

Beliefs About Cognitive Gender Differences: Accurate for Direction, Underestimated for Size

Diane F. Halpern · Carli A. Straight · Clayton L. Stephenson

Published online: 7 November 2010
© Springer Science+Business Media, LLC 2010

Abstract Although stereotype accuracy is a large, and often controversial, area of psychological research, surprisingly little research has examined the beliefs people have about gender differences in cognitive abilities. This study investigates the accuracy of these beliefs in a sample of 106 highly educated U.S. adults. Participants provided estimates of male and female performance for 12 cognitive tasks and games. These estimates were compared with published data on gender differences on the same 12 cognitive tasks and games. Results showed that participants were generally accurate about the direction of gender differences, but underestimated the size of gender differences.

Keywords Gender differences · Stereotype accuracy · Beliefs about gender differences

Introduction

Researchers argue about the extent to which females and males are similar or different in their cognitive abilities, resulting in a broad range of conclusions, from there are

(virtually) no differences (Spelke 2005) to females and males are so different that they learn in qualitatively different ways and need to be educated separately (Gurian et al. 2001). Despite the growing literature on gender similarities and differences in cognitive tasks, surprisingly little research has examined what people believe about these differences and the extent to which these beliefs are supported by research. Major theories in psychology, such as stereotype threat, predict that beliefs about gender differences in cognitive abilities can largely explain observed gender differences. Other prominent stereotype researchers base their predictions on the assumption that stereotypes are exaggerations of real group differences (Allport 1954; de Vries 2004; Fiske 1998; Operario and Fiske 2001; Stangor 2009; Taylor 1981). A second body of research, however, points to the overwhelming accuracy of stereotypes and, hence, challenges the assumptions of stereotypes as inaccurate and as exaggerations of real group differences (see Jussim et al. 2009 for a review). The research described in this article tests two assumptions: 1) people tend to have inaccurate beliefs about cognitive gender differences and 2) people tend to overestimate or exaggerate gender differences in cognitive abilities. These assumptions were tested using a sample of highly educated adults. An important point to mention prior to our review of current literature is that a majority of studies cited in this paper were conducted within the United States and may not generalize to non-western cultures. We note all cases in which research was conducted outside of the U.S., but where mention is not made, the sample is from the U.S.

We asked a group of highly educated U.S. adults to report their beliefs about the performance of males and females on cognitive tasks and games and then compared these data to published figures. The empirical literature on cognitive gender differences reveals that males and females

D. F. Halpern
Department of Psychology, Claremont McKenna College,
850 Columbia Avenue,
Claremont, CA 91711, USA
e-mail: diane.halpern@cmc.edu

C. A. Straight (✉) · C. L. Stephenson
School of Behavioral & Organizational Sciences, Claremont
Graduate University,
123 E. 8th Street,
Claremont, CA 91711, USA
e-mail: carli.straight@cgu.edu

C. L. Stephenson
e-mail: clayton.stephenson@cgu.edu

exhibit different average levels of performance on many, but not all, cognitive tasks. For instance, one of the most consistent findings is that males generally perform better when tasks include visuospatial representation of objects, especially if the task involves mental rotation. This finding holds true in Western (e.g., the U.S.—Johnson and Bouchard 2006; Levine et al. 1999; Loring-Meier and Halpern 1999; Masters and Sanders 1993; Moore and Johnson 2008; Norway—Nordvik and Amponsah 1998; and Canada—Voyer et al. 1995) as well as non-Western cultures such as China (Chan 2007), Ecuador (Flaherty 2005), and Japan (Flaherty). Another consistent finding is that females typically perform better than males when tasks involve writing and grammar skills (Hedges and Nowell 1995; Salahu-Din et al. 2008; U.S. Department of Education 2000), however, this trend is not as widely documented in cultures outside of the United States.

A mountain of research has been devoted to determining why these differences occur because the outcomes have serious implications for public policy decisions as well as the way people think about education, career choices, and the roles of males and females in society. For example, when reports began to emerge that American girls were being “shortchanged” in schools (e.g., American Association of University Women 1995; Sadker and Sadker 1994), educators immediately debated implementing girl-friendly classrooms to encourage the success of girls. A few years later, news articles touted recent research supporting what came to be known as a “boy crisis” (e.g., Sacks 2003). Once again, educational reform was advocated; however, this time it was suggested that teaching methods change to accommodate the short attention spans and “natural” need for high levels of activity for boys. Similarly, cognitive gender differences have been used to explain the dearth of women in math and science careers. For example, boys demonstrate an advantage in mental rotation, and mental rotation is thought to be related to some types of mathematics such as geometry and topology, thus, males are expected to be more successful in these fields (Halpern 2000; Wai et al. 2009).

Stereotypes as Exaggerations of Group Differences

There is a vast body of stereotype research that suggests that empirical cognitive gender differences can be largely explained by widespread beliefs about males and females. This line of reasoning relies on the assumption that stereotypes are, in fact, held by the majority of individuals and that most people believe that differences between groups are larger than they actually are. Specifically, many stereotype theorists suggest that stereotypes are consistently inaccurate (e.g., Brigham 1971; Katz and Braly 1933) and/or “a person overestimates the location of a group on a stereotypical

attribute and underestimates the location of a group on a counterstereotypical one” (de Vries 2004, p. 1286).

The notion of stereotypes as exaggerations of real group differences appears to stem from very early research on stereotyping. Gordon Allport, commonly cited as a founding father of intergroup relation research, originally defined a stereotype as “an exaggerated belief associated with a category” (Allport 1954, p. 191). More recently, researchers maintain that the very notion of group “categorization exaggerates between-group differences and minimizes within-group differences, increasing perceived homogeneity” (Fiske 1998, p. 375; Taylor 1981), and emphasize that “stereotypes overgeneralize, misattribute, prescribe, and often condemn the behavior and personal characteristics” of the targeted group (Operario and Fiske 2001, p. 23). In *The Handbook of Prejudice, Stereotyping and Discrimination*, stereotypes are described as negative, inaccurate, and unfair (Stangor 2009). One of the most popular stereotyping theories, stereotype threat, indeed, relies on the assumption that beliefs about groups influence behavior that would not occur in the absence of such beliefs, hence, exaggerating gender differences.

