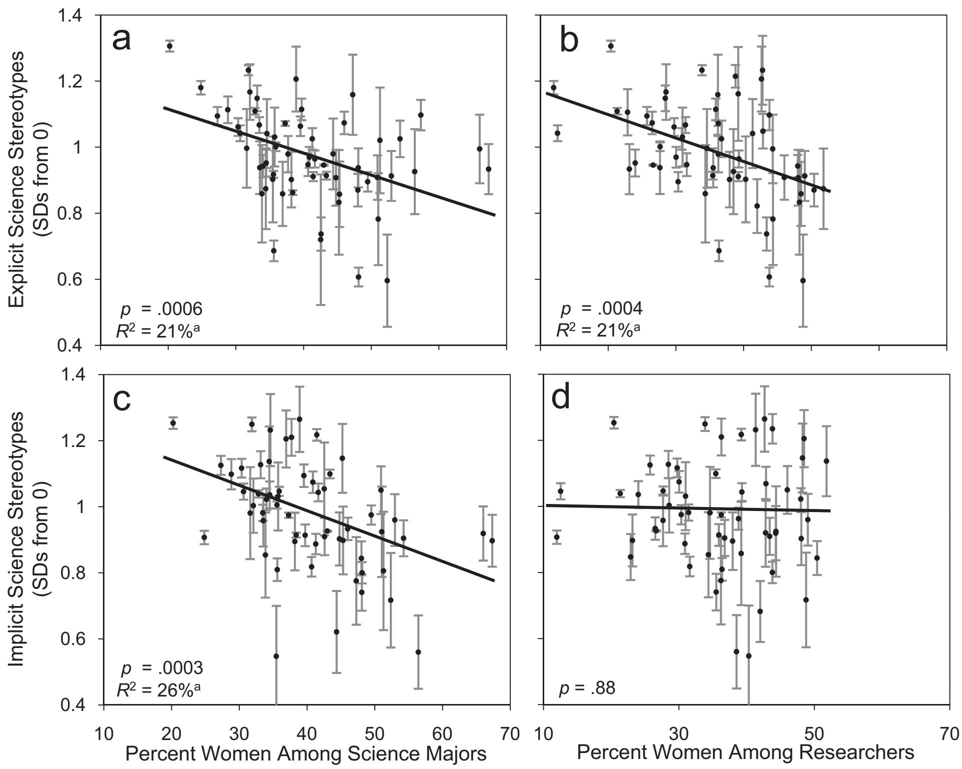

Figure 2. Cross-national relationships between women's participation in science and explicit (Panels a–b) and implicit (Panels c–d) gender-science stereotypes. Each data point reflects a nation's mean stereotypes after raw stereotype scores were standardized (see the Measures section); error bars represent standard errors. One influential outlier (Romania) was excluded from Panel c (see the Results section). [a]R^2 based on the percent reduction in estimated between-nation heterogeneity when adding women's participation in science to a meta-regression model with no covariates.

science and men. Notably, these mostly college-educated participants likely had less exposure to people employed as researchers than to science majors in universities, perhaps explaining why Panel d's relationship was not significant. To test this explanation, we investigated a corollary hypothesis: Panel c's relationship between women's science enrollment and implicit stereotypes should also be weaker among individuals less exposed to science majors than among those with more exposure. Additional analyses supported this hypothesis. As shown in Figure 3, Panel c's relationship between implicit stereotypes and women's enrollment was about half as strong for participants who had never attended college than for college-educated participants ($p = .001$), based on two-level hierarchical linear models (Raudenbush & Bryk, 2002). Presumably, participants without college education had less repeated exposure to female and male science majors. In contrast, relationships with explicit stereotypes (Panels a and b) did not differ by participants' level of education (all *p*s $> .10$). Finally, all significant relationships (Panels a–c) were approximately twice as strong for female than male participants (see Figure S1), consistent with other evidence that women are more sensitive to changes in gender diversity in STEM fields (Inzlicht & Ben-Zeev, 2000; Young et al., 2013). These differences by participant gender, however, were not as robust as differences by college education or the central findings in Figure 2 (see next section, Footnote 2).

How Robust Are Results Across Criteria for Selecting Nations?

Self-selected Internet samples such as ours have limited representativeness of national populations (Yeager et al., 2011). Consistent with other research (e.g., Lippa, Collaer, & Peters, 2010), we therefore selected nations on the basis of two variables (sample size and the population's percentage of Internet users) to maximize the likelihood of producing reasonably precise and representative national-level estimates. Rather than using a single criterion, we report results across many choices of selection criteria, as advocated by Simmons, Nelson, and Simonsohn (2011). Results in Figure 2 were robust across 36 choices in selection criteria based on minimum sample size ($n > 1$, $n > 10$, $n > 25$, $n > 50$, $n > 100$, $n > 200$) and percentage of Internet users (>0%, >1%, >5%, >10%, >25%, >50%). Across criteria, results were consistently replicated for the significant relationships in Panel a (all *p*s $< .005$), Panel b ($p < .05$ in 86% of cases), and Panel c ($p < .05$ in 86% of cases), as well as for the nonsignificant relationship in Panel d (all *p*s $> .28$). For Panels a–c, all relationships were in the predicted direction. Furthermore, consistent with results presented in the last section, Panel c's estimated relationship was always more than 50% stronger for individuals with a bachelor's degree

