Published in final edited form as:
Psychol Bull. 2010 November ; 136(6): 1123–1135. doi:10.1037/a0021276.

New Trends in Gender and Mathematics Performance: A Meta-Analysis

Sara M. Lindberg, **Janet Shibley Hyde**, and **Jennifer L. Petersen**
University of Wisconsin – Madison

Marcia C. Linn
University of California – Berkeley

Abstract

In this paper, we use meta-analysis to analyze gender differences in recent studies of mathematics performance. First, we meta-analyzed data from 242 studies published between 1990 and 2007, representing the testing of 1,286,350 people. Overall, *d* = .05, indicating no gender difference, and VR = 1.08, indicating nearly equal male and female variances. Second, we analyzed data from large data sets based on probability sampling of U.S. adolescents over the past 20 years: the NLSY, NELS88, LSAY, and NAEP. Effect sizes for the gender difference ranged between −0.15 and +0.22. Variance ratios ranged from 0.88 to 1.34. Taken together these findings support the view that males and females perform similarly in mathematics.

Keywords

mathematics performance; gender; meta analysis

Policy decisions, such as funding for same-sex education, as well as the continuing stereotype that girls and women lack mathematical ability, call for up-to-date information about gender differences in mathematical performance. Such stereotypes can discourage women from entering or persisting in careers in science, technology, engineering, and mathematics (STEM). Today women earn 45% of the undergraduate degrees in mathematics (NSF, 2008a), but women make up only 17% of university faculty in mathematics (NSF, 2008b). We report on a meta-analysis of recent studies of gender and mathematics. We estimate the magnitude of the gender difference and test whether it varies as a function of factors such as age and the difficulty level of the test.

Stereotypes about Gender and Mathematics

Mathematics and science are stereotyped as male domains (Fennema & Sherman, 1977; Hyde, Fennema, Ryan, Frost, & Hopp, 1990b, Nosek, et al, 2009). Stereotypes about female

---

Correspondence concerning this article should be addressed to Janet Hyde, University of Wisconsin, Department of Psychology, 1202 W. Johnson Street, Madison, WI 53706. jshyde@wisc.edu.
Sara M. Lindberg, School of Medicine & Public Health, University of Wisconsin – Madison; Janet Shibley Hyde and Jennifer L. Petersen, Department of Psychology, University of Wisconsin – Madison; Marcia C. Linn, Graduate School of Education, University of California, Berkeley.

inferiority in mathematics are prominent among children and adolescents, parents, and teachers. Although children may view boys and girls as being equal in mathematical ability, they nonetheless view adult men as being better at mathematics than adult women (Steele, 2003). Implicit attitudes that link males and mathematics have been demonstrated repeatedly in studies of college students (e.g., Kiefer & Sekaquaptewa, 2007; Nosek, Banaji, & Greenwald, 2002).

Parents believe that their sons' mathematical ability is higher than their daughters'. In one study, fathers estimated their sons' mathematical “IQ” at 110 on average, and their daughters' at 98; mothers estimated 110 for sons and 104 for daughters (Furnham et al., 2002; see also Frome & Eccles, 1998). Teachers, too, tend to stereotype mathematics as a male domain. In particular, they overrate boys' ability relative to girls' (Li, 1999; but see Helwig, Anderson, & Tindal, 2001).

These stereotypes are of concern for several reasons. First, in the language of cognitive social learning theory, stereotypes can influence competency beliefs or self-efficacy; correlational research does indeed show that parents' and teachers' stereotypes about gender and mathematics predict children's perceptions of their own abilities, even with actual mathematics performance controlled (Bouchey & Harter, 2005; Frome & Eccles, 1998; Keller, 2001; Tiedemann, 2000). Competency beliefs are important because of their profound effect on individuals' selection of activities and environments (Bandura, 1997; Bussey & Bandura, 1999). According to an earlier meta-analysis, girls report lower mathematics competence than boys do, although the difference is not large ($d = +.16$, Hyde et al., 1990b). In recent studies, elementary-school boys still report significantly higher mathematics competency beliefs than girls do (Else-Quest, Hyde, & Linn, 2010; Fredrick & Eccles, 2002; Lindberg, Hyde, & Hirsch, 2008; Watt, 2004).

A second concern is that stereotypes can have a deleterious effect on actual performance. Stereotype threat effects (Steele, 1997; Steele & Aronson, 1995) have been found for women in mathematics. In the standard paradigm, half the participants (talented college students) are told that the math test they are about to take typically shows gender differences (threat condition), and the other half is told that the math test is gender fair and does not show gender differences (control). Studies find that college women underperform compared with men in the threat condition but perform equal to men in the control condition, indicating that priming for gender differences in mathematics indeed impairs girls' math performance (e.g., Ben-Zeev, Fein, & Inzlicht, 2005; Cadinu, Maass, Rosabianca, & Kiesner, 2005; Johns, Schmader, & Martens, 2005; Quinn & Spencer, 2001; Spencer, Steele, & Quinn, 1999). Stereotype threat effects have been found in children as early as kindergarten (Ambady, Shih, Kim, & Pittinsky, 2001). Other research, measuring implicit stereotypes about gender and math, has found that these implicit stereotypes predict performance in a calculus course (Kiefer & Sekaquaptewa, 2007).

Stereotypes play a role in policy decisions as well as personal decision-making. For example, schools and states may base decisions to offer single-sex mathematics classes on the belief that these gender differences exist (Arms, 2007).

Gender and Mathematics Performance

The stereotypes about female inferiority in mathematics stand in distinct contrast to the scientific data on actual performance. A 1990 meta-analysis found an effect size of $d = 0.15$, males scoring higher, for gender differences in mathematics performance averaged over all samples; however, in samples of the general population (i.e. national samples, classrooms – as opposed to exceptionally precocious or low ability samples), females scored higher but by a negligible amount ($d = -0.05$; Hyde, Fennema, & Lamon, 1990a). Hedges and Nowell

(1995), using data sets representing large probability samples of American adolescents, found *d* = 0.03 to 0.26 across the different data sets. Moreover, girls earn better grades in mathematics courses than boys through the end of high school (Dwyer & Johnson, 1997; Kenney-Benson et al., 2006; Kimball, 1989). In short, previous research showed that gender differences in mathematics performance were very small and, depending on the sample and outcome measure, sometimes favored boys and sometimes favored girls.

Several features of the 1990 meta-analysis (Hyde et al., 1990a) warrant more detailed description. Using computerized literature searches, the researchers identified 100 useable studies, which yielded 254 independent effect sizes representing the testing of more than 3 million persons. One key moderator analysis examined the magnitude of gender differences as a function of age and cognitive level of the test (computation was considered the lowest level, understanding of concepts was considered intermediate, and complex problem-solving was considered the highest level). Girls performed better than boys at computation in elementary school and middle school but the differences were small (*d* = −0.20 and −0.22, respectively) and there was no gender difference in high school. There was no gender difference in understanding of mathematical concepts at any age. For complex problem solving, there was no gender difference in elementary or middle school, but a gender difference favoring males emerged in high school (*d* = 0.29). This last gender difference, although small, is of concern because complex problem solving is crucial for STEM careers.

A second moderator analysis examined the magnitude of gender differences in mathematics performance as a function of the ethnicity of the sample (Hyde et al., 1990a). The striking finding was that the small gender difference favoring males was found for Whites (*d* = 0.13), but not for Blacks (−0.02) or Latinos (0.00).

Depth of Knowledge

Traditionally, researchers maintained that girls might do as well as, or even better than boys on tests of computation, which require relatively simple cognitive processes (e.g., Anastasi, 1958). These same researchers concluded that male superiority emerged for tests requiring more advanced cognitive processing, such as complex problem solving. The 1990 meta-analysis by Hyde and colleagues provided some support for these ideas, although the gender difference in complex problem solving did not appear until the high school years and was not large even then.

Current mathematics education researchers conceptualize this issue of complexity of cognitive processes as a question of item demand and the depth of knowledge required to solve a particular problem. Webb (1999) developed a 4-level Depth of Knowledge framework to identify the cognitive difficulty of mathematics items on standardized assessments. In this framework, Level 1 (Recall) includes the recall of information such as facts or definitions, as well as performing simple algorithms. Level 2 (Skill/Concept) includes items that require students to make decisions about how to approach a problem. These items typically ask students to classify, organize, estimate, or compare information. Level 3 (Strategic Thinking) includes complex and abstract cognitive demands that require students to reason, plan, and use evidence. Level 4 (Extended Thinking) requires complex reasoning, planning, developing, and thinking over an extended period of time. Items at Level 4 require students to connect ideas within the content area or among content areas as they develop one problem-solving approach from many alternatives. This depth of knowledge framework was used to rate the cognitive demands of the tests that assess mathematics performance in the studies reviewed here.