Stereotype Threat

Stereotype threat research suggests that knowledge of negative stereotypes about one’s group leads to lowered performance on a valued task associated with that stereotype (Steele 1998). This theory is driven by the assumption that behavior is changed due to situational cues, but is not representative of an individual’s true potential. For example, Spencer et al. (1999) found that women performed substantially worse on a difficult mathematics test when under conditions of stereotype threat. Stereotype threat, the authors reasoned, was introduced into the test taking environment because the female test-takers were aware of a negative stereotype about their mathematic abilities and, consequently, performed worse than they would have if there were no threat. In fact, in a second study, these researchers assigned participants to one of two conditions, invoking the threat in one and removing the threat in the other. Females in the threat condition performed substantially worse than females in the non-threat condition (Spencer et al. 1999).

Recent meta-analyses of published stereotype threat literature have corroborated the findings outlined above, demonstrating large and consistent negative effects on performance when women are in situations that would confirm a negative stereotype of their performance (Nguyen and Ryan 2008). In addition, stereotype threat research has been conducted across many different situations and used a variety of priming targets (e.g., race, age, gender, etc.). The central premise of this research is that stereotypes about group differences are based on exaggerations of actual group differences.

Stereotypes as Accurate Representations of Group Differences

Contrary to the stereotype research described above, another body of literature suggests that individuals do not consistently exaggerate empirically tested gender differences. Instead, most people are consistently accurate in their understanding about whether, when, and how much males and females differ. Studies have demonstrated, for example, that participants are generally accurate in their beliefs about nonverbal gender differences (Briton and Hall 1995), male and female attitudes on social and political issues (Diekmann et al. 2002), differences in the motivation and achievement of seventh-grade students by gender, social class, and ethnicity (Madon et al. 1998), the percentage of males and females in gender-typed occupations (McCauley and Thangavelu 1991; McCauley et al. 1998), along with many other phenomena.

Very little research has been conducted on the accuracy of beliefs about gender differences in cognitive abilities. The authors are aware of two such studies. Swim (1994) measured the accuracy of gender stereotypes by comparing perceptions of the sizes of gender differences in social behaviors, nonverbal behaviors, and cognitive abilities with actual meta-analytic findings. She concluded that “subjects did not uniformly overestimate gender differences. The predominant tendency was to be either accurate or to underestimate differences” (p. 30). In fact, other research has found trends toward underestimation as well (Cejka and Eagly 1999; McCauley and Thangavelu 1991). A second study is that of Hall and Carter (1999). These researchers asked college students to decide if each of 77 different traits, including personality, small group behaviors, and cognitive abilities, was more commonly associated with males or females. They found that participants’ ratings of their own beliefs and their ratings of differences based on research were highly correlated. Furthermore, accuracy was high for all 77 traits and ratings were similar for female and male participants.

The Current Study

The current study expands on previous research to measure the accuracy of gender stereotypes about cognitive abilities by comparing participants’ judgments of male and female performance on cognitive tasks and games against empirical data. Stereotypes of this kind are consensual, or group-based, stereotypes, rather than personal, or person-perception stereotypes (Jussim et al. 2009). Participants provided their best estimates about the size and direction of cognitive differences between females and males on measures where there are empirical data against which to compare their judgments. Domains that were included in the survey were deliberately selected where there are empirical data on the differences between male and female performance. Items were also

included where we have no data to support the existence of gender differences. We tested the following predictions, derived from two major assumptions of stereotyping theories, as they apply to beliefs about cognitive gender differences:

- H1: Participants will inaccurately predict the direction of actual gender differences in performance on cognitive tasks and games.
- H2: Participants will exaggerate the size of actual gender differences in performance on cognitive tasks and games.

To test the first hypothesis, we compared participants’ estimates of the relative performance of males and females on 12 cognitive tasks and games (i.e., males perform better than females, females perform better than males, or males and females perform equally) with published data on the actual relative performance of males and females. These tasks and games were 1) the number of words males and females can say with meaning at age 2, 2) the percentage of males and females with reading disorders, 3) males’ and females’ 12th grade writing assessment scores, 4) the percentage of male and female professors in English, history, and foreign languages, 5) males’ and females’ 8th grade science assessment scores, 6) the age at which males and females learn to count to 10, 7) males’ and females’ 4th grade math assessment scores, 8) males and females SAT-M scores, 9) the percentage of male and female professors in math, physics, and engineering, 10) the percent of correct scores on a mental rotation task, 11) the number of past winners in the National Geography Bee, and 12) the number of past winners in the National Scrabble Tournament.

To test the second hypothesis, that participants would exaggerate cognitive gender differences, we compared participants’ estimates of the size of the difference in males’ and females’ performances to the size of differences reported in published literature on the same 12 cognitive tasks and games outlined above.

Although we did not render predictions about differences in the stereotyping of male and female participants, we tested for this effect in all analyses. Our method of asking for the size of a possible gender difference in addition to the direction of the difference is a significant extension of past research because it allows participants to express the belief that some differences are small or nonexistent and others large.

Method

Participants and Recruitment

Participants ($N=106$) were recruited through email, phone, or in person by students enrolled in a graduate course on gender similarities and differences in cognition. Students contacted adult friends, family, and acquaintances via email

asking them to click on a link that connected to an online survey. These initial contacts were encouraged to send the survey link to other individuals, thus initializing a snowball sampling technique. The ensuing sample included 77 females (73%) and 29 males (27%). The education of participants was generally high: 59% attended some graduate/professional school or held a graduate/professional degree, 26% held a 4-year college degree, 3% held a 2-year college degree, 10% completed some college, and 1% completed a high school diploma or GED. Participants ranged in age from 19 to 65 years, with a mean of 32 years. Seventy-six percent identified their ethnicity as White; 9% Asian American/Pacific Islander; 6% Hispanic; 5% Biracial/Multi-Racial; 2% African American; and 8% other. One participant did not specify ethnicity. Because this research is concerned with gender issues, we tested whether there were any differences in education level, age, or race/ethnicity by gender before we proceeded with analyses on our dependent variables. *T*-test and chi-square statistics indicated that there were no differences between males and females on any of these demographic variables.

Materials and Procedures

Participants indicated their consent to participate in the survey by checking a box on the first screen in the survey. The survey required judgments in response to 12 items designed to capture stereotypes concerning cognitive abilities as well as performance in cognitive competitions (e.g., Scrabble tournaments and geography bees). The survey took approximately 15–20 min to complete.

Measures

A short context was provided for each of the 12 items, which we believed would stimulate more thoughtful responses. For example, one item read “Children learn to count in their preschool years. At what age do most girls (boys) learn to count to 10?” This is an example for which we have no data that supports the notion that there are differences between girls and boys.

- Item 1: Participants estimated the average number of words that most 2-year-old boys (girls) can say with meaning. Because we believed that many participants would respond that they did not know, participants were provided with an 8-point scale, anchored by the values 150 and 325, and increasing by intervals of 25 in between. Participants placed a check in the box that corresponded to their answers for both boys and girls.
- Item 2: Participants wrote in their estimates of the percentage of girls (boys) diagnosed with reading disorders.