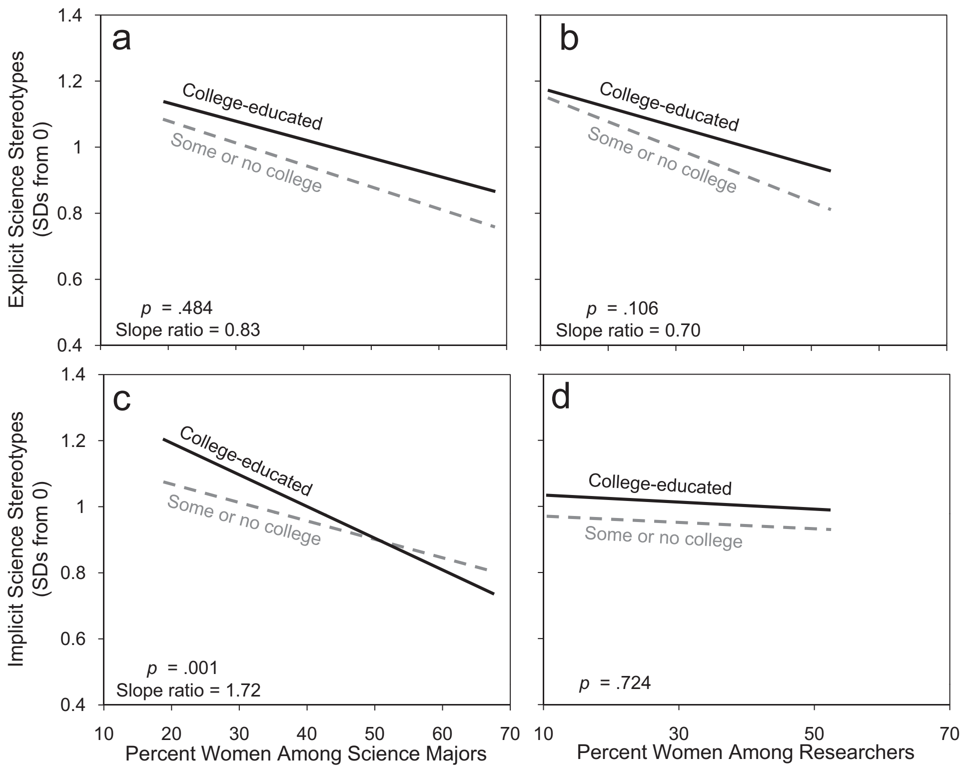

Figure 3. Moderation of cross-national relationships by participant's level of college education. The p values concern differences in the regression slopes, and "Slope ratio" is the slope for college-educated participants divided by the slope for participants with some or no college.

compared with those who never attended college ($p < .05$ in 72% of cases).[4] Also consistent with results presented earlier, Panel a's and b's estimated relationships never differed by college education (all ps $> .098$). Finally, Figure 2's relationships were also robust to exclusion of outliers. For instance, across selection criteria, Panel c's relationship was significant in 86% versus 78% of cases when including versus excluding outliers, respectively. Romania was an outlier in Figure 2's Panel c and therefore was excluded from that panel and subsequent analyses of that relationship; results were similar with and without the outlier. This robustness across selection criteria strengthens our central findings.

Can Covariates Explain Relationships Between Gender Diversity and Stereotypes?

Multiple regression models tested whether other national attributes could have accounted for Figure 2's relationships between women's representation in science and gender-science stereotypes. Closely following Bryk and Thum's (1989) analytic approach, we first developed separate regression models that each contained only one group of covariates (e.g., composite indices of gender equity). These initial models helped identify specific covariates that were most related to stereotypes. Consistent with Bryk and Thum, a composite model then included those covariates that significantly predicted stereotypes in the initial models. This approach maximized statistical power while investigating a wide range of covariates.

Multiple regression analyses generally indicated that (a) covariates such as national gender equity did not independently predict implicit or explicit gender-science stereotypes and (b) inclusion of covariates did not nullify relationships between women's representation in science and these stereotypes (see Table S1 for detailed results). For example, two widely used composite indices of national gender equity—the Gender Empowerment Measure and Gender Gap Index—did not independently predict explicit or implicit gender-science stereotypes (all ps $> .38$). When controlled for these measures, all relationships between women's science participation and gender-science stereotypes that were previously significant (see Figure 2, Panels a–c) remained significant (all ps $< .002$). The Netherlands was a particularly dramatic example of composite equity indices not predicting gender-science stereotypes. Despite scoring high on composite indices of gender equity, this nation (sample size $n \sim 3{,}000$) had the strongest explicit and second strongest implicit gender-science stereotypes among the nations in Figure 1. This seemingly paradoxical result, however, makes sense because of high domain-specific sex segregation in the Netherlands, whereby male scientists outnumbered

[4] The moderating effect of gender was less robust. Across selection criteria, our focal relationships (Panels a–c) were stronger for women than men in 98% of cases and twice as strong in 32% of cases. These trends were consistent but significant ($p < .05$) in only 17% of cases and marginal ($.05 < p < .10$) in 21% of cases.

female scientists nearly four to one in both employment and educational enrollment.

Furthermore, indicating discriminant validity, the percent of women among science majors or researchers did not predict explicit stereotypes about liberal arts (all *p*s $> .06$). Women's representation in science therefore did not predict gender stereotypes that are not related to science. Additionally, average explicit stereotypes for liberal arts and science were generally not related across nations (e.g., $r = .09$ among the 66 nations in Figure 1). In summary, covariate and discriminant validity analyses together support the domain specificity of relationships between women's representation in science and national gender-science stereotypes.

How Do Achievement Differences, Compared With Gender Diversity, Relate to Stereotypes?

Nosek et al. (2009) presented evidence that gender differences in science achievement related to national implicit gender-science stereotypes (see also Hamamura, 2012; Pope & Sydnor, 2010). Our covariate analyses, however, revealed that these achievement differences did not independently relate to stereotypes after controlling for women's enrollment in science education. Hence, although both gender differences in achievement and in enrollment sometimes related to cross-national differences in gender-science stereotypes, gender differences in enrollment may be more relevant to explaining differences in stereotypes. To investigate further, we compared the strength of stereotype–achievement relationships across time, selection criteria, participant gender, inclusion of covariates, and international data sources (for further detail, see the supplemental materials).