New Trends

Cultural shifts have occurred since the 1980s that call for the reexamination of gender differences in mathematics. In the 1980s, a prominent explanation of male superiority in complex problem solving beginning in high school was gender differences in course choice (Meece, Eccles-Parsons, et al, 1982). Girls were less likely than boys to take advanced mathematics courses and advanced science courses. Because mathematical problem solving is an important component of chemistry and physics courses, students may learn those skills in science courses as much as in mathematics courses. Today, however, the gender gap in course taking has disappeared in all areas except physics. For the high school graduating class of 2005, 7.7% of boys and 7.8% of girls took calculus; 57.8% of girls and 50.6% of boys took chemistry; and 32.8% of girls and 36.8% of boys took physics (NSF, 2008c). Insofar as courses taken by students influence their mathematics performance, we would expect that the gender difference in complex problem solving in high school would have narrowed.

In addition, cross-national data show that the gender gap in mathematics performance narrows or even reverses in societies with more gender equality (e.g., Sweden and Iceland), compared with those with more gender inequality (e.g., Turkey) (Else-Quest, Hyde, & Linn, 2010; Guiso, Monte, Sapienza, & Zingales, 2008). Insofar as the United States has moved toward gender equality over the past 30 to 40 years, the gender gap in mathematics performance should have narrowed.

Findings from a recent analysis of data from state assessments of mathematics performance provide evidence that the gender gap in mathematics performance in the U.S. has indeed diminished or even vanished (Hyde, Lindberg, Linn, Ellis, & Williams, 2008). Those data had several limitations and raised some questions that deserve analysis in a larger study. First, they were based just on tests administered by the states to satisfy the requirements of No Child Left Behind Legislation. Items tapping Levels 3 or 4 depth of knowledge were notably absent. Second, the data were derived only from students in grades 2 through 11; trends in gender differences beyond grade 11 (age 17) could therefore not be assessed. Third, the distributions of male and female performance were available for only part of the sample. The available results raised intriguing questions about gender differences in complex problem solving since in some subgroups females outperformed males at the high end of the distribution.

Gender and Variability

Most of the research has focused on mean-level gender differences, but variability (variance) remains an issue even when means are similar. The greater male variability hypothesis was originally proposed in the 1800s and advocated by scientists such as Charles Darwin and Havelock Ellis, to explain why there was an excess of men both in homes for the mentally deficient and among geniuses (Shields, 1982). In modern statistical terms, the hypothesis is that, independent of mean-level differences, males have a greater variance than females do on the intellectual trait of interest. Thus, the hypothesis states that men are more likely than women to be at both the top and the bottom of the statistical distribution of mathematics performance. Typically the statistic that is computed is the variance ratio (VR), the ratio of the male variance to the female variance. Thus values > 1.0 indicate greater male variability. Based on test norming data, Feingold (1992a) found a VR of 1.11 for the DAT numerical ability, 1.20 for the SAT-Math, and 1.02 for the WAIS Arithmetic subtest. Hedges and Nowell (1995) found VR's ranging between 1.05 and 1.25 for mathematics tests administered to national samples of adolescents such as the NLSY and NELS:88. The analysis of recent state assessment data, described earlier, found VR's ranging between 1.11 and 1.21 (Hyde et al., 2008). Thus there is some evidence of greater male variability in mathematics performance, although the variance ratios are not terribly lopsided. The greater male variability hypothesis, of course, is a description of the data, not an explanation for it, but if true, it could partially account for findings of an excess of males at very high levels of mathematical performance (Hedges & Friedman, 1993). One goal of this meta-analysis was to re-assess the greater male variability hypothesis for mathematics performance using contemporary data.

The Current Study

Several factors warrant a new meta-analysis of research on gender and mathematics performance. First, approximately 18 years of new data have accumulated since the 1990 meta-analysis (Hyde et al., 1990a). Second, cultural shifts have occurred over the last two decades. Specifically, girls are now taking advanced mathematics courses and some science courses in high school at the same rate as boys are, closing the gap in course choice. The magnitude of gender differences in mathematics performance is expected to be even smaller than it was in the 1990 meta-analysis; of particular interest is the gender difference favoring boys in complex problem-solving in high school, and whether this difference has narrowed in recent years. Third, statistical methods of meta-analysis have advanced. At the time of the 1990 meta-analysis, only fixed-effects models were available. Fixed-effects models have since been criticized, and random-effects and mixed models have been developed (Hedges & Vevea, 1998; Lipsey & Wilson, 2001). The current meta-analysis used a mixed-effects model, the advantages of which are detailed below.

Our goals in these meta-analyses were to provide answers to the following questions:

1. What is the magnitude of the gender difference in mathematics performance, using the *d* metric?
2. Does the direction or magnitude of the gender difference vary as a function of the depth of knowledge tapped by the test?
3. Developmentally, at what ages do gender differences appear or disappear?
4. Are there variations across U. S. ethnic groups, or across nations, in the direction or magnitude of the gender difference?
5. Has the magnitude of gender differences in mathematics performance declined from 1990 to 2007?
6. Do males display greater variance in scores and, if so, by how much?

Study 1 addressed these questions using traditional methods of meta-analysis that involve identifying all possible studies using article databases. Study 2 addressed the questions using an alternative method advocated by Hedges and Nowell (1995), which involves the analysis of recent large, national U.S. data sets based on probability sampling. Cross-national, probability sampled data sets have been analyzed by others (Else-Quest, Hyde, & Linn, 2010; Guiso et al., 2008; Penner, 2008).

Study 1

Method

Identification of studies—Computerized database searches of ERIC, PsycINFO, and Web of Knowledge were used to generate a pool of potential articles. To identify all articles that investigated mathematics performance, the following search terms were used: (math* or calculus or algebra or geometry) AND (performance or achievement or ability) NOT (mathematical model). This broad term was selected to capture the widest range possible of research conducted on this topic while avoiding studies that used computational modeling methodology to study unrelated phenomena. Search limits restricted the results to articles that discussed research with human populations and that were published in English between 1990 and 2007. The three database searches identified 10,816, 9,577, and 18,244 studies, respectively, which were considered for inclusion. Given the tremendous volume of studies identified, we decided to rely solely on this method of identification, a potential limitation.

Abstracts and citations were imported into RefWorks citation manager, and then each article was evaluated for inclusion based on the following criteria: (a) the title or abstract alluded to a measure of mathematics performance; (b) the study appeared to contain original data; (c) the study was conducted on a human population; (d) the sample included at least five males and five females. 3,941 studies met the aforementioned criteria. These articles were then printed and examined to determine whether they presented sufficient statistics for an effect size calculation. The final sample of studies included in study 1 utilized data from 242 articles, comprising 441 samples and 1,286,350 people.

Although the final sample of studies is large, some readers may wonder why so many potential studies were lost during the coding processes. Most cases of exclusion were based on one of three reasons: (1) Many of the articles identified in our search reported on data from the large national datasets included in Study 2 (NAEP, NELS, LSAY, NLSY) or from large international datasets (SIMS, TIMSS, PISA) that had been covered by other meta-analyses of gender differences in mathematics performance (Else-Quest, Hyde, & Linn, 2010; Guiso et al., 2008; Penner, 2008). Those articles were excluded from study 1 to avoid redundancy. (2) The searches also picked up reports published by state, local, and federal educational organizations, few of which contained individual-level data. (3) Finally, many articles did not contain enough data to compute effect sizes (e.g., results were not disaggregated by gender) and additional data could not be obtained from the study authors.

Coding the studies—If studies reported data from large datasets or longitudinal studies that were likely to create multiple publications, this was noted so as to avoid inclusion of non-independent effect sizes.

Several characteristics of each sample were coded as potential moderators: (a) age of the participants; (b) nationality of participants (American, Canadian, European, Australian/New Zealander, Asian, African, Latin American, or Middle Eastern); (c) for US samples, majority ethnic group of participants (Euro-American, African American, Asian American, Hispanic, other; mixed; or unreported), and (d) ability level of the participants (low ability, general ability, selective, highly selective).

Several aspects of the mathematics tests were also coded as potential moderators: (a) whether the test was time-limited; (b) whether the test included each of several problem types (multiple choice, short answer, open ended); (c) which types of mathematical content were included in the test (numbers and operations, algebra, geometry, measurement, data analysis and probability); (d) depth of knowledge (level 1 recall, level 2 skills/concepts, level 3 strategic thinking, level 4 extended thinking; levels are described in greater detail above, see also Webb, 1999) and (e) whether the test was specific to the local curriculum (i.e. based on published curricular standards for the region or developed in collaboration with local teachers, textbooks, and syllabi) or was relatively independent of the curriculum. Publication year was also coded as a potential moderator.

50 articles were double coded to determine inter-rater reliability. Percentage agreement for coding of the moderators was > 95% for all variables.

Effect size computation—Cohen's *d* (Cohen, 1988) is the effect size for the standardized mean difference between two groups on a continuous variable (e.g., the mean difference between males and females on a continuous measure of mathematics performance). Thus, for each independent sample within an article, *d* was computed, with $d = (M_m - M_f) / s_w$ and M_m = the mean for males, M_f = the mean for females, and s_w = the pooled within-gender standard deviation. Whenever possible, separate effect sizes were computed for independent groups within each sample (e.g., different age groups, Blacks and Whites). If means and standard deviations were not available, the effect size was computed from other statistics such as *t* or *F*, using formulas provided by Lipsey and Wilson (2001). The complete list of samples, with corresponding effect sizes and variance ratios, can be found in the supplemental online material for this article.