- Item 3: Participants estimated the relative performance of males and females on a national educational assessment of writing at grade 12. Participants were provided with three options, “Girls and boys scored equally,” “Girls were ahead of boys,” and “Boys were ahead of girls.” They placed a check in the box that corresponded with their answer choice. In addition, if participants answered that either gender was ahead, they also specified by how many months or years.
- Item 4: Participants wrote in their best estimates of the percentage of female college professors at United States research universities in English, history, and foreign languages combined.
- Item 5: Participants wrote in the number of countries, out of 33, in which males outperformed females, and vice versa, on an 8th grade (or equivalent) international science achievement test.
- Item 6: Participants estimated the age at which most girls (boys) learn to count to 10. Participants were provided with a 10-point scale of response choices anchored by 24 and 60 month, and increasing at 4 month intervals in between. They placed a check in the box that corresponded with their answer choices for both boys and girls.
- Item 7: Participants estimated the relative performance of males and females on a test of 4th grade mathematical skills. They were provided with three options, “Girls and boys scored equally,” “Girls were ahead of boys,” and “Boys were ahead of girls.” They placed a check in the box that corresponded with their answer choice. In addition, if participants answered that either gender was ahead, they also specified by how many months or years.
- Item 8: Participants estimated the average score for males (females) on the mathematics portion of the Scholastic Aptitude Test (SAT-M). Participants were instructed to write in a score between 200 (lowest possible score) and 800 (perfect) for females and males.
- Item 9: Participants wrote in their best estimates of the percentage of female college professors at United States research universities in mathematics, physics, and engineering combined.
- Item 10: Participants were asked to think about tasks requiring the ability to imagine what an object would look like from a different angle (i.e., mental rotation). They estimated the percentage of tasks that assessed this ability that females (on average) would perform correctly, followed by what percentage males (on average) would perform correctly.

Items 11 and 12: Two items assessed beliefs about the performance of males and females in various intellectual competitions. Participants were presented with the titles of two national contests in the United States: Scrabble Tournament and Geography Bee. They were asked to consider the last 10 years for each of these competitions and to estimate how many of those years the competitions were won by females and how many were won by males. They wrote their answers in the spaces provided.

Results

To test if there was an effect of the gender of participants on responses to the outcome variables, we conducted a multivariate analysis of variance (MANOVA) with gender as the fixed factor and each of the outcome variables as dependent variables. Using Hotelling's trace statistic, there was no significant effect of gender on estimates of performance on cognitive tasks and games, $T=.679$, $F(37, 68)=1.25$, $p>.05$. Nonetheless, we report results for each item by gender of participant.

Means and standard deviations are reported for estimates of the performance of males and females on each item, with the exception of the two items for which participants selected from the answer choices "boys and girls scored equally," "boys were ahead of girls," and "girls were ahead of boys." In these cases, the distribution of answer choices is reported.

Descriptive data are followed by tests of significance (t -tests or chi-square analyses with Bonferroni correction for multiple tests) to determine whether participants predicted a difference in the scores of males and females on cognitive tasks and games and, if so, in what direction. These tests assess whether Hypothesis One, that participants will be inaccurate about the direction of cognitive gender differences, can be accepted or rejected for each item. In instances where participants estimated significant differences between the groups, we looked to see whether males or females were predicted to have a higher performance. We then compared these predictions with published data to determine whether participants accurately predicted the direction of empirical gender differences. We concluded that estimates were inaccurate where significant differences were predicted between the performance of males and females in a direction that was not consistent with that published in the literature (e.g., females were predicted to perform significantly better than males on the SAT-M, whereas published data indicate that males perform better than females on this test). Similarly, estimates were considered inaccurate when they predicted

differences where published literature does not support any difference. Estimates were considered accurate where the estimated direction of differences (or similarities) accurately reflected the direction of differences (or similarities) reported in published data.

In addition, raw mean difference scores are presented as measures of the size of the difference in male and female performance. These data assess whether Hypothesis Two, that participants will exaggerate the size of cognitive gender differences, can be accepted or rejected for each item. Thus, actual and estimated difference scores are compared to determine over-, under-, or accurate estimations. Where estimates were within 10% of actual mean differences, we determined participants were accurate. Where estimated raw differences differed from actual differences by more than 10%, we determined that participants either over- or underestimated the size of gender differences. Experts recommend raw mean difference comparisons where the data are meaningful and the studies included in the comparison use the same measurement scale (Borenstein et al. 2009).

Means, standard deviations, and raw difference scores are presented in Table 1. P -values are considered significant if they are below .002, which is the Bonferroni adjusted value for conducting 24 tests at the .05 alpha level. Actual p -values are also reported, when they are greater than .001, to allow for alternative interpretations of the level of significance due to the conservative nature of Bonferroni adjustments.

Item 1: According to Cole and Cole (2001), children learn to use 200–300 words by age 2. Girls use an average of 275.1 words, whereas boys use an average of 196.8 words (Lutchmaya et al. 2002). Our data show that participants believed girls ($M=221.7$) use significantly more words with meaning at age 2 than boys ($M=193.9$), $t(105)=8.69$, $p<.001$. Thus, they accurately identified the direction of gender differences, a finding that does not support Hypothesis One. Furthermore, male and female participants did not differ significantly in their estimates of either the number of words 2-year old girls (Male $M=206.9$ and Female $M=227.3$), $t(104)=1.86$, $p=.066$, or boys (Male $M=187.9$ and Female $M=196.1$), $t(104)=.929$, $p=.355$, can say with meaning.

Participants were not accurate in their estimates of the size of these differences, as illustrated by a comparison of the raw difference scores. The actual raw mean difference is equal to 78.3, whereas the estimated raw mean difference is 27.8. Participants' estimates were just 36% of the actual size of differences and, thus, do not support Hypothesis Two because they represent an underestimation.