Consistent with Nosek et al. (2009), stereotype–achievement relationships were found in data from the Trends in International Mathematics and Science Study (TIMSS), which focuses on assessing what students learn in science classrooms. However, these results for TIMSS were somewhat inconsistent over time (e.g., not replicated in the year 2007), as shown in the top-left corner of Table 1. Averaging across four testing administrations helped to identify overall trends. For instance, indicating some robustness, time-averaged gender differences in TIMSS science achievement significantly related to implicit gender-science stereotypes in 39% of cases of selection criteria after excluding one influential outlier. These cross-national relationships were somewhat more robust for the stereotypes of female than male participants (see bottom-left corner of Table 1). For instance, time-averaged TIMSS gender differences related to women's implicit stereotypes in 58% of cases of selection criteria after excluding one influential outlier. When controlled for women's enrollment in science education, however, this relationship remained significant in only 8% of cases (and in the predicted direction in 89% of cases), whereas women's enrollment continued to significantly predict stereotypes in 67% of cases (see Tables S2–S6 for more detailed results). Finally, our analysis identified another novel finding that relationships between achievement gender differences and stereotypes were generally *not* found in data from the Programme for International Student Assessment (PISA), which focuses more on assessing how well students apply science to everyday contexts than does TIMSS (Else-Quest et al., 2010; but see Fensham, 2008). See right half of Table 1 for results for PISA. Hence, achievement differences independently predicted stereotypes in some cases when specifically analyzing women's implicit stereotypes and TIMSS (not PISA) data. However, evidence for this relationship was considerably less robust than for relationships between gender-science stereotypes and women's representation in science.

Discussion

Results indicated robust relationships between women's representation in science and *national gender-science stereotypes*, defined as associations connecting science with men more than women. These relationships tended to be stronger for female participants and remained after controlling for many covariates such as national gender equity. Even nations with high overall gender equity had strong gender-science stereotypes if men dominated science fields specifically (see also Charles & Bradley,

Table 1
Robustness of Stereotype–Achievement Relationships

	TIMSS							PISA						
Variable	1999	2003	2007	2011	Ave	Ave[a]	Ave[b]	2000	2003	2006	2009	Ave	Ave[c]	Ave[d]
							Predicting mean implicit stereotypes							
$p < .05$	25%	44%	0%	17%	8%	39%	0%	3%	0%	0%	0%	0%	0%	0%
$.05 < p < .10$	11%	6%	0%	17%	19%	11%	8%	25%	0%	0%	0%	3%	3%	0%
$p > .10$	64%	50%	100%	67%	72%	50%	92%	72%	100%	100%	100%	97%	97%	100%
							Predicting women's implicit stereotypes							
$p < .05$	17%	50%	0%	53%	31%	58%	8%	0%	0%	0%	0%	28%	0%	0%
$.05 < p < .10$	14%	17%	0%	28%	17%	17%	25%	0%	0%	0%	6%	6%	17%	0%
$p > .10$	69%	33%	100%	19%	53%	25%	67%	100%	100%	100%	94%	67%	83%	100%
Max N	38	43	46	44	62	61	51	42	40	55	68	69	68	61

Note. Each column displays results across selection criteria (e.g., with 1999 TIMSS data, stereotype–achievement relationships were significant across 25% of choices in selection criteria). TIMSS = Trends in Mathematics and Science Study; PISA = Programme for International Student Assessment; Ave = time-averaged gender differences in science achievement; Max N = number of nations analyzed with the most liberal selection criteria (sample size $n > 1$).
[a] Outlier Colombia excluded. [b] Controlling for percent women among science majors. Outliers Colombia and Romania excluded. [c] Outlier Malta excluded. [d] Controlling for percent women among science majors. Outliers Malta and Romania excluded.

2009). In support of the specificity to science fields, women's representation in science did not predict explicit gender stereotypes about liberal arts. Furthermore, compared with gender differences in science achievement (Nosek et al., 2009), women's representation in science more robustly predicted explicit and implicit stereotypes.

Women's representation in science predicted national gender-science stereotypes in three of four cases. As an informative boundary condition, women's employment in the researcher workforce predicted only explicit, but not implicit, gender-science stereotypes. This result suggests that repeated and varied exposures to counterstereotypic women may be necessary to stably change implicit gender-science stereotypes, consistent with broader literature on implicit social cognition (Gawronski & Bodenhausen, 2006, 2011). The implicit stereotypes of these mostly college-educated participants thus likely related more to their frequent exposure to female and male science majors and less to their rarer exposure to female and male employed researchers. If this reasoning is valid, then cross-national relationships between implicit stereotypes and gender diversity among science majors should also be weaker among participants who never attended college. Our analyses supported these predictions.

Repeated counterstereotypic exposure may be less critical to changing explicit stereotypes because they also respond to more abstract propositional information such as statistics about women's representation in science (Gawronski & Bodenhausen, 2006, 2011). Consistent with these hypotheses, relationships between gender diversity and explicit stereotypes were similar for participants with and without college education, even though participants without college education likely had less repeated exposure to female and male science majors. These results align with other findings that people often are highly accurate in explicitly estimating gender compositions of occupations. For instance, in one study, undergraduates had high accuracy across 80 occupations, despite little direct exposure to women and men in those occupations (Cejka & Eagly, 1999).

These cross-national findings extend previous research investigating the psychological effects of encountering female role models in STEM fields (Dasgupta, 2011; González de San Román & de la Rica Goiricelaya, 2012; Riegle-Crumb & Moore, 2014). Such role models show promise for weakening stereotypes, especially among female students (Beilock et al., 2010; Galdi et al., 2014; although see Lenton, Bruder, & Sedikides, 2009) and students who strongly identify with the role models (Young et al., 2013). However, these counterstereotypic examples could be subtyped because they occur along with pervasive stereotypic evidence from the broad cultural environment (Richards & Hewstone, 2001; Stout et al., 2011). Changes in broader cultural environments such as women's increasing representation in science fields in the United States. (Hill et al., 2010) might have stronger, more robust effects on gender-science stereotypes. Role models may be one of the first steps in changing stereotypes over time, especially because female STEM peers and mentors can help protect girls and women against the negative effects of current stereotypes (Dasgupta, 2011; Stout et al., 2011). Cultural stereotypes could then change as more women enter STEM fields and gender compositions change at the national level (Beaman, Chattopadhyay, Duflo, Pande, & Topalova, 2009). Future research can help understand how individual differences in counterstereotypic exposure contribute to these cultural trends. For instance, our analysis of the relationship between individuals' educational attainment and gender-science stereotypes could be extended by measuring how closely college-educated individuals identified with female science peers and professors (Young et al., 2013).