In a few cases, more than one effect size was available for the same sample. When that happened, the following decision rules were used to select or calculate a single effect size for inclusion in the analyses: (1) If a study used a longitudinal design, we used the effect size from the most recent measurement. (2) If multiple effect sizes were available from the same time point but missing data produced different sample sizes for different measures, then the effect size with the largest corresponding sample size was selected for inclusion in the analyses. (3) If there were multiple effect sizes from the same time point with the same sample size, means and standard deviations were pooled, and the pooled values were used to compute the composite effect size that was included in the analyses.

Raw effect sizes were corrected for bias, and standard errors were calculated. Specifically, estimated population effect sizes were used, which adjust for upward-bias of effect sizes among small samples (Hedges, 1981). In addition, inverse variance weights were calculated for each effect size so that analyses could be weighted by inverse variance, a procedure that allows large samples to have more leverage in the analyses than small samples have (Lipsey & Wilson, 2001).

Variance ratio computation—Variance ratios were also computed for each study, with $VR = \text{variance}_{\text{males}} / \text{variance}_{\text{females}}$. Thus, a $VR > 1$ denotes greater male variability, and a $VR < 1$ denotes greater female variability. Standard errors and inverse variance weights were calculated for each variance ratio. Analyses of the variance ratios were conducted as outlined by Katzman and Alliger (1992).

Data analyses—Results were analyzed using a mixed-effects model (Lipsey & Wilson, 2001) and SPSS macros by Wilson (2005). The mixed-effects model assumes that variability among effect sizes can be explained by both fixed factors (i.e. systematic differences due to moderators) and random factors (i.e. error variance). This approach is preferable to fixed-effects or random-effects models, which assume that all variability among effect sizes is accounted for by moderators or by error, respectively. In mixed-effects models, a random-effects variance component is computed based on the residual homogeneity after moderator effects have been accounted for. Then, inverse variance weights are recalculated with the random-effects variance component added, and the model is refit.

To measure heterogeneity of variance of the effect sizes, Q statistics were computed using weights based on the mixed-effects model. When homogeneity analyses indicated that there was significant heterogeneity among the effect sizes, moderator analyses were conducted to test whether characteristics of the mathematics assessments or characteristics of the samples could explain the variability among effects. These moderator analyses allowed us to explore whether systematic differences among the studies led to reliably different effects. Moderators were tested for significance using an analog to ANOVA in the case of categorical moderator variables (e.g., ethnicity), or an analog to multiple regression in the case of continuous moderator variables (e.g., publication year). The level of missing data for moderator variables (due to vague descriptions of mathematics measures and study procedures) made it untenable to conduct a simultaneous analysis of all moderators. The piece-wise approach is not ideal, but we believe that it is warranted in this case, because we were testing a priori hypotheses based on the findings of previous meta-analyses.

Results

Magnitude of gender differences—The overall weighted effect size, averaged over all studies, was $d = +0.05$, representing a negligible gender difference. Figure 1 shows the distribution of effect sizes, which is approximately normal and centered around 0. Heterogeneity analysis revealed that the set of effect sizes was significantly heterogeneous, Q_t (427) = 11478.74, $p < .001$. The random effects variance component was .070.

Moderator analyses—Given the heterogeneity among the effect sizes, we conducted analyses for suspected moderator variables. Table 1 displays the analyses for variations in effect size as a function of characteristics of the tests. Problem type (presence of multiple choice, short answer, and open ended questions) was the only aspect of the tests that significantly predicted heterogeneity among the effects, Q_b (3, 178) = 8.11, $p < .05$. The presence of multiple choice questions on exams predicted relatively better performance by males (β = .16), whereas the presence of short answer and open ended questions predicted relatively better performance by females (βs = −.02 and −.06, respectively).

The magnitude of the gender difference did not depend on whether there was a time limit for the test Q_b (1, 136) = 2.21, $p = .14$, or whether the test was curriculum-focused, Q_b (1, 422) = 3.01, $p = .08$. Similarly, there were no variations in the magnitude of the effect size as a function of problem content (numbers and operations, algebra, geometry, measurement, probability), Q_b (5, 204) = 8.55, $p = .13$, or depth of knowledge, Q_b (3, 119) = 1.51, $p = .68$.

Table 2 displays the analyses for variations in effect size as a function of characteristics of the sample. The magnitude of the gender difference varied significantly as a function of the selectivity of the sample, Q_b (3, 424) = 27.06, $p < .001$. For samples of the general population, $d = +0.07$, but $d = +0.40$ for highly selective samples.

Nationality was not a significant predictor of effect sizes, Q_b (7, 421) = 10.12, $p = .18$. All effects were small or negligible. Among US studies, effect sizes varied as a function of ethnicity, Q_b (1, 89) = 10.00, $p < .01$. Samples composed mainly of Whites showed $d = +0.13$, whereas for ethnic-minority samples, $d = −0.05$. Although we would have preferred to report results from these groups separately by ethnicity, the number of samples was too small to permit this.

The analysis also indicated that age was a significant moderator, Q_b (5, 423) = 44.75, $p < .001$. Gender differences were negligible in elementary-school and middle-school-aged children and reached a peak of $d = +0.23$ in high school. The gender difference then declined for college-age samples and adults.

To test for trends over time from 1990 to 2007, an analog to multiple regression was performed, using year of publication to predict effect size. This analysis indicated that publication year was not a significant predictor of effect size, Q_b (1, 427) = 1.05, $p = .31$.

A final analysis explored the interaction of age and depth of knowledge. Thus, our next analysis focused on both depth of knowledge and different age groups. This analysis was limited by the fact that many articles identified in our original search provided insufficient information to be able to code depth of knowledge, so they were not useable in this analysis.

Furthermore, of those that had usable information about test content, only a small proportion of tests included items that tapped complex problem solving (Level 3 or 4). Some studies located in the literature search involved problem solving at Level 3 or 4 but reported only qualitative data on students' approach to the problem, with no data on actual performance. Although these studies could not be included in the meta-analysis, they suggest the value of looking at more complex tasks. Table 3 provides a breakdown of effect sizes by age and depth of knowledge. Of particular interest in this analysis was whether a gender difference in complex problem solving would be seen among high school and college students, as was found in Hyde's 1990 meta-analysis. Our results showed that there was a small gender difference favoring high school males on tests that included problems at Levels 3 or 4 ($d = +0.16$), but the effect was reversed among college students (d = −0.11). However, these findings are based on small numbers of studies, and therefore cannot be considered robust.

Gender differences in variability—A mixed-effects analysis of gender differences in variance was conducted in parallel to the effect size analyses reported above. The overall weighted variance ratio, averaged over all studies, was $VR = 1.07$, indicating a slightly larger variance for males than for females. The residual variance component was .073.

Study 2

Method

Large United States datasets—Study 1 excluded articles that reported secondary data analyses from large national datasets, because original data from those studies were acquired directly for a separate analysis, which constitutes Study 2. Datasets were included in Study 2 if (a) they included relevant information about math performance, (b) represented data collected after 1990, (c) were nationally representative with a large sample size, and (d) provided statistics for both males and females. International datasets were excluded from Study 2 because they have been thoroughly reviewed elsewhere in the literature (see Else-Quest, Hyde, & Linn, 2010). The following large U.S. datasets were analyzed in Study 2: The National Longitudinal Survey of Youth - 1997 (NLSY97, U.S. Bureau of Labor Statistics, n.d.), The National Educational Longitudinal Study (NELS88, National Center for Educational Research n.d.a), The Longitudinal Study of American Youth (LSAY, n.d.), and the National Assessment of Educational Progress (NAEP, National Center for Education Research, n.d.b).

The *NLSY - 97* began data collection in 1997 and followed students each year until 2002. At the first assessment 58.3% of participants were White, 27.0% were Black, 1.7% were Asian American, 0.8% were American Indian, and 12.4% did not report ethnicity. Math achievement was measured using the PIAT-R (Markwardt, 1998). During round one of data collection, the test was administered to all participants who were in the ninth grade or lower. During round two of data collection, the test was administered to all participants who had taken the test during round one, as well as those who were at least 12 years old on December 31, 1996. The PIAT-R consists of multiple choice items about three areas of math content: 1) foundations (i.e. number, size, and shape discretion), 2) basic facts (i.e. addition, subtraction, multiplication, division), and 3) applications (i.e. algebra, geometry, fractions, word problems, and numerical relationships). The PIAT-R math assessment begins with an age-appropriate question and increases or decreases in difficulty until the youth establishes a “basal,” that is, when the youth correctly answers five consecutive questions. Once a basal is reached the questions increase in difficulty until a “ceiling” is reached when the youth incorrectly answers five of seven consecutive questions. The ceiling is then adjusted for incorrect responses given between the basal and the ceiling and is standardized with a mean of 100 and a standard deviation of 15. The current study determines math achievement using the standardized score of the PIAT-R math for each assessment.