Table 1 Means, standard deviations, and raw difference scores of actual and estimated gender differences in cognitive abilities and competitions

Survey items	Means (SD)		Raw mean difference scores	
	Actual	Estimated	Actual	Estimated
# Words say with meaning				
Females	275.1 (121.6)	221.7 (50.8)***	78.3	27.8
Males	196.8 (126.8)	193.9 (40.4)		
Reading disorders				
% Females	9.8	12.1 (8.8)***	10.8	5.0
% Males	20.6	17.1 (10.7)		
12th grade writing assessment scores				
Females	160 (37)	Females ahead		
Males	136 (39)			
# Months females ahead			36	13.6 (7.7)
Professors in english, history, and foreign languages				
% Females	40.4	55.0 (17.6)**	19.2	10.0
8th grade science assessment scores				
Females	6	8.7 (6.8)***	22	9.5
Males	28	18.2 (8.9)		
Age learn to count to 10				
Females	36 months	31.7 (7.0)***	0	1.1
Males	36 months	32.8 (7.2)		
4th grade math assessment scores				
Females	239 (28)	Undetermined	2	–
Males	237 (29)			
SAT-M scores				
Females	504 (111)	532.2 (63.3)***	34	33.7
Males	538 (116)	565.9 (68.6)		
Professors in math, physics, and engineering				
% Females	16.2	19.9 (13.7)***	67.6	60.2
% Correct on mental rotation task				
Females	30	62.6 (17.7)***	40	6.7
Males	70	69.3 (16.6)		
Geography bee				
Females	0	4.4 (1.2)***	10	1.2
Males	10	5.6 (1.2)		
Scrabble tournament				
Females	0	5.4 (1.6)**	10	0.8
Males	10	4.6 (1.6)		

Tests of significance are reported in this table within each condition to compare estimates of male and female performance on each cognitive task and game

* $p < .002$ (Bonferroni adjusted p -value)

** $p < .05$ (unadjusted p -value)

^a # of Words Say with Meaning values are means on an 8-point scale, where 1 = “150 words” and 8 = “325 words”, and values increased by 25 points. Reading Disorders values are means on a continuous scale with no anchors. 12th Grade Writing Assessment Scores are reported as the results of a chi-square analysis based on responses to a 3-point scale, where 1 = “Girls and boys scored equally”, 2 = “Girls were ahead of boys”, and 3 = “Boys were ahead of girls”. Professors in English, History, and Foreign Languages values are means on a continuous scale with no anchors. 8th Grade Science Assessment Scores values are means on a continuous scale, with a maximum value of 33. Age Learn to Count to 10 values are means on a 10-point scale, where 1 = “24 months” and 10 = “60 months”, and values increased by 4 points. 4th Grade Math Assessment Scores are reported as the results of a chi-square analysis based on responses to a 3-point scale, where 1 = “Girls and boys scored equally”, 2 = “Girls were ahead of boys”, and 3 = “Boys were ahead of girls”. SAT-M Scores values are means on a continuous scale, with a minimum value of 200 and a maximum value of 800. Professors in Math, Physics, and Engineering values are means on a continuous scale with no anchors. % Correct on Mental Rotation Task values are means on a continuous scale with no anchors. Geography Bee values are means on a continuous scale, with a maximum value of 10. Scrabble Tournament values are means on a continuous scale, with a maximum value of 10

- Item 2: The reported percentages of boys and girls diagnosed with a reading disorder are 20.6% and 9.8%, respectively (Rutter et al. 2004). Consistent with these data, participants indicated a general belief that boys ($M=17.1\%$) are more often diagnosed with reading disorders than girls ($M=12.1\%$), $t(105)=7.30$, $p<.001$. Thus, again, the direction of differences was accurately identified and Hypothesis One was not supported. Male and female participants did not differ significantly in their estimates of either the percentage of girls (Male $M=12.6$ and Female $M=11.9$), $t(104)=-.34$, $p=.735$, or boys (Male $M=16.8$ and Female $M=17.1$), $t(104)=.14$, $p=.89$, diagnosed with reading disorders.
- Participants substantially underestimated the size of the difference by more than half, with an estimated raw difference score equal to 5.0% and an actual raw difference score equal to 10.8%. These data do not support Hypothesis Two.
- Item 3: The most recent report at the time of data collection by the National Assessment of Educational Progress (NAEP), a division of the U. S. Department of Education, indicated that 12th grade girls ($M=160$) outperformed boys ($M=136$) in a writing assessment by 24 points, on a scale ranging from 0 to 300 (Salahu-Din et al. 2008). A chi-square analysis was conducted on our data to compare the distribution of participant responses across three category choices: girls and boys scored equally, girls were ahead of boys, and boys were ahead of girls. The statistic was significant, $\chi^2(2)=98.11$, $p<.001$, and the distribution of responses indicate that participants overwhelmingly believed that girls were ahead of boys on this measure (77% selected girls were ahead of boys, 2% selected boys were ahead of girls, and 21% selected girls and boys scored equally). These data do not support Hypothesis One because the majority of participants were accurate about the direction of differences. In addition, there were no differences in the distribution of responses across category choices by gender of participant, as indicated by a non-significant chi-square statistic, $\chi^2(2)=.773$, $p=.68$.
- Other data indicate that girls in their senior year of high school are approximately 36 months ahead of boys in writing skills (U.S. Department of Education 2000). Participants who selected that girls are ahead in this skill provided an estimate of how far ahead they thought girls were, in months. On this measure, participants underestimated the actual figure ($M=13.6$), $t(81)=26.53$, $p<.001$, by 22 months. These data do not support Hypothesis Two.
- Item 4: Data from the U.S. Department of Education (Forrest Cataldi et al. 2005) indicate that females comprise 40.4% of professors at United States research universities in English, history, and foreign languages. Participants predicted that females ($M=55.0\%$) represent a higher percentage of the total in these positions than males ($M=45.0\%$), $t(105)=2.96$, $p=.004$. Yet, using Bonferroni corrections, this difference was not statistically significant. These data support Hypothesis One because participants predicted that there were no differences in the representation of males and females in these professions, whereas actual data show that males are more highly represented than females. Male and female participants did not differ significantly in their estimates of the percentage of female (Male $M=51.6$ and Female $M=56.3$), $t(104)=1.23$, $p=.220$, professors at United States research universities in English, history, and foreign languages.
- Participants underestimated the size of the difference, estimating a raw difference score equal to 10.0, whereas the actual raw difference score is equal to 19.2. These data do not support Hypothesis Two.
- Item 5: According to data from the Trends in International Mathematics and Science Study (Martin et al. 2005) boys are ahead of girls in tests of 8th grade scientific achievement in 28 of 34 countries measured. Participants accurately identified the direction of gender differences, predicting that boys ($M=18.2$) would perform higher on this test in more countries than girls ($M=8.7$), $t(105)=7.95$, $p<.001$. These data do not support Hypothesis One. Furthermore, male and female participants did not differ significantly in their estimates of the percentage of boys (Male $M=19.7$ and Female $M=17.7$), $t(104)=1.03$, $p=.305$, and girls (Male $M=10.3$ and Female $M=8.1$), $t(104)=1.51$, $p=.134$, who are ahead in tests of 8th grade scientific achievement.
- The mean raw difference score estimated by participants, 9.5, is again a substantial underestimate of the actual raw difference score, 22. These data do not support Hypothesis Two.
- Item 6: Data show that there are no differences in the age at which boys and girls learn to count to ten; both learn this skill around 36 months (Fuson 1988; Geary 2006; Lipton and Spelke 2006; Wynn 1990, 1992). Participants predicted that girls ($M=31.7$) learn at a younger age than boys ($M=32.8$), $t(105)=4.37$, $p<.001$, therefore, participants were not accurate in their estimates of the direction of