Throughout this article, we have primarily considered how gender diversity might influence stereotypes. However, as noted earlier, because the data are correlational, a bidirectional relationship is plausible. For instance, as Nosek et al. (2009) suggested, implicit stereotypes could cause women to underperform on science achievement tests because of a phenomenon known as *stereotype threat* (Schmader, Johns, & Forbes, 2008; Walton & Spencer, 2009). This lower achievement could, in turn, limit women's access to science fields. Consistent with this reasoning, women's implicit gender-science stereotypes related to male advantages in the TIMSS test, which was designed to assess students' learning of science curriculum. However, this evidence was less robust than our central findings relating stereotypes and women's representation in science. Stereotypes could also influence women's representation in science through other factors such as women's identification with STEM fields (Dasgupta, 2011; Nosek & Smyth, 2011). Both causal directions between gender composition in science and gender-science stereotypes are thus plausible, although gender composition likely influences stereotypes more directly than stereotypes influence gender composition. The impact of stereotypes on gender compositions would be mediated over many years as women enroll in STEM courses and seek employment in STEM fields, whereas the impact of gender compositions on stereotypes can be more immediate (Lenton et al., 2009).

Furthermore, some of our study's results would be difficult to explain if gender composition did not influence stereotypes in some way. For instance, if the gender composition of science majors in college did not affect stereotypes, then stereotypes of individuals with and without college education should not differ. Another alternative hypothesis is that individuals with and without college education might differ on average if other correlated individual-level variables (e.g., age or socioeconomic status) influence stereotypes. However, our data supported neither hypothesis because college education predicted stronger implicit stereotypes, but only in nations where men dominated science majors (see Figure 3). In contrast, college education predicted weaker implicit stereotypes in nations where women dominated science majors. Compared with those alternative hypotheses, the associative-propositional model (Gawronski & Bodenhausen, 2006, 2011) can more parsimoniously account for the cross-level interactions with women's representation in science, as discussed earlier.

Limitations

Our correlational design revealed the possible impact of cultural environments, but experimental manipulations offer greater potential for causal inference. However, the effects of experimental manipulations may be weakened by broader sociocultural messages (e.g., Stout et al., 2011). Hence, the effects of cultural environments are inherently challenging to study because they generally cannot be experimentally manipulated. Although investigating changes over time could strengthen cross-cultural analyses such as ours (Brandt, 2011), this study's time period of data collection was too small (2000–2008) to meaningfully test for such longitudinal changes.

This study used self-selected Internet samples, which have limited representativeness of national populations (Yeager et al., 2011), especially if the percentage of Internet users is low. These concerns were somewhat lessened because our central results were robust across a wide range of minimum percentages of Internet users. Nevertheless, participants also self-selected into our study by finding the Project Implicit website and choosing the gender-science task. For instance, participants especially interested in gender issues might have been more likely to choose the gender-science task. Self-selection is therefore a general methodological concern. However, it is unclear how much self-selection affected our specific empirical findings (e.g., the regression slopes in Figure 2).

Another limitation was that available statistics did not permit analysis of how results might have differed by field of science (e.g., physics vs. biology). Future research should address this issue because both stereotypes and gender diversity vary substantially by field (Nosek & Smyth, 2011; Smyth & Nosek, 2013). However, as noted in the introduction, our analyses that averaged across science fields may be justified because female dominance in only one field (e.g., biology) could be regarded as an exception to the usual pattern of male dominance in science (Richards & Hewstone, 2001). Other limitations were that explicit stereotypes were measured by a single survey item, participants without college education were underrepresented in our sample, nations characterized as low in human development and/or had low Internet usage rates were underrepresented in our sample, and the number of nations was small ($N = 66$) even if the number of participants was large. However, in defense of our findings, they proved to be robust despite these limitations.

Educational Implications and Future Research

Our results indicated that participants across 66 nations strongly associated science with men more than women, including in nations where women were approximately half of the nation's science majors and employed researchers (see also Nosek et al., 2009). Hence, across the world, gender-science stereotypes present concerns for science educators and students to the extent that these associations affect the experiences of women and men pursuing science degrees and occupations. For instance, such stereotypes negatively impact women by causing underachievement in introductory undergraduate STEM courses (Miyake et al., 2010), disidentification with and negative attitudes toward science (Good, Rattan, & Dweck, 2012; Nosek & Smyth, 2011; Steffens et al., 2010), and gender discrimination (Rueben, Sapienza, & Zingales, in press). Despite their ubiquity, gender-science stereotypes also demonstrated cultural variability and therefore potential for change. Yet, gender-science stereotypes were still strong even in nations with small gender differences in overall science participation.

Although it is not yet clear how best to weaken these stereotypes, a number of promising strategies can be explored. Science educators could help weaken stereotypes by highlighting diverse examples of female scientists. Presenting single or infrequent examples of female scientists will likely not substantially change gender-STEM stereotypes (Stout et al., 2011), especially if such women are presented as token examples (Shachar, 2000). A more effective strategy to weaken stereotypes could be to integrate many examples of female scientists as part of teachers' normal classroom instruction. For instance, teachers could motivate the learning of specific scientific concepts by discussing how they relate to the research of currently practicing female and male scientists (Linn & Eylon, 2011). Related prior research has indicated the benefits of integrating narrative information about scientists into instruction (Arya & Maul, 2012). For instance, in one experimental study (Hong & Lin-Siegler, 2012), learning how scientists struggled in their research increased students' interest in the science lesson and students' content understanding (e.g., about Newtonian mechanics). Learning how both female and male scientists struggle could also help protect female students against the negative effects of gender-STEM stereotypes (Asgari, Dasgupta, & Stout, 2012; Good et al., 2012). Future research should extend these approaches to understand how repeated examples of female scientists might weaken gender-STEM stereotypes over time.

Our study might also have implications for social policies such as affirmative action. For the recent U.S. Supreme Court case *Fisher v. Texas*, social psychologists prepared an amicus brief outlining the implications of stereotype threat for affirmative action (Brief of Experimental Psychologists, 2012). The brief argued that diversity in college populations helps minorities reach their maximum potential because ingroup peers can inoculate minorities against the negative effects of cultural stereotypes (Dasgupta, 2011; Murphy, Steele, & Gross, 2007; Richman, vanDellen, & Wood, 2011). Increasing the diversity of college populations might also change underlying stereotypes about science fields. Future research should investigate this possibility. These efforts to weaken stereotypes could then have cascading influence by encouraging more women to pursue and excel in fields in which they have been historically underrepresented.