The *NELS88* is a longitudinal study which began examining 8th graders in 1988 and followed these youth in 10th and 12th grade in 1990 and 1992. At the first assessment 67.0% of participants were White, 12.2% were Black, 12.7% were Latino, 6.3% were Asian American, and 1.6% did not report ethnicity. The NELS math assessment was developed by Educational Testing Services and consisted of multiple choice questions. Item content included arithmetic, algebra, geometry, data and probability, and advanced topics. One version of the test was administered at the base year, and three versions of the test were administered at the first and second follow up. Based on their performance on the math test in the base year, students were divided into three groups (low, moderate, and high ability) and were assigned versions of the math test at the first and second follow up in accordance with their ability. Each test, regardless of ability, assessed skill/knowledge, comprehension, and problem solving in all five content areas described above. In addition to the multiple choice test, a constructed response test was given to 12th graders at the 1992 assessment; this test involved items examining measurement, geometry, and data analysis.

The *LSAY* followed youth from 1987 to 1992 with assessments at each year from grades 7 to 12. At the first assessment 70% of participants were White, 11% were Black, 9.2% were Latino, 3.5% were Asian American, 1.5% were American Indian, and 4.8% did not report ethnicity. At each assessment students completed a multiple-choice math test that assessed skills in geometry, measurement, data analysis, algebra, and simple operations (for a complete list of problems, see http://lsay.msu.edu/instruments_006.html).

The *NAEP* math assessment was the only large U.S. database in the current study that was not longitudinal. It consisted of two different studies: the long-term trend assessment, and the main assessment.

The *NAEP long-term trend assessment* was given every four years from 1992 to 2004 to students aged 9, 13, and 17. Ethnic diversity was different for each assessment and age group, but consisted of Whites (M = 74.07%, SD = 4.87), Blacks (M = 14.78%, SD = 1.31), Latinos (M = 7.82%, SD = 3.31) and those who did not report ethnicity (M = 3.29%, SD = 1.38). The math assessment in this study has remained virtually unchanged since its inception in 1978. It included both multiple-choice and short constructed response items which focus on math skills including number operations, measurement, algebra, and geometry.

In contrast to the long-term trend data, the *NAEP main assessment* selected students by grade rather than by age. The main assessment was given to 4th and 8th graders every two years from 1990 to 2007. Twelfth graders were included in 1990 through 2000. Ethnic diversity was different for each assessment and age group, but consisted of Whites (M = 67.07%, SD = 6.43), Blacks, (M =15.94%, SD = 0.99), Latinos (M = 11.89%, SD = 4.89), Asian Americans (M = 3.17%, SD = 1.66), American Indians (M = 0.94%, SD = 0.42) and those who did not report ethnicity (M = 3.29%, SD = 1.38). The math portion of the NAEP main assessment used multiple choice, short constructed response, and long constructed response items. In addition to the math skills assessed in the long-term trend analysis, the main analysis also included skills in data analysis and probability.

A unique aspect of the NAEP assessments is that performance scores are constructed via IRT (item response theory). Thus, they are not a direct reflection of the *number* of problems any student got right or wrong. Rather, the *pattern* of correct responses is used to construct a “probable value” score that reflects the student's overall understanding of mathematics. See Mislevy, Johnson, and Muraki (1992) for a further discussion of this approach and its advantages in estimating population values. Our analyses are based on these probable value scale scores, along with the corresponding weights and standard errors reported by NAEP.

Data analysis—Data analysis for Study 2 was similar to Study 1. All effect sizes were calculated, and a mixed effects model was used to determine whether the effect sizes within each dataset were heterogeneous. If effect sizes were heterogeneous, a weighted ordinary least squares regression was applied to predict gender differences in math performance.

Moderating variables used in Study 2 were age, publication year, percentage of each type of problem (number sense, algebra, geometry, measurement), percentage of problems in each type of format (multiple choice, short answer, open ended), and percentage of Whites, Blacks, and Latinos in each sample. Similar to Study 1, variance ratios were also computed. More information about sample and test characteristics were available for the large datasets than were available for the studies uncovered in the literature reviews in Study 1. Therefore, with the exception of depth of knowledge, we were able to code moderator variables with more detail and many moderators that were coded as categorical variables in Study 1 were considered continuous variables in Study 2. For example, Study 1 coded whether the tests used multiple choice, short answer, or open ended questions, whereas Study 2 coded the percentage of question in each format.

Results

Across all datasets in Study 2, the average weighted effect size was $d = +0.07$. The average weighted variance ratio across all datasets was 1.09. The effect sizes were heterogeneous, $Q_t(55) = 393.04$, $p < 0.001$, with a random effects variance component of .001. Differences among the national datasets were a significant source of heterogeneity $Q_b(4, 55) = 43.12$, $p < .001$; therefore we describe findings from each dataset in turn.

Effect sizes for each assessment of the NLSY are presented in Table 4. The mean weighted effect size for all six assessments of the NLSY was $d = +0.08$. The average weighted variance ratio was 1.05. Effects sizes for NLSY-97 were homogenous, $Q_w(5) = 5.72$, $p = .33$.

Effect sizes for each assessment of the NELS:88 are presented in Table 5. The average weighted effect size across all eight assessment was $d = +0.10$. The average weighted variance ratio for the NELS:88 was 0.94. Effect sizes within the NELS:88 were heterogeneous, $Q_w(7) = 18.66$, $p < 0.01$.

Effect sizes for each assessment of the LSAY are presented in Table 6. Results for the LSAY indicated small or negligible gender differences for each assessment. The average weighted effect size for all six assessments was $d = -0.07$. The weighted average variance ratio was 1.26. Effect sizes were homogenous, $Q_w(5) = 2.50$, $p = .78$.

Effect sizes for each assessment of the NAEP are presented in Table 7. Results for NAEP indicated small or negligible gender differences at all grades. The average weighted effect size across all 18 assessments of the long-term trend data was $d = +0.09$. The average weighted effect size across all 18 of the main assessments was $d = +0.06$. The average weighted variance ratio for the long-term trend data was 1.13 and for the main assessment was 1.04. Effect sizes for both the long-term trend data and the main assessment were homogenous, $Q_w(17) = 16.05$, $p = .52$ and $Q_w(17) = 12.88$, $p = .74$, respectively.

The heterogeneity of the effect sizes across datasets indicates that these studies are not replications of each other but rather vary along some dimension(s). We therefore conducted additional moderation analyses to examine whether sample characteristics or test characteristics could explain the heterogeneity among effect sizes across datasets. These analyses were conducted using an analog to multiple regression, in which each level of each moderator was entered as a separate predictor of the studies' effect sizes. The resulting betas can be interpreted much like correlation coefficients, with positive values indicating an increase in effect size as the value of the moderator increases (relative advantage for males) and negative values indicating a decrease in effect size as the value of the moderator increases (relative advantage for females).

As seen in Table 8, two aspects of the tests accounted for heterogeneity among studies: problem type and mathematical content. With regard to problem type, tests with a higher proportion of multiple choice and open-ended items yielded smaller gender effect sizes, whereas tests with a higher proportion of short answer items yielded larger gender effect sizes. This finding surprised us, given that three of the big datasets (LSAY, NLSY, NELS) were 100% multiple choice, and only one of them had a negative overall effect size; if multiple choice items confer a significant female advantage, we might have expected negative effects across all three of those studies. Therefore, we conducted an additional analysis, looking just at the 36 NAEP effect sizes, which have variation in the proportion of multiple choice, short answer, and open-ended questions. When looking at just the NAEP effect sizes, we found a different pattern of results, such that males did better on tests with a greater proportion of multiple choice items ($\beta = +.29$), and females did better on tests with a greater proportion of short answer and open-ended items (βs = $-.32$ and $-.19$, respectively). Thus, problem type had a similar effect on gender differences in the NAEP as were found in study 1.

With regard to mathematical content, tests with a higher proportion of algebra items yielded smaller effect sizes (females performed relatively better), and tests with a higher proportion of measurement items yielded larger effect sizes (males performed relatively better). The other three types of mathematical content were not significant predictors of effect size in this analysis.

With regard to depth of knowledge, tests containing items at levels 3 or 4 yielded larger effect sizes (males performed better or females performed worse). All of the tests contained items at levels 1 and 2, and therefore we were not able to examine the specific effects of items at those levels.

The ethnic composition of the samples did not have an effect on the magnitude of the gender difference in mathematics performance. However, age was a marginally significant predictor of effect size ($p = .0516$), with older samples yielding relatively larger gender differences favoring males.

Discussion

We proposed to answer six questions with these meta-analyses. We take up each question in turn.