differences. Hypothesis One is supported because the actual data indicate that boys and girls learn this skill at approximately the same age. Male and female participants did not differ significantly in their estimates of the age at which boys (Male $M=33.7$ and Female $M=32.5$), $t(104)=.72$, $p=.471$, and girls (Male $M=32.4$ and Female $M=31.4$), $t(104)=.68$, $p=.500$, learn to count to 10.

Participants slightly overestimated the size of the difference between the average age at which boys and girls learn to count to 10. Actual data show that there is no difference, whereas participants predicted that boys were older by 1.1 months. These data support Hypothesis Two.

Item 7: The National Assessment of Educational Progress Report (Perie et al. 2005) indicates that boys and girls score equally well in mathematical achievement at the 4th grade level (237 vs. 239 on a 500-point scale). Similar to the 12th grade writing assessment item, participants selected from three response choices: girls and boys scored equally, girls were ahead of boys, and boys were ahead of girls. The statistic was not significant, $\chi^2(2)=4.09$, $p=.129$, indicating that participants were equally distributed among responses (42% selected girls and boys scored equally, 27% selected girls were ahead of boys, and 30% selected boys were ahead of girls). As a follow-up, we combined the responses that estimated that boys or girls were ahead and tested the new variable against responses that estimated that girls and boys scored equally and conducted another chi-square analysis. The statistic was again not significant, $\chi^2(1)=2.42$, $p=.12$, thus, participants were no more likely to believe boys or girls were ahead than they were to believe that boys and girls scored equally. These results, however, are inconclusive as to whether or not participants were accurate in their estimates of the direction of gender differences. Furthermore, there were no differences in the distribution of responses across category choices by gender of participant, as indicated by a non-significant chi-square statistic, $\chi^2(2)=3.22$, $p=.20$.

Item 8: The 2005 College Bound Report established that girls score an average of 504 points and boys an average of 538 points on the mathematics portion of the Scholastic Aptitude Test (SAT-M; College Board SAT 2005). Participants accurately estimated that boys ($M=565.9$) performed significantly better than girls ($M=532.2$) on this exam, $t(105)=6.11$, $p<.001$. Thus, they accurately estimated the direction of gender differences, which does not support Hypothesis One. In addition, accounting

for Bonferroni adjustments, male and female participants did not differ significantly in their estimates of average SAT-M scores for boys (Male $M=541.9$ and Female $M=575.0$), $t(104)=2.26$, $p=.03$, or girls (Male $M=531.0$ and Female $M=532.6$), $t(104)=.11$, $p=.91$.

Participants were accurate with their estimates of the size of differences, as evidenced by estimated raw mean difference scores equal to 33.7 and actual raw mean difference scores equal to 34.0. These data do not support Hypothesis Two.

Item 9: Because The National Center for Education Statistics (Forrest Cataldi et al. 2005) classifies mathematics and physics under a single natural science label and categorizes engineering by itself, a mean percentage of female professors across the natural sciences and engineering was calculated, resulting in a figure of 16.2%. Participants accurately estimated that the percentage of females ($M=10.9\%$) in these professions is smaller than the percentage of males ($M=80.2\%$), $t(105)=22.66$, $p<.001$. Thus, Hypothesis One is not supported. Male and female participants did not differ significantly in their estimates of the percentage of female (Male $M=22.6$ and Female $M=18.8$), $t(104)=1.27$, $p=.21$, professors at United States research universities in math, physics, and engineering.

Participants slightly overestimated the size of the difference, estimating a raw mean difference equal to 69.3, whereas the actual raw mean difference is equal to 67.6. Although these data appear to support Hypothesis Two, the difference between estimated and actual differences is less than 10%, which was our criterion for determining a meaningful difference. Thus, these data do not support Hypothesis Two, as estimated and actual mean difference scores are similar.

Item 10: The percentages correct on the Vandenberg and Kuse Mental Rotation Test (MRT; Vandenberg and Kuse 1978) are 30% for girls and 70% for boys, respectively. Consistent with these data, participants predicted that boys ($M=69.3\%$) would answer more items correctly than girls ($M=62.6\%$), $t(105)=4.09$, $p<.001$. Thus, they correctly identified the direction of observed gender differences, which does not support Hypothesis One. Male and female participants did not differ significantly in their estimates of the percentage of items males (Male $M=66.7$ and Female $M=70.3$), $t(104)=1.00$, $p=.32$, or females (Male $M=60.9$ and Female $M=63.3$), $t(104)=.62$, $p=.54$, answer correctly on a test of mental rotation.

However, they greatly underestimated the size of the actual difference, estimating a very small mean difference, 6.7, when in fact a very large difference has been reported in the literature, 40.0. These data do not support Hypothesis Two.

- Item 11: No female had won the geography bee in the 10 years prior to 2006 (National Geographic Bee 2006). Consistent with actual data, participants estimated that boys ($M=5.6$) had won more of these tournaments than girls ($M=4.4$), $t(105)=4.99$, $p<.001$. These findings do not support Hypothesis One. Male and female participants did not differ significantly in their estimates of the number of Geography Bee winners who are male (Male $M=5.3$ and Female $M=5.7$), $t(104)=1.62$, $p=.11$, or female (Male $M=4.7$ and Female $M=4.3$), $t(104)=1.67$, $p=.10$.

Although participants identified the correct direction of the difference, estimates of the size of the difference, 0.8, were much smaller than the size of the actual difference, 10.0. These data do not support Hypothesis Two.

- Item 12: The National Scrabble Championship tournament is separated into 6 divisions, where a champion is crowned in each division. We isolated Division 1, which is the most challenging and whose winner receives the highest reward, and recorded the gender of these winners. In this highest ranked division, no female had won in the 10 years prior to 2006 (National Scrabble Championship 2006). Participants predicted that the number of females ($M=5.4$) who had won this tournament was equal to the number of males ($M=4.6$), $t(105)=2.70$, $p=.008$, thus, they were inaccurate in their estimates of the direction of differences. These data support Hypothesis One. Male and female participants did not differ in their estimates of the number of Scrabble Tournament winners who are male (Male $M=4.2$ and Female $M=4.7$), $t(104)=1.38$, $p=.17$, or girls (Male $M=5.8$ and Female $M=5.3$), $t(104)=1.38$, $p=.172$.