References

Arya, D. J., & Maul, A. (2012). The role of scientific discovery narrative in middle school science education: An experimental study. *Journal of Educational Psychology, 104,* 1022–1032. doi:10.1037/a0028108

Asgari, S., Dasgupta, N., & Stout, J. G. (2012). When do counterstereotypic ingroup members inspire versus deflate? The effect of successful professional women on young women's leadership self-concept. *Personality and Social Psychology Bulletin, 38,* 370–383. doi:10.1177/0146167211431968

Barbercheck, M. (2001). Mixed messages: Men and women in advertisements inScience. In M. Wyer, M. Barbercheck, D. Geisman, H. O. Ozturk, & M. Wayne (Eds.), *Women, science, and technology: A reader in feminist science studies* (pp. 117–131). New York, NY: Routledge.

Barone, C. (2011). Some things never change: Gender segregation in higher education across eight nations and three decades. *Sociology of Education, 84,* 157–176. doi:10.1177/0038040711402099

Bazler, J. A., & Simonis, D. A. (1991). Are high school chemistry textbooks gender fair? *Journal of Research in Science Teaching, 28,* 353–362. doi:10.1002/tea.3660280408

Beaman, L., Chattopadhyay, R., Duflo, E., Pande, R., & Topalova, P. (2009). Powerful women: Does exposure reduce bias? *The Quarterly Journal of Economics, 124,* 1497–1540. doi:10.1162/qjec.2009.124.4.1497

Beilock, S. L., Gunderson, E. A., Ramirez, G., & Levine, S. C. (2010). Female teachers' math anxiety affects girls' math achievement. *Proceedings of the*

National Academies of Sciences, 107, 1060–1063. doi:10.1073/pnas.0910967107

Bergkvist, L., & Rossiter, J. R. (2007). The predictive validity of multiple-item versus single-item measures of the same constructs. *Journal of Marketing Research, 44,* 175–184. doi:10.1509/jmkr.44.2.175

Bigler, R. S., & Liben, L. S. (2006). A developmental intergroup theory of social stereotypes and prejudice. In R. V. Kail (Ed.), *Advances in child development and behavior* (Vol. 34, pp. 39–89). San Diego, CA: Elsevier. doi:10.1016/S0065-2407(06)80004-2

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Chichester, England: Wiley. doi:10.1002/9780470743386

Bowen, C. E., & Skirbekk, V. (2013). National stereotypes of older people's competence are related to older adults' participation in paid and volunteer work. *Journals of Gerontology, Series B: Psychological Sciences and Social Sciences, 68,* 974–983. doi:10.1093/geronb/gbt101

Brandt, M. J. (2011). Sexism and gender inequality across 57 societies. *Psychological Science, 22,* 1413–1418. doi:10.1177/0956797611420445

Brief of Experimental Psychologists et al. as Amici Curiae Supporting Respondents, Fisher v. University of Texas, August 13, 2012 (No. 01–1015).

Brotman, J. S., & Moore, F. M. (2008). Girls and science: A review of four themes in the science education literature. *Journal of Research in Science Teaching, 45,* 971–1002. doi:10.1002/tea.20241

Bryk, A. S., & Thum, Y. M. (1989). The effects of high school organization on dropping out: An exploratory investigation. *American Educational Research Journal, 26,* 353–383. doi:10.3102/00028312026003353

Ceci, S. J., Williams, W. M., & Barnett, S. M. (2009). Women's underrepresentation in science: Sociocultural and biological considerations. *Psychological Bulletin, 135,* 218–261. doi:10.1037/a0014412

Cejka, M. A., & Eagly, A. H. (1999). Gender-stereotypic images of occupations correspond to the sex segregation of employment. *Personality and Social Psychological Bulletin, 25,* 413–423. doi:10.1177/0146167299025004002

Chambers, D. W. (1983). Stereotypic images of the scientist: The draw-a-scientist test. *Science Education, 67,* 255–265. doi:10.1002/sce.3730670213

Charles, M., & Bradley, K. (2009). Indulging our gendered selves? Sex segregation by field of study in 44 countries. *American Journal of Sociology, 114,* 924–976. doi:10.1086/595942

Chimba, M., & Kitzinger, J. (2010). Bimbo or boffin? Women in science: An analysis of media representations and how female scientists negotiate cultural contradictions. *Public Understanding of Science, 19,* 609–624. doi:10.1177/0963662508098580

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Dasgupta, N. (2011). Ingroup experts and peers as social vaccines who inoculate the self-concept: The stereotype inoculation model. *Psychological Inquiry, 22,* 231–246. doi:10.1080/1047840X.2011.607313

Eagly, A. H., & Wood, W. (2012). Social role theory. In P. van Lange, A. Kruglanski, & E. T. Higgins (Eds.), *Handbook of theories of social psychology* (pp. 458–476). Thousand Oaks, CA: Sage. doi:10.4135/9781446249222.n49

Else-Quest, N. M., & Grabe, S. (2012). The political is personal: Measurement and application of nation-level indicators of gender equity in psychological research. *Psychology of Women Quarterly, 36,* 131–144. doi:10.1177/0361684312441592

Else-Quest, N. M., Hyde, J. S., & Linn, M. C. (2010). Cross-national patterns of gender differences in mathematics: A meta-analysis. *Psychological Bulletin, 136,* 103–127. doi:10.1037/a0018053

Farland-Smith, D. (2009). How does culture shape students' perceptions of scientists? Cross-national comparative study of American and Chinese elementary students. *Journal of Elementary Science Education, 21,* 23–42. doi:10.1007/BF03182355

Fensham, P. J. (2008). Context or culture: Can TIMSS and PISA teach us about what determines educational achievement in science? In B. Atweh, A. C. Barton, M. Borba, N. Gough, C. Keitel, C. Vistro-Yu, & R. Vithal (Eds.), *Internationalisation and globalization in mathematics and science education* (pp. 151–172). Dordrecht, the Netherlands: Springer. doi:10.1007/978-1-4020-5908-7_9