First, what is the magnitude of the gender difference in mathematics performance, based on contemporary studies? Taking Study 1 and Study 2 together, the answer appears to be that there is no longer a gender difference in mathematics performance. For Study 1, d values averaged +0.05 based on data from 1,286,350 persons. For Study 2, d values averaged +0.07 based on data from 1,309,587 persons. These results are consistent with a recent analysis of U.S. data from state assessments of youth in grades 2 through 11, which found that girls had reached parity with boys in math performance (Hyde, Lindberg, Linn, Ellis, & Williams, 2008).

Second, does the direction or magnitude of the gender difference vary as a function of the depth of knowledge tapped by the test? By itself, depth of knowledge was not a significant predictor of differences in effect sizes in study 1. However, study 2 indicated that there may be a modest effect of depth of knowledge on gender differences in mathematics performance, with those containing a greater proportion of items at levels 3 or 4 favoring males. These results are inconsistent with a recent analysis of NAEP data, examining gender differences for items categorized as difficult by NAEP, and as at Level 3 or 4 Depth of Knowledge; at 12th grade, the average *d* = +0.07, or a negligible difference (Hyde et al., 2008). However, an examination of depth of knowledge and age simultaneously in Study 1 (Table 3) indicates a male advantage (d = 0.16) in Level 3 or 4 problems in high school, a finding that is consistent with the earlier meta-analysis by Hyde and colleagues (1990). This finding, however, is based on only 3 studies, so it should be interpreted with caution. Very few studies used items requiring this greater depth of knowledge, yet it is precisely the skill that is required for high-level STEM careers.

Third, developmentally, at what ages do gender differences appear or disappear? Consistent with previous meta-analyses, are gender differences larger in high school than in elementary or middle school? The data sets reviewed in Study 2 showed a marginally significant increase in effect sizes as age increased ($\beta = +.24$). This is consistent with the results of Study 1, which found gender differences close to 0 for elementary and middle school students, and small effects favoring males for high-school and college students (*d*s= +0.23 and +0.18, respectively). These results, too, are inconsistent with the Hyde et al. (2008) analysis of data from state assessments, which showed no gender difference in performance at any grade level through grade 11. Again, though, it is important to consider age and depth of knowledge required by the test simultaneously.

Overall, we conclude that a small gender difference favoring males in complex problem solving is still present in high school. Multiple factors may account for this gender gap. As noted earlier, girls are less likely to take physics than boys are, and complex problem solving is taught in physics classes, perhaps even more than in math classes. Gender differences in patterns of interest may play a role (Su, Rounds, & Armstrong, 2009), although these patterns, too, are shaped by culture. Moreover, even in very recent studies, parents and teachers give higher ability estimates to boys than to girls (Lindberg, Hyde, & Hirsch, 2008), and the effects of parents' and teachers' expectations on children's estimates of their own ability and their course choices are well documented (Eccles, 1994; Jacobs, Davis-Kean, Bleeker, Eccles, & Malanchuk, 2005).

Fourth, are there variations across U.S. ethnic groups, or across nations, in the direction or magnitude of the gender difference? In regard to ethnicity, Table 2 shows that, for Study 1, a small gender difference was found favoring males among Whites, *d* = +0.13, but for all ethnic minorities combined, *d* = −.05. This result is similar to the one found in the 1990 meta-analysis (Hyde et al., 1990a), which found a small gender difference favoring males among Whites, but no difference for ethnic minority groups. Table 2 also shows variation in effect sizes according to nationality or region of the world. The largest gender difference favoring males was found in studies from Africa, *d* = +0.21, but even this difference is small. The largest difference favoring females was found in Central/South America and Mexico, *d* = −0.06. These variations in the magnitude and direction of the gender difference in math performance are consistent with those found in analyses of international data sets such as PISA and TIMSS. These other studies have, in addition, found that values of *d* for nations correlate significantly with measures of gender inequality for those nations (e.g., Else-Quest et al., 2010; Guiso et al., 2008; Penner, 2008).

Fifth, has the magnitude of gender differences in mathematics performance declined from 1990 to 2007? Study 1 found no relation between year of publication and effect sizes, indicating no discernible trend over time toward smaller gender differences. This may be because of the fact that, even in 1990, gender differences were already small (Hyde et al., 1990a), leaving little room for further decline.

Sixth, do males display greater variance in scores and, if so, by how much? The overall variance ratio in Study 1 was 1.07. That is, males displayed a somewhat larger variance, but the *VR* was not far from 1.0 or equal variances. In Study 2, the average variance ratio was 1.09, again not far from 1.0. In addition, the NELS:88 data (Table 3) show several VR's that are < 1.0, indicating that greater male variability is not ubiquitous. Variance ratios less than 1.0 have also been found in some national and international data sets (Hyde et al, 2008; Hyde & Mertz, 2009).

Overall, to put these findings in a broader context, gender can be conceptualized as one of many predictors of mathematics performance. Other factors include socioeconomic status (SES), parents' education, and the quality of schooling. Melhuish and colleagues (2008) compared the effect sizes of 9 predictors of children's mathematics performance at age 10: birth weight, gender, SES, mother's education, father's education, family income, quality of the home learning environment, preschool effectiveness, and elementary school effectiveness. The striking finding was that gender was the weakest of these 9 predictors, i.e., it had the smallest effect size. Mother's education, quality of the home learning environment, and elementary school effectiveness were far stronger predictors. Our findings are consistent with those of Melhuish and colleagues; gender is not a strong predictor of mathematics performance.

Implications

Overall, the results of these two studies provide strong evidence of gender similarities in mathematics performance. The heterogeneity of the findings suggests that there are moderator variables that might clarify the pattern of effect sizes. Detecting consistent moderators of gender differences would be strengthened by measures that tap the full range of mathematical reasoning, including items that require sustained reasoning about complex problems. The existence and magnitude of gender differences in mathematics performance varies as a function of many factors, including nation, ethnicity, and age.

These findings have several policy implications. First, these findings call into question current trends toward single-sex math classrooms. Advocates of single-sex education base their argument in part on the assumption that girls lag behind boys in mathematics performance and need to be in a protected, all-girls environment to be able to learn math (e.g., Streitmatter, 1999). The data, however, show that girls are performing as well as boys in mathematics, based on 242 separate studies (Study 1) and 4 large, well-sampled national U. S. data sets (Study 2). The great majority of these girls and boys did their learning in coeducational classrooms. Thus, the argument that girls' mathematics performance suffers in gender-integrated classrooms simply is not supported by the data. If we wish to improve students' mathematics performance, we would do better to focus not on gender, but on factors that have larger effects, such as the quality and implementation of the curriculum (Tarr et al., 2008) as well as the quality of the elementary school and the quality of the home learning environment (Melhuish et al., 2008).

Second, the dearth of Level 3 or 4 items in assessments has a serious consequence. Given the importance of mathematics tests for school evaluation under the No Child Left Behind legislation, it is common for teachers to teach to the test (Au, 2007). If the test fails to emphasize the skills that citizens need, American students are disadvantaged. In addition, without evidence concerning student progress on these important forms of mathematical reasoning, teachers, administrators, and policy makers cannot determine which curriculum materials or teaching strategies contribute to mathematical proficiency. Finally, tests that fail to emphasize complex problem solving or sustained reasoning communicate an inaccurate picture of mathematics to students.

These findings also have implications for dispelling stereotypes. Overall, it is clear that, in the U.S. and some other nations, girls have reached parity with boys in mathematics performance. It is crucial that this information be made widely known, to counteract stereotypes about female math inferiority held by gatekeepers such as parents and teachers, and by students themselves.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was funded by the National Science Foundation (REC 0635444 to Hyde) and the Eunice Kennedy Schriver National Institute of Child Health And Human Development (T32 HD049302 to Lindberg). The content of this paper is solely the responsibility of the authors and does not necessarily represent the official views of the funding organizations.

References

Ambady N, Shih M, Kim A, Pittinsky TL. Stereotype susceptibility in children: Effects of identity activation on quantitative performance. Psychological Science 2001;12:385–390. [PubMed: 11554671]

Anastasi, A. Differential psychology. 3rd ed.. Macmillan; New York: 1958.

Arms, E. Gender equity in coeducational and single-sex environments. In: Klein, S., editor. Handbook for achieving equity through education. Lawrence Erlbaum Associates; Mahwah, NJ: 2007. p. 171-190.

Au W. High-stakes testing and curricular control: A qualitative metasynthesis. Educational Researcher 2007;36:258–267.

Bandura, A. Social foundations of thought and action: A social cognitive theory. Prentice-Hall; Englewood Cliffs, NJ: 1986.

Bandura, A. Self-efficacy: The exercise of control. Freeman; New York: 1997.

Ben-Zeev T, Fein S, Inzlicht M. Arousal and stereotype threat. Journal of Experimental Social Psychology 2005;41:174–181.

Bouchey HA, Harter S. Reflected appraisals, academic self-perceptions, and math/science performance during early adolescence. Journal of Educational Psychology 2005;97:673–686.