Participants substantially underestimated the size of the difference, estimating a difference of 0.8, compared to an actual difference of 10.0. These data do not support Hypothesis Two.

To summarize, participants responded to 12 items concerning their beliefs about cognitive gender differences and similarities. When the estimated and actual direction of differences were compared, participants were accurate with regard to which gender, if any, performed better on 9 of 12 items. Thus, 75% of the items provided evidence

that did not support Hypothesis One. If we consider that one of the items produced inconclusive results, the evidence is even more compelling. The items for which participants inaccurately predicted the direction of gender differences were the percentage of female professors at U.S. universities in English, history, and foreign languages (the percentage of female faculty was predicted to be higher than the percentage of male faculty, whereas the opposite pattern is true) and the number of male and female National Scrabble champions (females and males were predicted to win an equal number of tournaments, whereas males win more tournaments than females).

Results also show that participants underestimated the size of cognitive gender differences in 8 of 12 cases, overestimated in 1 case, and were accurate in 2 cases. Thus, Hypothesis Two was supported by only 1 item, whereas evidence was collected to refute Hypothesis Two by 10 items. The item for which participants overestimated the size of cognitive gender differences was the age at which boys and girls learn to count to 10.

Discussion

Participants demonstrated general knowledge about the relative performance of males and females on cognitive tasks and games. However, they also showed a tendency to underestimate the size of the differences between males and females. These results corroborate the findings of other researchers who have found group-based stereotypes to be either accurate or underestimates of actual differences in both the U. S. and Canada (e.g., Ashton and Esses 1999; Cejka and Eagly 1999; McCauley and Thangavelu 1991).

Why should most people be correct about the direction of cognitive gender differences, but underestimate the size of the differences? The topic of human differences, of any sort, makes many people uncomfortable because differences have historically been the basis for prejudice and discrimination. If females and males really do differ in some aspects of cognition, this information could be and has been misused to justify different educational opportunities and/or affirmative action for either males or females. We understand the justifiable concerns about the misuse and misinterpretation of data on gender differences, and we believe that these concerns influence the way people think about differences. A compromise position for anyone who is uncomfortable with the idea that cognitive gender differences exist is to acknowledge that there are differences, but to decide that they are generally small. We found that most people know whether females or

males tend to excel at different cognitive tasks. Even in those few instances where our participants got the direction of the gender difference wrong, correct knowledge about gender differences can be seen as a possible reason for the error. For example, most people overestimated the percentage of professors in English, history, and foreign languages who are women. These are three academic domains where, in fact, women tend to excel (Willingham and Cole 1997), so it would logically be expected that there would be more women who are professors in these fields than men. Most people do not know that across all fields, there are more men who are professors than women, so they did not use the base rate information to inform their estimates. Similarly, most people overestimated the percentage of winners in Scrabble tournaments who are women. Again, this error reflects the correct knowledge that women tend to have superior language skills (Hedges and Nowell 1995; Willingham and Cole 1997), so once again, this error in predicting the direction of a gender difference can be attributed to a fairly accurate knowledge of the domains in which women and men excel.

Another reason why most people underestimate the size of cognitive gender differences is that, at least for some of the measures, the size has been diminishing over the last few decades and the research literature is mixed in the way that findings are reported. For example, there are many more males than females who achieve the highest scores on the SAT-M, but in the 1980s the ratio of high scoring males to females was 12:1; it is now between 4:1 and 3:1 (Wai et al. 2010). In addition, most people tend to think of differences in broad domains such as mathematics when, in fact, the way males and females vary depends on what is measured and the developmental period of the lifespan that is being assessed. There is little or no gender difference on recent international measures of mathematics such as the TIMSS and PISA (Else-Quest et al. 2010) and females get better grades in school in all subjects including many areas of mathematics (Snyder et al. 2009). Thus, when asked about gender differences on the SAT-M, it is likely that people use what they know about gender and mathematics to reason about the answer.

Implications

An often implicit assumption in stereotype research contends that stereotypes are exaggerations of real group differences. However, our research presents a challenge to this assumption with respect to gender differences in performance on cognitive tasks and games. Rather than exaggerations of empirical gender differences, beliefs about the performance of males and females on cognitive tasks were generally underestimates of these differences. Our results show that stereotypes, or generalized beliefs about

groups of people, in this case females and males, are not necessarily based on unreasonable distinctions between groups. These data support the use of a neutral definition of a stereotype as “a set of beliefs about the personal attributes of a social group” (Ashmore and Del Boca 1981, p. 21).

Socio-cultural explanations of gender differences are certainly valuable and by no means does this research attempt to refute the very robust findings generated by such theories. In fact, gender role stereotyping theory offers a plausible explanation for our findings. Gender role stereotypes are societal beliefs about behaviors and characteristics that are appropriate for each gender (Singleton 1987). These stereotypes are considered to be largely responsible for gender-typed behavior because they both describe and prescribe what is expected of females and males. In that sense, gender role stereotypes accurately reflect the gender-typed roles in a society and at the same time enforce those gender-typed roles (Fiske 1998). Future research should explore the relationship between gender role stereotypes and beliefs about gender differences in cognitive abilities.

Our results bring attention to two common assumptions that are often taken for granted in stereotype research. The current findings that beliefs about groups are largely accurate, but underestimations of the size of empirically verified gender differences should be incorporated into future research in this area. Research on stereotype threat has grown exponentially in recent years with large numbers of studies showing how beliefs about the ways in which males and females differ can affect cognitive performance (Nguyen and Ryan 2008). But without understanding what most people believe about the ways the genders differ, the research paradigm is missing a fundamental component. For example, do beliefs about the size of cognitive gender difference predict the size of a stereotype threat? More specifically, will women who believe that there are large differences in the mathematical abilities of women and men show a greater reduction in their own performance on a high stakes test of advanced mathematics when their gender is made salient than women who believe that the difference is small or nonexistent? This is a key question for future research on the intersection of gender-related stereotypes and test performance. The empirical literature on cognitive gender differences is also growing at a rapid rate, but it may be as important to know what people believe to be true as it is to know about the ways the genders differ and are similar.