Fishbein, M., & Ajzen, I. (1974). Attitudes towards objects as predictors of single and multiple behavioral criteria. *Psychological Review, 81,* 59–74. doi:10.1037/h0035872

Flicker, E. (2003). Between brains and breasts—Women scientists in fiction film: On the marginalization and sexualization of scientific competence. *Public Understanding of Science, 12,* 307–318. doi:10.1177/0963662503123009

Fralick, B., Kearn, J., Thompson, S., & Lyons, J. (2009). How middle schoolers draw engineers and scientists. *Journal of Science Education and Technology, 18,* 60–73. doi:10.1007/s10956-008-9133-3

Furnham, A., & Paltzer, S. (2010). The portrayal of men and women in television advertisements: An updated review of 30 studies published since 2000. *Scandinavian Journal of Psychology, 51,* 216–236. doi:10.1111/j.1467-9450.2009.00772.x

Galdi, S., Cadinu, M., & Tomasetto, C. (2014). The roots of stereotype threat: When automatic associations disrupt girls' math performance. *Child Development, 85,* 250–263. doi:10.1111/cdev.12128

Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin, 132,* 692–731. doi:10.1037/0033-2909.132.5.692

Gawronski, B., & Bodenhausen, G. V. (2011). The associative-propositional evaluation model: Theory, evidence, and open questions. In M. Zanna & J. M. Olson (Eds.), *Advances in experimental social psychology* (Vol. 44, pp. 59–127). Amsterdam, the Netherlands: Elsevier. doi:10.1016/B978-0-12-385522-0.00002-0

González de San Román, A., & de la Rica Goiricelaya, S. (2012). *Gender gaps in PISA test scores: The impact of social norms and the mother's transmission of role attitudes* (IZA Discussion Paper). Bonn, Germany: Institute for the Study of Labor (IZA). Retrieved from http://ftp.iza.org/dp6338.pdf

Gonzales, P., Guzman, J. C., Partelow, L., Pahlke, E., Jocelyn, L., Kastberg, D., & Williams, T. (2004). *Highlights from the Trends in International Mathematics and Science Study (TIMSS) 2003* (NCES Report No. 2005-005). Washington, DC: U. S. Department of Education, National Center for Education Statistics.

Good, C., Rattan, A., & Dweck, C. S. (2012). Why do women opt out? Sense of belonging and women's representation in mathematics. *Journal of Personality and Social Psychology, 102,* 700–717. doi:10.1037/a0026659

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology, 85,* 197–216. doi:10.1037/0022-3514.85.2.197

Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L., & Banaji, M. R. (2009). Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology, 97,* 17–41. doi:10.1037/a0015575

Gunderson, E. A., Ramirez, G., Levine, S. C., & Beilock, S. L. (2012). New directions for research on the role of parents and teachers in the development of gender-related math attitudes. *Sex Roles, 66,* 153–166. doi:10.1007/s11199-011-9996-2

Hamamura, T. (2012). Power distance predicts gender differences in math performance across societies. *Social Psychological and Personality Science, 3,* 545–548. doi:10.1177/1948550611429191

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences, 33,* 61–83.

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

Hofstede, G., Hofstede, G. J., & Minkov, M. (2010). *Cultures and organizations: Software of the mind*. New York, NY: McGraw-Hill.

Hong, H.-Y., & Lin-Siegler, X. (2012). How learning about scientists' struggles influences students' interest and learning in physics. *Journal of Educational Psychology, 104,* 469–484. doi:10.1037/a0026224

Inzlicht, M., & Ben-Zeev, T. (2000). A threatening intellectual environment: Why females are susceptible to experiencing problem-solving deficits in the presence of males. *Psychological Science, 11,* 365–371. doi:10.1111/1467-9280.00272

Kimball, M. M. (1986). Television and sex-role attitudes. In T. M. Williams (Ed.), *The impact of television: A natural experiment in three communities* (pp. 265–301). Orlando, FL: Academic Press.

Knapp, G., & Hartung, J. (2003). Improved tests for a random effects meta-regression with a single covariate. *Statistics in Medicine, 22,* 2693–2710.

Koenig, A. M., & Eagly, A. H. (in press). Evidence for the social role theory of stereotype content: Observations of groups' roles shape stereotypes. *Journal of Personality and Social Psychology*.

Koenig, A. M., Eagly, A. H., Mitchell, A. A., & Ristikari, T. (2011). Are leader stereotypes masculine? A meta-analysis of three research paradigms. *Psychological Bulletin, 137,* 616–642. doi:10.1037/a0023557

Lenton, A. P., Bruder, M., & Sedikides, C. (2009). A meta-analysis on the malleability of automatic gender stereotypes. *Psychology of Women Quarterly, 33,* 183–196. doi:10.1111/j.1471-6402.2009.01488.x

Linn, M. C., & Eylon, B-S. (2011). *Science learning and instruction: Taking advantage of technology to promote knowledge integration*. New York, NY: Routledge.

Lippa, R. A., Collaer, M. L., & Peters, M. (2010). Sex differences in mental rotation and line angle judgments are positively associated with gender equality and economic development across 53 nations. *Archives of Sexual Behavior, 39,* 990–997. doi:10.1007/s10508-008-9460-8

Long, M., Steinke, J., Applegate, B., Lapinski, M. K., Johnson, M. J., & Ghosh, S. (2010). Portrayals of male and female scientists in television programs popular among middle school-age children. *Science Communication, 32,* 356–382. doi:10.1177/1075547009357779

Martin, M. O., Mullis, I. V. S., Gonzalez, E. J., & Chrostowski, S. J. (2004). *TIMSS 2003 international mathematics report: Findings from IEA's trends in international mathematics and sciene study at the fourth and eighth grades*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Mason, C. L., Kahle, J. B., & Gardner, A. L. (1991). Draw-a-scientist test: Future implications. *School Science and Mathematics, 91,* 193–198. doi:10.1111/j.1949-8594.1991.tb12078.x