Bussey K, Bandura A. Social cognitive theory of gender development and differentiation. Psychological Review 1999;106:676–713. [PubMed: 10560326]

Cadinu M, Maass A, Rosabianca A, Kiesner J. Why do women underperform under stereotype threat? Evidence for the role of negative thinking. Psychological Science 2005;16:572–578. [PubMed: 16008792]

Cohen, J. Statistical power analysis for the behavioral sciences. Erlbaum; Hillsdale, NJ: 1988.

Dwyer, CA.; Johnson, LM. Grades, accomplishments and correlates. In: Willingham, WA.; Cole, NS., editors. Gender and fair assessment. Erlbaum; Mahwah, NJ: 1997. p. 127-156.

Eccles JS. Understanding women's educational and occupational choices: Applying the Eccles et al. model of achievement-related choices. Psychology of Women Quarterly 1994;18:585–610.

Else-Quest NM, Hyde JS, Linn MC. Cross-national patterns of gender differences in mathematics and gender equity: A meta-analysis. Psychological Bulletin 2010;136:103–127. [PubMed: 20063928]

Feingold A. Sex differences in variability in intellectual abilities: A new look at an old controversy. Review of Educational Research 1992a;52:61–84.

Feingold A. The additive effects of differences in central tendency and variability are important in comparisons between groups. American Psychologist 1995;50:5–13.

Fennema E, Sherman J. Sex-related differences in mathematics achievement, spatial visualization, and affective factors. American Educational Research Journal 1977;14:51–71.

Fredricks JA, Eccles JS. Children's competence and value beliefs from childhood through adolescence: Growth trajectories in two male-sex-typed domains. Developmental Psychology 2002;38:519–533. [PubMed: 12090482]

Frome PM, Eccles JS. Parents' influence on children's achievement-related perceptions. Journal of Personality and Social Psychology 1998;74:435–452. [PubMed: 9491586]

Furnham A, Reeves E, Budhani S. Parents think their sons are brighter than their daughters: Sex differences in parental self-estimations and estimations of their children's multiple intelligences. Journal of Genetic Psychology 2002;163:24–39. [PubMed: 11952262]

Guiso L, Monte F, Sapienza P, Zingales L. Culture, gender, and math. Science 2008;320:1164–1165. [PubMed: 18511674]

Hedges LV. Distribution theory for Glass's estimator of effect size and related estimators. Journal of Educational Statistics 1981;6:107–128.

Hedges LV, Friedman L. Gender differences in variability in intellectual abilities: A reanalysis of Feingold's results. Review of Educational Research 1993;63:94–105.

Hedges LV, Nowell A. Sex differences in mental test scores, variability, and numbers of high-scoring individuals. Science 1995;269:41–45. [PubMed: 7604277]

Hedges LV, Vevea JL. Fixed- and random-effects models in meta-analysis. Psychological Methods 1998;3:486–504.

Helwig R, Anderson L, Tindal G. Influence of elementary student gender on teachers' perceptions of mathematics achievement. Journal of Educational Research 2001;95:93–102.

Hyde JS, Fennema E, Lamon S. Gender differences in mathematics performance: A meta-analysis. Psychological Bulletin 1990a;107:139–155. [PubMed: 2138794]

Hyde JS, Fennema E, Ryan M, Frost LA, Hopp C. Gender comparisons of mathematics attitudes and affect. Psychology of Women Quarterly 1990b;14:299–324.

Hyde JS, Lindberg SM, Linn MC, Ellis AB, Williams CC. Gender similarities characterize math performance. Science 2008;321:494–495. [PubMed: 18653867]

Hyde JS, Mertz JE. Gender, culture, and mathematics performance. Proceeding of the National Academy of Sciences 2009;106:8801–8807.

Inzlicht M, Ben-Zeev T. A threatening intellectual environment: Why females are susceptible to experiencing problem-solving deficits in the presence of males. Psychological Science 2000;11:365–371. [PubMed: 11228906]

Jacobs, J.; Davis-Kean, P.; Bleeker, M.; Eccles, J.; Malanchuk, O. “I can, but I don't wan to”: The impact of parents, interests, and activities on gender differences in math. In: Gallagher, A.; Kaufman, J., editors. Gender differences in mathematics: An integrative psychological approach. Cambridge University Press; New York: 2005. p. 73-98.

Johns M, Schmader T, Martens A. Knowing is half the battle: Teaching stereotype threat as a means of improving women's math performance. Psychological Science 2005;16:175–179. [PubMed: 15733195]

Katzman S, Alliger GM. Averaging untransformed variance ratios can be misleading: A comment on Feingold. Review of Educational Research 1992;62:427–428.

Keller C. Effect of teachers' stereotyping on students' stereotyping of mathematics as a male domain. The Journal of Social Psychology 2001;14:165–173. [PubMed: 11372563]

Kenney-Benson G, Pomerantz E, Ryan A, Patrick H. Sex differences in math performance: The role of children's approach to schoolwork. Developmental Psychology 2006;42:11–26. [PubMed: 16420115]

Kiefer AK, Sekaquaptewa D. Implicit stereotypes, gender identification, and math-related outcomes: A prospective study of female college students. Psychological Science 2007;18:13–18. [PubMed: 17362371]

Kimball MM. A new perspective on women's math achievement. Psychological Bulletin 1989;105:198–214.

Li Q. Teachers' beliefs and gender differences in mathematics: A review. Educational Research 1999;41:63–76.

Lindberg SM, Hyde JS, Hirsch LM. Gender and mother-child interactions during mathematics homework. Merrill-Palmer Quarterly 2008;54:232–255.

Lipsey, MW.; Wilson, DB. Practical meta-analysis. Sage; Thousand Oaks, CA: 2001.

Longitudinal Study of American Youth (n.d.). Retrieved January 29, 2009, from http://lsay.msu.edu/

Markwardt, FC, Jr.. Peabody individual achievement test-revised. American Guidance Service; Circle Pines, MN: 1998.

Meece JL, Eccles-Parsons J, et al. Sex differences in math achievement: Toward a model of academic choice. Psychological Bulletin 1982;91:324–348.

Melhuish EC, Sylva K, Sammons P, Siraj-Blatchford I, Taggart B, Phan MB, Malin A. The early years: Preschool influences on mathematics achievement. Science 2008;321:1161–1162. [PubMed: 18755959]

Mislevy RJ, Johnson EG, Muraki E. Scaling Procedures in NAEP. Journal of Educational Statistics 1992;17:131–154.

National Center for Education Research. (n.d.a.). National Educational Longitudinal Study of 1988 (NELS 88). Retrieved January 29, 2009, from http://nces.ed.gov/surveys/NELS88/

National Center for Education Research. (n.d.b.). The Nations Report Card: Mathematics. Retrieved January 29, 2009, from http://nces.ed.gov/nationsreportcard/mathematics/

Nosek BA, Banaji MR, Greenwald AG. Math = male, me = female, therefore math? me. Journal of Personality and Social Psychology 2002;83:44–59. [PubMed: 12088131]

Nosek BA, Smyth FL, Sriram N, Lindner NM, Devos T, Ayala A, Bar-Anan Y, Bergh R, Cai H, Gonsalkorale K, Kesebir S, Maliszewski N, Neto F, Olli E, Park J, Schnabel K, Shiomura K, Tulbure B, Wiers RW, Somogyi M, Akrami N, Ekehammar B, Vianello M, Banaji MR, Greenwald AG. National differences in gender-science stereotypes predict national sex differences in science and math achievement. Proceeding of the National Academy of Sciences 2009;106:10593–10597.

NSF. Women, minorities, and persons with disabilities in science and engineering. 2008a. www.nsf.gov/statistics/wmpd. Retrieved May 26, 2008

NSF. Thirty-three years of women in S&E faculty positions. 2008b. http://www.nsf.gov/statistics/infbrief/nsf08308/nsf08308.pdf. Retrieved November 2, 2009

NSF. Science and engineering indicators 2008. 2008c. www.nsf.gov/statistics/seind08. Retrieved May 29, 2008

Penner AJ. Gender differences in extreme mathematical achievement: An international perspective on biological and social factors. American Journal of Sociology 2008;114:S138–S170.

Quinn DN, Spencer SJ. The interference of stereotype threat with women's generation of mathematical problem-solving strategies. Journal of Social Issues 2001;57:55–72.

Shields SA. The variability hypothesis: The history of a biological model of sex differences in intelligence. Signs: Journal of Women in Culture and Society 1982;7:769–797.

Spencer SJ, Steele CM, Quinn DM. Stereotype threat and women's math performance. Journal of Experimental Social Psychology 1999;35:4–28.

Steele CM. A threat in the air: How stereotypes shape intellectual identity and performance. American Psychologist 1997;52:613–629. [PubMed: 9174398]

Steele CM, Aronson J. Stereotype threat and the intellectual test performance of African Americans. Journal of Personality and Social Psychology 1995;69:797–811. [PubMed: 7473032]

Steele J. Children's gender stereotypes about math: The role of stereotype stratification. Journal of Applied Social Psychology 2003;33:2587–2606.