The primary limitation to this study relates to the sample, mainly that it was comprised of highly educated adults who may have been more aware of observed gender differences in cognitive abilities than the general population. To generalize beyond the sample used in this research, this study should be replicated in a more representative sample of the general population. Also, future research on stereotype accuracy should be conducted in other cultures

besides the United States and Canada. It is plausible that beliefs about cognitive abilities are influenced by factors that vary by culture, such as social equality, division of labor, and exposure to data regarding gender differences.

Acknowledgement We thank Dr. Gabriel Cook, at Claremont McKenna College, for his comments and suggestions on an earlier draft of this paper.

We thank the students in the graduate course on Cognitive Sex Differences at Claremont Graduate University for their assistance with data collection.

References

- Allport, G. W. (1954/2000). The nature of prejudice. Reprinted in C. Stangor (Ed.), *Stereotypes and prejudice: Essential readings* (pp. 20–48). Philadelphia: Psychology Press.
- American Association of University Women. (1995). *How schools shortchange girls: The AAUW report (1992)*. New York: Marlowe and Company.
- Ashmore, R. D., & Del Boca, F. K. (1981). Conceptual approaches to stereotypes and stereotyping. In D. L. Hamilton (Ed.), *Cognitive processes in stereotyping and intergroup behavior* (pp. 1–35). Hillsdale: Erlbaum.
- Ashton, M. C., & Esses, V. M. (1999). Stereotype accuracy: Estimating the academic performance of ethnic groups. *Personality and Social Psychology Bulletin*, 25, 225–236. doi:10.1177/0146167299025002008.
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. West Sussex: Wiley.
- Brigham, J. C. (1971). Ethnic stereotypes. *Psychological Bulletin*, 76, 15–38. doi:10.1037/h0031446.
- Briton, N. J., & Hall, J. A. (1995). Beliefs about female and male nonverbal communication. *Sex Roles*, 32, 79–90. doi:10.1007/BF01544758.
- Cejka, M. A., & Eagly, A. H. (1999). Gender-stereotypic images of occupations correspond to the sex segregation of employment. *Personality and Social Psychology Bulletin*, 25, 413–423. doi:10.1177/0146167299025004002.
- Chan, D. W. (2007). Gender differences in spatial ability: Relationship to spatial experience among Chinese gifted students in Hong Kong. *Roeper Review: A Journal on Gifted Education*, 29, 277–282. doi:10.1080/02783190709554423.
- Cole, M., & Cole, S. R. (2001). *The development of children* (4th ed.). New York: Worth.
- College Board SAT. (2005). *2005 College bound seniors: Total group profile reports*. Retrieved from <http://professionals.collegeboard.com/data-reports-research/sat/archived/2005>.
- de Vries, R. E. (2004). The use of personality chiasms in stereotype research. *Journal of Applied Social Psychology*, 34, 1281–1305. doi:10.1111/j.1559-1816.2004.tb02007.x.
- Diekmann, A. B., Eagly, A. H., & Kulesa, P. (2002). Accuracy and bias in stereotypes about the social and political attitudes of women and men. *Journal of Experimental Social Psychology*, 38, 268–282. doi:10.1006/jesp.2001.1511.
- Else-Quest, N. M., Hyde, J. S., & Kinn, M. C. (2010). Cross-national patterns of gender differences in mathematics: A meta-analysis. *Psychological Bulletin*, 136, 103–127. doi:10.1037/a0018053.
- Fiske, S. T. (1998). Stereotyping, prejudice, and discrimination. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (4th ed., pp. 357–411). Boston: The McGraw-Hill Companies.
- Flaherty, M. (2005). Gender differences in mental rotation ability in three cultures: Ireland, Ecuador, and Japan. *Psychologia: An International Journal of Psychology in the Orient*, 48, 31–38. doi:10.2117/psysoc.2005.31.
- Forrest Cataldi, E., Fahimi, M., & Bradburn, E. M. (2005). *2004 National Study of Postsecondary Faculty (NSOPF:04) report on faculty and instructional staff in fall 2003* (NCES 2005–172). U.S. Department of Education. Washington: National Center for Education Statistics. Retrieved from <http://nces.ed.gov/pubsearch>.
- Fuson, K. C. (1988). *Children's counting and concepts of number*. New York: Springer-Verlag.
- Geary, D. C. (2006). Development of mathematical understanding. In D. Kuhn, R. Siegler, D. William, & R. M. Lerner (Eds.), *Handbook of child psychology: Vol 2, cognition, perception, and language* (6th ed., pp. 777–810). Hoboken: Wiley.
- Gurian, M., Henley, P., & Trueman, T. (2001). *Boys and girls learn differently!: A guide for teachers and parents*. San Francisco: Jossey-Bass.
- Hall, J. A., & Carter, J. D. (1999). Gender-stereotype accuracy as an individual difference. *Journal of Personality and Social Psychology*, 77, 350–359. doi:10.1037/0022-3514.77.2.350.
- Halpern, D. F. (2000). *Sex differences in cognitive abilities* (3rd ed.). Mahwah: Lawrence Erlbaum Associates.
- Hedges, L. V., & Nowell, A. (1995). Sex differences in mental test scores, variability, and numbers of high-scoring individuals. *Science*, 269, 41–45. doi:10.1126/science.7604277.
- Johnson, W., & Bouchard, T. J. (2006). Gender differences in mental abilities: G masks the dimensions on which they lie. *Intelligence*, 35, 23–59. doi:10.1016/j.intell.2006.03.012.
- Jussim, L., Cain, T. R., Crawford, J. T., Harber, K., & Cohen, F. (2009). The unbearable accuracy of stereotypes. In T. D. Nelson (Ed.), *Handbook of prejudice, stereotyping, and discrimination* (pp. 199–225). New York: Psychology Press.
- Katz, D., & Braly, K. (1933). Racial stereotypes of one hundred college students. *Journal of Abnormal and Social Psychology*, 28, 280–290. doi:10.1037/h0074049.
- Levine, S. C., Huttenlocher, J., Tayler, A., & Langrock, A. (1999). Early sex differences in spatial skill. *Developmental Psychology*, 35, 940–949. doi:10.1037/0012-1649.35.4.940.
- Lipton, J. S., & Spelke, E. S. (2006). Preschool children master the logic of number word meaning. *Cognition*, 98, B57–B66. doi:10.1016/j.cognition.2004.09.013.
- Loring-Meier, S., & Halpern, D. F. (1999). Sex differences in visuospatial working memory: Components of cognitive processing. *Psychonomic Bulletin & Review*, 6, 464–471.
- Lutchmaya, S., Baron-Cohen, S., & Raggatt, P. (2002). Foetal testosterone and vocabulary size in 18- and 24-month-old infants. *Infant Behavior & Development*, 24, 418–424. doi:10.1016/S0163-6383(02)00087-5.
- Madon, S., Jussim, L., Keiper, S., Eccles, J., Smith, A., & Palumbo, P. (1998). The accuracy and power of sex, social class, and ethnic stereotypes: A naturalistic study in person. *Personality and Social Psychology Bulletin*, 24, 1304–1318. doi:10.1177/01461672982412005.
- Martin, M. O., Mullis, I. V. S., Gonzalez, E. J., & Chrostowski, S. J. (2005). TIMSS 2003 international science report. Retrieved from http://timss.bc.edu/PDF/t03_download/T03INTLSCIRPT.pdf.
- Masters, M. S., & Sanders, B. (1993). Is the gender difference in mental rotation disappearing? *Behavior Genetics*, 23, 337–341. doi:10.1007/BF01067434.
- McCaughey, C., & Thangavelu, K. (1991). Individual differences in sex stereotyping of occupations and personality traits. *Social Psychology Quarterly*, 54, 267–279. doi:10.2307/786655.
- McCaughey, C., Thangavelu, K., & Rozin, P. (1998). Sex stereotyping of occupations in relation to television representations and census facts. *Basic and Applied Social Psychology*, 9, 197–212. doi:10.1077/s15324834basp0903_3.