Milford, T. M., & Tippett, C. D. (2013). Preservice teachers' images of scientists: Do prior science experiences make a difference? *Journal of Science Teacher Education, 24,* 745–762. doi:10.1007/s10972-012-9304-1

Miller, D. I., & Halpern, D. F. (2014). The new science of cognitive sex differences. *Trends in Cognitive Sciences, 18,* 37–45. doi:10.1016/j.tics.2013.10.011

Miyake, A., Kost-Smith, L. E., Finkelstein, N. D., Pollock, S. J., Cohen, G. L., & Ito, T. A. (2010). Reducing the gender achievement gap in college science: A classroom study of values affirmation. *Science, 330,* 1234–1237. doi:10.1126/science.1195996

Murphy, M. C., Steele, C. M., & Gross, J. J. (2007). Signaling threat how situational cues affect women in math, science, and engineering settings. *Psychological Science, 18,* 879–885. doi:10.1111/j.1467-9280.2007.01995.x

National Science Board. (2014). *Science and engineering indicators 2014* (Report No. NSB 14–01). Arlington, VA: National Science Foundation.

Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002). Harvesting implicit group attitudes and beliefs from a demonstration web site. *Group Dynamics: Theory, Research, and Practice, 6,* 101–115. doi:10.1037/1089-2699.6.1.101

Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2005). Understanding and using the Implicit Association Test: II. Method variables and construct validity. *Personality and Social Psychology Bulletin, 31,* 166–180. doi:10.1177/0146167204271418

Nosek, B. A., Hawkins, C. B., & Frazier, R. S. (2011). Implicit social cognition: From measures to mechanisms. *Trends in Cognitive Sciences, 15,* 152–159. doi:10.1016/j.tics.2011.01.005

Nosek, B. A., & Smyth, F. L. (2011). Implicit social cognitions predict sex differences in math engagement and achievement. *American Educational Research Journal, 48,* 1125–1156. doi:10.3102/0002831211410683

Nosek, B. A., Smyth, F. L., Hansen, J. J., Devos, T., Lindner, N. M., Ranganath, K. A., . . . Banaji, M. R. (2007). Pervasiveness and correlates of implicit and explicit attitudes and stereotypes. *European Review of Social Psychology, 18,* 36–88. doi:10.1080/10463280701489053

Nosek, B. A., Smyth, F. L., Sriram, N., Lindner, N. M., Devos, T., Ayala, A., . . . Greenwald, A. G. (2009). National differences in gender–science stereotypes predict national sex differences in science and math achievement. *Proceedings of the National Academies of Science, 106,* 10593–10597. doi:10.1073/pnas.0809921106

Núñez-Peña, M. I., Guilera, G., & Suárez-Pellicioni, M. (in press). The single-item math anxiety scale: An alternative way of measuring mathematics anxiety. *Journal of Psychoeducational Assessment*. doi:10.1177/0734282913508528

Oswald, F. L., Mitchell, G., Blanton, H., Jaccard, J., & Tetlock, P. E. (2013). Predicting ethnic and racial discrimination: A meta-analysis of IAT criterion studies. *Journal of Personality and Social Psychology, 105,* 171–192. doi:10.1037/a0032734

Pope, D. G., & Sydnor, J. R. (2010). Geographic variation in the gender differences in test scores. *Journal of Economic Perspectives, 24,* 95–108. doi:10.1257/jep.24.2.95

Ramsey, L. R., Betz, D. E., & Sekaquaptewa, D. (2013). The effects of an academic environment intervention on science identification among women in STEM. *Social Psychology of Education, 16,* 377–397. doi:10.1007/s11218-013-9218-6

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Reilly, D. (2012). Gender, culture, and sex-typed cognitive abilities. *PLoS ONE, 7,* e39904. doi:10.1371/journal.pone.0039904

Richards, Z., & Hewstone, M. (2001). Subtyping and subgrouping: Processes for the prevention and promotion of stereotype change. *Personality and Social Psychology Review, 5,* 52–73. doi:10.1207/S15327957PSPR0501_4

Richman, L. S., vanDellen, M., & Wood, W. (2011). How women cope: Being a numerical minority in a male-dominated profession. *Journal of Social Issues, 67,* 492–509. doi:10.1111/j.1540-4560.2011.01711.x

Riegle-Crumb, C., King, B., Grodsky, E., & Muller, C. (2012). The more things change, the more they stay the same? Prior achievement fails to explain gender inequality in entry into STEM college majors over time. *American Educational Research Journal, 49,* 1048–1073. doi:10.3102/0002831211435229

Riegle-Crumb, C., & Moore, C. (2014). The gender gap in high school physics: Considering the context of local communities. *Social Science Quarterly, 95,* 253–268. doi:10.1111/ssqu.12022

Rueben, E., Sapienza, P., & Zingales, L. (in press). How stereotypes impair women's careers in science. *Proceedings of the National Academies of Science*. doi:10.1073/pnas.1314788111

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

Shachar, O. (2000). Spotlighting women scientists in the press: Tokenism in science journalism. *Public Understanding of Science, 9,* 347–358. doi:10.1088/0963-6625/9/4/301

Sherman, J. W., Gawronski, B., & Trope, Y. (Eds.). (2014). *Dual-process theories of the social mind*. New York, NY: Guilford Press.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science, 22,* 1359–1366. doi:10.1177/0956797611417632

Smith, J. L., Lewis, K. L., Hawthorne, L., & Hodges, S. D. (2013). When trying hard isn't natural: Women's belonging with and motivation for male-dominated STEM fields as a function of effort expenditure concerns. *Personality and Social Psychology Bulletin, 39,* 131–143. doi:10.1177/0146167212468332

Smyth, F. L., & Nosek, B. A. (2013). *Male and female scientists' implicit gender-science stereotypes vary with scientific identity, not gender-ratios*. Unpublished manuscript, Department of Psychology, University of Virginia, Charlottesville, VA.

Steffens, M. C., Jelenec, P., & Noack, P. (2010). On the leaky math pipeline: Comparing implicit math-gender stereotypes and math withdrawal in female and male children and adolescents. *Journal of Educational Psychology, 102,* 947–963. doi:10.1037/a0019920

Steinke, J. (2013). Portrayals of female scientists in the mass media. In A. Valdivia & S. R. Mazzarella (Eds.), *The international encyclopedia of media studies* (pp. 1–18). Oxford, England: Blackwell.