Streitmatter, JL. For girls only: Making a case for single-sex schooling. State University of New York Press; Albany, NY: 1999.

Su R, Rounds J, Armstrong PI. Men and things, women and people: A meta-analysis of sex differences in interests. Psychological Bulletin 2009;135

Tarr JE, Reys RE, Reys BJ, Chávez Ó, Shih J, Osterlind SJ. The impact of middle-grades mathematics curricula and the classroom learning environment on student achievement. Journal for Research in Mathematics Education 2008;39:247–280.

Tiedemann J. Parents' gender stereotypes and teachers' beliefs as predictors of children's concept of their mathematical ability in elementary school. Journal of Educational Psychology 2000;92:144–151.

United States Bureau of Labor Statistics. (n.d.). National Longitudinal Surveys: The NLSY 97. Retrieved January 29, 2009, from http://www.bls.gov/nls/nlsy97.htm

Watt HMG. Development of adolescents' self-perceptions, values, and task perceptions according to gender and domain in 7th- through 11th-grade Australian students. Child Development 2004;75:1556–1574. [PubMed: 15369531]

Webb, NL. Alignment of science and mathematics standards and assessments in four states. Council of Chief State School Officers; Washington, DC: 1999. Research Monograph No. 18

Wilson, DB. Meta-analysis macros for SAS, SPSS, and Stata. 2005. Retrieved July 20, 2006, from http://mason.gmu.edu/~dwilsonb/ma.html

References Included in the Study 1 Meta-Analysis

Abdel-Khalek AM, Lynn R. Sex differences on the standard progressive matrices and in educational attainment in Kuwait. Personality and Individual Differences 2006;40:175–182.

Abedi J, Lord C. The language factor in mathematics tests. Applied Measurement in Education 2001;14:219–234.

Akerman BA. Twins at puberty: A follow-up study of 32 twin pairs. Psychology: The Journal of the Hellenic Psychological Society 2003;10:228–236.

Alkhateeb HM. Gender differences in mathematics achievement among high school students in the united arab emirates, 1991–2000. School Science and Mathematics 2001;101:5–9.

Alkhateeb HM. A preliminary study of achievement, attitudes toward success in mathematics, and mathematics anxiety with technology-based instruction in brief calculus. Psychological Reports 2002;90:47–57. [PubMed: 11899011]

Alkhateeb HM, Jumaa M. Cooperative learning and algebra performance of eighth grade students in united arab emirates. Psychological Reports 2002;90:91–100. [PubMed: 11899018]

Anderman EM, Midgley C. Changes in achievement goal orientations, perceived academic competence, and grades across the transition to middle-level schools. Contemporary Educational Psychology 1997;22:269–298. [PubMed: 9237829]

Arigbabu AA, Mji A. Is gender a factor in mathematics performance among Nigerian preservice teachers? Sex Roles 2004;51:749–753.

Atkins M, Rohrbeck CA. Gender effects in self-management training: Individual cooperative interventions. Psychology in the Schools 1993;30:362–368.

Aunio P, Hautamäki J, Heiskari P, Van Luit JEH. The early numeracy test in Finnish: Children's norms. Scandinavian Journal of Psychology 2006;47:369–378. [PubMed: 16987206]

Austin JT, Hanisch KA. Occupational attainment as a function of abilities and interests: A longitudinal analysis using project TALENT data. Journal of Applied Psychology 1990;75:77–86. [PubMed: 2307634]

Badian NA. Persistent arithmetic, reading, or arithmetic and reading disability. Annals of Dyslexia 1999;49:45–70.

Bandalos DL, Yates K, ThorndikeChrist T. Effects of math self-concept, perceived self-efficacy, and attributions for failure and success on test anxiety. Journal of Educational Psychology 1995;87:611–623.

Battista MT. Spatial visualization and gender differences in high school geometry. Journal for Research in Mathematics Education 1990;21:47–60.

Baumert J, Demmrich A. Test motivation in the assessment of student skills: The effects of incentives on motivation and performance. European Journal of Psychology of Education 2001;16:441–462.

Bell SM, Mccallum RS, Bryles J, Driesler K, Mcdonald J, Park SH, Williams A. Attributions for academic-success and failure : An individual difference investigation of academic-achievement and gender. Journal of Psychoeducational Assessment 1994;12:4–13.

Bempechat J, Graham SE, Jimenez NV. The socialization of achievement in poor and minority students - A comparative study. Journal of Cross-Cultural Psychology 1999;30:139–158.

Bennett RE, Morley M, Quardt D, Rock DA. Graphical modeling: A new response type for measuring the qualitative component of mathematical reasoning. Applied Measurement in Education 2000;13:303–322.

Benson J, Bandalos D, Hutchinson S. Modeling test anxiety among men and women. Anxiety, Stress, and Coping 1994;7:131–148.

Bibby PA, Lamb SJ, Leyden G, Wood D. Season of birth and gender effects in children attending moderate learning difficulty schools. British Journal of Educational Psychology 1996;66:159–168. [PubMed: 8652356]

Bielinski J, Davison ML. Gender differences by item difficulty interactions in multiple-choice mathematics items. American Educational Research Journal 1998;35:455–476.

Birenbaum M, Gutvirtz Y. The relationship between test anxiety and seriousness of errors in algebra. Journal of Psychoeducational Assessment 1993;11:12–19.

Birenbaum M, Nasser F. Ethnic and gender differences in mathematics achievement and in dispositions towards the study of mathematics. Learning and Instruction 2006;16:26–40.

Birenbaum M, et al. Stimulus features and sex differences in mental rotation test performance. Intelligence 1994;19:51.

Bliwise NG. Web-based tutorials for teaching introductory statistics. Journal of Educational Computing Research 2005;33:309.

Bolger N, Kellaghan T. Method of measurement and gender differences in scholastic achievement. Journal of Educational Measurement 1990;27:165–174.

Borg MG. Sex and age differences in the scholastic attainment of grammar school children in the first three years of secondary schooling: A longitudinal study. Research in Education 1996;(56):1–20.

Borg MG, Falzon JM. Birth date and sex effects on the scholastic attainment of primary schoolchildren: A cross-sectional study. British Educational Research Journal 1995;21:61.

Borg MG, Falzon JM, Sammut A. Age and sex differences in performance in an 11-plus selective examination. Educational Psychology 1995;15:433–443.

Bornholt LJ, Goodnow JJ, Cooney GH. Influences of gender stereotypes on adolescents perception of their own achievement. American Educational Research Journal 1994;31:675–692.

Brennan RT, Kim J, Wenz-Gross M, Siperstein GN. The relative equitability of high-stakes testing versus teacher-assigned grades: An analysis of the Massachusetts comprehensive assessment system (MCAS). Harvard Educational Review 2001;71:173–216.

Bridgeman B, Harvey A, Braswell J. Effects of calculator use on scores on a test of mathematical reasoning. Journal of Educational Measurement 1995;32:323–340.

Bridgeman B, Wendler C. Gender differences in predictors of college mathematics performance and in college mathematics course grades. Journal of Educational Psychology 1991;83:275–284.

Bruce CK, Lawrenz FP. Actual and teacher perceptions of the abilities of mathematical high school chemistry students in Minnesota. School Science and Mathematics 1991;91:1–5.

Bull R, Johnston RS. Children's arithmetical difficulties: Contributions from processing speed, item identification, and short-term memory. Journal of Experimental Child Psychology 1997;65:1–24. [PubMed: 9126630]

Busato VV, Dam G. T. M. t. Eeden P. v. d. Gender-related effects of co-operative learning in a mathematics curriculum for 12–16-year-olds. Journal of Curriculum Studies 1995;27:667–686.

Byrnes JP, Hong L, Xing S. Gender differences on the math subtest of the scholastic aptitude test may be culture-specific. Educational Studies in Mathematics 1997;34:49–66.

Byrnes JP, Takahira S. Explaining gender differences on SAT-math items. Developmental Psychology 1993;29:805–810.

Byrnes JP, Takahira S. Why some students perform well and others perform poorly on SAT math items. Contemporary Educational Psychology 1994;19:63–78.

Campbell E, Schellinger T, Beer J. Relationships among the ready or not parental checklist for school readiness, the Brigance Kindergarten and 1-grade screen, and SRA scores. Perceptual and Motor Skills 1991;73:859–862.

Cankoy O, Tut MA. High-stakes testing and mathematics performance of fourth graders in north Cyprus. Journal of Educational Research 2005;98:234–243.

Cardelle-Elawar M. Effects of feedback tailored to bilingual students' mathematics needs on verbal problem solving. Elementary School Journal 1990;91:165–175.

Chan DW. Assessing giftedness of Chinese secondary students in Hong Kong: A multiple intelligences perspective. High Ability Studies 2001;12:215–234.

Chen PP. Exploring the accuracy and predictability of the self-efficacy beliefs of seventh-grade mathematics students. Learning and Individual Differences 2002;14:77–90.

Cherian VI. Gender, socioeconomic-status, and mathematics achievement by Xhosa children. Psychological Reports 1993;73:771–778.