- Moore, D. S., & Johnson, S. P. (2008). Mental rotation in human infants: A sex difference. *Psychological Science*, *19*, 1063–1066. doi:10.1111/j.1467-9280.2008.02200.x.
- National Geographic Bee. (2006). *Past National Geographic Bee winners*. Retrieved from http://www.nationalgeographic.com/geographybee/past_winners.html.
- National Scrabble Championship. (2006). *2006 USSO championship*. Retrieved from <http://www.scrabble-assoc.com/tournaments/2006/usso/history.html>.
- Nguyen, H.-H. D., & Ryan, A. M. (2008). Does stereotype threat affect test performance of minorities and women? A meta-analysis of experimental evidence. *Journal of Applied Psychology*, *93*, 1314–1334. doi:10.1037/a0012702.
- Nordvik, H., & Amponsah, B. (1998). Gender differences in spatial abilities and spatial activity among university students in an egalitarian educational system. *Sex Roles*, *38*, 1009–1023. doi:10.1023/A:1018878610405.
- Operario, D., & Fiske, S. T. (2001). Stereotypes: Content, structure, processes, and context. In R. Brown & S. Gaertner (Eds.), *Blackwell handbook of social psychology: Intergroup processes* (pp. 22–44). Malden: Blackwell Publishing.
- Perie, M., Grigg, W., & Dion, G. (2005). *The Nation's Report Card: Mathematics 2005* (NCES 2006-453). U. S. Department of Education, National Center for Education Statistics. Washington: U.S. Department of Education.
- Rutter, M., Caspi, A., Fergusson, D., Horwood, L. J., Goodman, R., Maughan, B., et al. (2004). Sex differences in developmental reading disability: New findings from 4 epidemiological studies. *Journal of the American Medical Association*, *291*, 2007–2012. doi:10.1001/jama.291.16.2007.
- Sacks, G. (2003). The boy crisis in education: What can be done about it. *Mensight Magazine*. Retrieved from <http://mensightmagazine.com/Articles/Sacks/boycrisis.htm>.
- Sadker, M., & Sadker, D. (1994). *Failing in fairness: How American schools cheat girls*. New York: C. Scribner's Sons.
- Salahu-Din, D., Persky, H., & Miller, J. (2008). *The Nation's Report Card: Writing 2007* (NCES 2008-468). National Center for Education Statistics, Institute of Education Sciences. Washington: U.S. Department of Education.
- Singleton, C. H. (1987). Biological and social explanations of sex-role stereotyping. In D. J. Hargreaves & A. M. Colley (Eds.), *The psychology of sex roles* (pp. 3–26). Washington: Hemisphere Publishing.
- Snyder, T. D., Dillow, S. A., & Hoffman, C. M. (2009). *Digest of educational statistics. Institutes for educational science: National center for education statistics (Table 146)*. Washington: Author.
- Spelke, E. S. (2005). Sex differences in intrinsic aptitude for mathematics and science? A critical review. *The 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.
- Stangor, C. (2009). The study of stereotyping, prejudice, and discrimination within social psychology: A quick history of theory and research. In T. D. Nelson (Ed.), *Handbook of prejudice, stereotyping, and discrimination* (pp. 1–12). New York: Psychology Press.
- Steele, C. (1998). A threat in the air: How stereotypes shape intellectual identity and performance. In J. L. Eberhardt & S. T. Fiske (Eds.), *Confronting racism: The problem and the response* (pp. 202–233). Thousand Oaks: Sage.
- Swim, J. K. (1994). Perceived versus meta-analytic effect sizes: An assessment of the accuracy of gender stereotypes. *Journal of Personality and Social Psychology*, *66*, 21–36. doi:10.1037/0022-3514.66.1.21.
- Taylor, S. E. (1981). A categorization approach to stereotyping. In D. L. Hamilton (Ed.), *Cognitive processes in stereotyping and intergroup behavior* (pp. 83–114). Hillsdale: Erlbaum.
- U.S. Department of Education, National Center for Education Statistics. (2000). *Trends in educational equity for girls and women (NCES 2000-030)*. Washington: U.S. Printing Office.
- Vandenberg, S. G., & Kuse, A. R. (1978). Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual and Motor Skills*, *47*, 599–604.
- Voyer, D., Voyer, S., & Bryden, M. P. (1995). Magnitude of sex differences in spatial abilities: A meta-analysis and consideration and consideration of critical variables. *Psychological Bulletin*, *117*, 250–270. doi:10.1037/0033-2909.117.2.250.
- Wai, J., Lubinski, D., & Benbow, C. P. (2009). Spatial ability for STEM domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology*, *101*, 817–835.
- Wai, J., Cacchio, M., Putallaz, M., & Makel, M. C. (2010). Sex differences in the right tail of cognitive abilities: A 30 year examination. *Intelligence*, *38*, 412–423. doi:10.1016/j.intell.2010.04.006.
- Willingham, W. W., & Cole, N. S. (1997). *Gender and fair assessment*. Mahwah: Erlbaum.
- Wynn, K. (1990). Children's understanding of counting. *Cognition*, *36*, 155–193. doi:10.1016/0010-0277(90)90003-3.
- Wynn, K. (1992). Children's acquisition of the number words and the counting system. *Cognitive Psychology*, *24*, 20–251. doi:10.1016/0010-0285(92)90008-P.