Steinke, J., Lapinski, M. K., Crocker, N., Zietsman-Thomas, A., Williams, Y., Evergreen, S. H., & Kuchibhotla, S. (2007). Assessing media influences on middle school-aged children's perceptions of women in science using the Draw-A-Scientist Test (DAST). *Science Communication, 29,* 35–64. doi:10.1177/1075547007306508

Stout, J. G., Dasgupta, N., Hunsinger, M., & McManus, M. A. (2011). STEMing the tide: Using ingroup experts to inoculate women's self-concept in science, technology, engineering, and mathematics (STEM). *Journal of Personality and Social Psychology, 100,* 255–270. doi:10.1037/a0021385

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metaphor package. *Journal of Statistical Software, 36,* 1–48.

Viechtbauer, W., & Cheung, M. W. L. (2010). Outlier and influence diagnostics for meta-analysis. *Research Synthesis Methods, 1,* 112–125. doi:10.1002/jrsm.11

Walton, G. M., & Spencer, S. J. (2009). Latent ability: Grades and test scores systematically underestimate the intellectual ability of negatively stereotyped students. *Psychological Science, 20,* 1132–1139. doi:10.1111/j.1467-9280.2009.02417.x

Wanous, J. P., Reichers, A. E., & Hudy, M. J. (1997). Overall job satisfaction: How good are single-item measures? *Journal of Applied Psychology, 82,* 247–252. doi:10.1037/0021-9010.82.2.247

Wood, W., & Eagly, A. H. (2012). Biosocial construction of sex differences and similarities in behavior. In J. M. Olson & M. P. Zanna (Eds.), *Advances in experimental social psychology* (Vol. 46, pp. 55–123). London, England: Elsevier. doi:10.1016/B978-0-12-394281-4.00002-7

Yeager, D. S., Krosnick, J. A., Chang, L., Javitz, H. S., Levendusky, M. S., Simpser, A., & Wang, R. (2011). Comparing the accuracy of RDD telephone surveys and internet surveys conducted with probability and non-probability samples. *Public Opinion Quarterly, 75,* 709–747. doi:10.1093/poq/nfr020

Young, G. M., Rudman, L. A., Buettner, H. M., & McLean, M. C. (2013). The influence of female role models on women's implicit science cognitions. *Psychology of Women Quarterly, 37,* 283–292. doi:10.1177/0361684313482109

(*Appendix follows*)

Appendix

Covariates Included in Multiple Regression Analyses

Variable name	Description
Composite indices of gender equity	
GGI	Gender Gap Index. Based on four subindices for gender gaps in economic participation and opportunity, educational attainment, political empowerment, and health/survival.
GEM	Gender Empowerment Measure. Based on gender gaps in earned income; women's representation in parliament; and women's employment in managerial, professional, and technical occupations.
Domain-specific gender equity	
GGI_eco	Economic subindex of Gender Gap Index. Based on gender gaps in labor force participation rates, wage equality for similar work, earned income, and high labor employment.
GGI_edu_log	Education subindex of Gender Gap Index. Based on gender gaps in literacy rates and enrollment rates in primary, secondary, and tertiary education. Reflected and log-transformed to reduce negative skew.
TertArtsF	Percentage of women among liberal arts majors (humanities and arts).
TertTeachF	Percentage of teachers in tertiary education who are women.
Achievement differences	
TIMSS_diff	Gender difference in Grade 8 TIMSS science achievement averaged across years 1999, 2003, 2007, 2011. Positive values indicate male advantages.
PISA_diff	Gender difference in PISA science achievement averaged across years 2000, 2003, 2006, 2009. Positive values indicate male advantages.
Cultural dimensions	
PowerDist	Power Distance. Represents "the extent to which the less powerful members of institutions and organizations within a country expect and accept that power is distributed unequally" (Hofstede, Hofstede, & Minkov, 2010, p. 61).
UncertAvoid	Uncertainty Avoidance. Represents "the extent to which the members of a culture feel threatened by ambiguous or unknown situations" (Hofstede et al., 2010, p. 191).
MascFem	Masculinity minus Femininity. Masculinity represents "when emotional gender roles are clearly distinct: Men are supposed to be assertive, tough, and focused on material success, whereas women are supported to be more modest, tender, and concerned with the quality of life," whereas femininity represents "when emotional gender roles overlap" (Hofstede et al., 2010, p. 140).
IndivCollect	Individualism minus Collectivism. Individualism represents "societies in which the ties between individuals are loose," whereas collectivism represents "societies in which people from birth onward are integrated into strong, cohesive in-groups" (Hofstede et al., 2010, p. 92).
Atheism_log	Percentage of population that does not believe in a God. Log-transformed to reduce positive skew.
Human development	
HDI_log	Human Development Index. Based on life expectancy at birth, mean years of schooling, expected years of schooling, and gross national income per capita. Reflected and log-transformed to reduce negative skew.
IQ	Nation's average IQ.
Prevalence of scientists	
Rsrcher_log	Number of employed researchers (based on head counts) per one million people. Log-transformed to reduce positive skew.
TertSciPrct	Percentage of tertiary students in science.
World region	
Asia	Dummy code comparing nations in Asia with nations in the Americas.
Europe	Dummy code comparing nations in Europe with nations in the Americas.
Other	Dummy code comparing nations in other world regions (Africa; Oceania) with nations in the Americas. These other regions were combined into one dummy code because of their low frequency in our sample of nations.
Sample characteristics	
critlat_mean	Average trial latency collapsed across experimental conditions of the implicit measure.
prct_male	Percentage of men in the stereotype sample.
prct_college	Percentage of stereotype sample with bachelor's degree or higher.
age_mean	Average age of the stereotype sample.
corr_iatexp	Correlation between implicit and explicit gender-science stereotypes.

Note. TIMSS = Trends in Mathematics and Science Study; PISA = Programme for International Student Assessment.

Received July 7, 2014
Revision received July 28, 2014
Accepted July 28, 2014 ■