Cherian VI, Cherian L. Relationship of divorce, gender, socioeconomic-status and unhappiness to mathematics achievement of children. Journal of Family Welfare 1995;41:30–37.

Chipman SF, Marshall SP, Scott PA. Content effects on word problem performance: A possible source of test bias? American Educational Research Journal 1991;28:897–915.

Clariana RB, Schultz CW. Gender by content achievement differences in computer-based instruction. The Journal of Computers in Mathematics and Science Teaching 1993;12:277–288.

Collaer ML, Hill EM. Large sex difference in adolescents on a timed line judgment task: Attentional contributors and task relationship to mathematics. Perception 2006;35:561–572. [PubMed: 16700296]

Connors MA. Achievement and gender in computer-integrated calculus. Journal of Women and Minorities in Science and Engineering 1995;2:113.

Crosser SL. Summer birth date children: Kindergarten entrance age and academic achievement. Journal of Educational Research 1991;84:140.

Davies J, Brember I. Boys outperforming girls: An 8-year cross-sectional study of attainment and self-esteem in year 6. Educational Psychology 1999;19:5–16.

Davis H, Carr M. Gender differences in mathematics strategy use - the influence of temperament. Learning and Individual Differences 2001;13:83–95.

Davis-Dorsey J, Ross SM, Morrison GR. The role of rewording and context personalization in the solving of mathematical word problems. Journal of Educational Psychology 1991;83:61–68.

De Brauwer J, Verguts T, Fias W. The representation of multiplication facts: Developmental changes in the problem size, five, and tie effects. Journal of Experimental Child Psychology 2006;94:43–56. [PubMed: 16376370]

De Lisle J, Smith P, Jules V. Which males or females are most at risk and on what? an analysis of gender differentials within the primary school system of Trinidad and Tobago. Educational Studies 2005;31:393–418.

DeMars CE. Gender differences in mathematics and science on a high school proficiency exam: The role of response format. Applied Measurement in Education 1998;11:279–299.

Diamante T. Unitarian validation of a mathematical problem-solving exercise for sales occupations. Journal of Business and Psychology 1993;7:383–401.

Dickhäuser O, Meyer W. Gender differences in young children's math ability attributions. Psychology Science 2006;48:3–16.

Duffy J, Gunther G, Walters L. Gender and mathematical problem solving. Sex Roles 1997;37:477–494.

Eid GK, Koushki PA. Secondary education programs in Kuwait: An evaluation study. Education (Chula Vista, Calif.) 2005;126:181–200.

El Hassan K. Gender issues in achievement in Lebanon. Social Behavior and Personality 2001;29:113–123.

Elliott JC. Affect and mathematics achievement of nontraditional college students. Journal for Research in Mathematics Education 1990;21:160–165.

Entwisle DR, Alexander KL, Olson LS. The gender-gap in math – Its possible origins in neighborhood effects. American Sociological Review 1994;59:822–838.

Evans EM, Schweingruber H, Stevenson HW. Gender differences in interest and knowledge acquisition: The United States, Taiwan, and Japan. Sex Roles 2002;47:153–167.

Ewers CA, Wood NL. Sex and ability differences in childrens math self-efficacy and accuracy. Learning and Individual Differences 1993;5:259–267.

Feldman R, Guttfreund D, Yerushalmi H. Parental care and intrusiveness as predictors of the abilities-achievement gap in adolescence. Journal of Child Psychology and Psychiatry and Allied Disciplines 1998;39:721–730.

Fennema E, Carpenter TP, Jacobs VR. A longitudinal study of gender differences in young children's mathematical thinking. Educational Researcher 1998;27:6–11.

Fink B, Brookes H, Neave N, Manning JT, Geary DC. Second to fourth digit ratio and numerical competence in children. Brain and Cognition 2006;61:211–218. [PubMed: 16466837]

Fischbein S. Biosocial influences on sex differences for ability and achievement test results as well as marks at school. Intelligence 1990;14:127–139.

Fuller B, Hua H, Snyder C Jr. When girls learn more than boys: The influence of time in school and pedagogy in Botswana. Comparative Education Review 1994;38:347–376. [PubMed: 12318867]

Gallagher AM, De Lisi R, Holst PC, McGillicuddy-De Lisi AV, Morely M, Cahalan C. Gender differences in advanced mathematical problem solving. Journal of Experimental Child Psychology 2000;75:165–190. [PubMed: 10666324]

Galler JR, Ramsey FC, Harrison RH, Taylor J, Cumberbatch G, Forde V. Postpartum maternal moods and infant size predict performance on a national high school entrance examination. Journal of Child Psychology and Psychiatry 2004;45:1064–1075. [PubMed: 15257663]

Garner M, Engelhard G. Gender differences in performance on multiple-choice and constructed response mathematics items. Applied Measurement in Education 1999;12:29–51.

Geary DC, Salthouse TA, Chen GP, Fan L. Are east Asian versus American differences in arithmetical ability a recent phenomenon? Developmental Psychology 1996;32:254–262.

Geary DC, Saults SJ, Liu F, Hoard MK. Sex differences in spatial cognition, computational fluency, and arithmetical reasoning. Journal of Experimental Child Psychology 2000;77:337–353. [PubMed: 11063633]

Gilbert WS. Bridging the gap between high school and college. Journal of American Indian Education 2000;39:36.

Glutting JJ, Oh HJ, Ward T, Ward S. Possible criterion-related bias of the WISC-III with a referral sample. Journal of Psychoeducational Assessment 2000;18:17–26.

Gottesman RL, Bennett RE, Nathan RG, Kelly MS. Inner-city adults with severe reading difficulties: A closer look. Journal of Learning Disabilities 1996;29:589–597. [PubMed: 8942303]

Gouchie C, Kimura D. The relationship between testosterone levels and cognitive ability patterns. Psychoneuroendocrinology 1991;16:323–334. [PubMed: 1745699]

Gresky DM, Eyck LLT, Lord CG, McIntyre RB. Effects of salient multiple identities on women's performance under mathematics stereotype threat. Sex Roles 2005;53:703–716.

Grewal AS. Sex differences in algebra by senior secondary school students in Transkei, South Africa. Psychological Reports 1998;83:1266–1266.

Halat E. Sex-related differences in the acquisition of the van hiele levels and motivation in learning geometry. Asia Pacific Education Review 2006;7:173–183.

Hall CW, Davis NB, Bolen LM, Chia R. Gender and racial differences in mathematical performance. Journal of Social Psychology 1999;139:677–689. [PubMed: 10646303]

Hay I, Ashman AF, van Kraayenoord CE. The influence of gender, academic achievement and non-school factors upon pre-adolescent self-concept. Educational Psychology 1998;18:461–470.

Hein J, Bzufka MW, Neumarker KJ. The specific disorder of arithmetic skills. prevalence studies in a rural and an urban population sample and their cliniconeuropsychological validation. European Child & Adolescent Psychiatry 2000;9:87–101.

Held JD, Alderton DL, Foley PP, Segall DO. Arithmetic reasoning gender differences – Explanations found in the armed services vocational aptitude battery (ASVAB). Learning and Individual Differences 1993;5:171–186.

Helwig R, Anderson L, Tindal G. Influence of elementary student gender on teachers' perceptions of mathematics achievement. Journal of Educational Research 2001;95:93–102.

Ho CH, Eastman C, Catrambone R. An investigation of 2D and 3D spatial and mathematical abilities. Design Studies 2006;27:505–524.

Ho HZ, Senturk D, Lam AG, Zimmer JM, Hong S, Okamoto Y, et al. The affective and cognitive dimensions of math anxiety: A cross-national study. Journal for Research in Mathematics Education 2000;31:362–379.

Hong E, Aqui Y. Cognitive and motivational characteristics of adolescents gifted in mathematics: Comparisons among students with different types of giftedness. Gifted Child Quarterly 2004;48:191–201.

Hosenfeld I, Koller O, Baumert J. Why sex differences in mathematics achievement disappear in German secondary schools: A reanalysis of the German TIMSS-data. Studies in Educational Evaluation 1999;25:143–161.

Huang J. An investigation of gender differences in cognitive abilities among Chinese high school students. Personality and Individual Differences 1993;15:717–719.

Iben MF. Attitudes and mathematics. Comparative Education 1991;27:135–151.

Isiksal M, Askar P. The effect of spreadsheet and dynamic geometry software on the achievement and self-efficacy of 7th-grade students. Educational Research 2005;47:333–350.

Jinabhai CC, Taylor M, Rangongo MF, Mkhize NJ, Anderson S, Pillay BJ, et al. Investigating the mental abilities of rural Zulu primary school children in South Africa. Ethnicity & Health 2004;9:17–36. [PubMed: 15203463]

Jones ML, Rowsey RE. The effects of immediate achievement and retention of middle school students involved in a metric unit designed to promote the development of estimating skills. Journal of Research in Science Teaching 1990;27:901–913.

Kahn M. A class act - mathematics as filter of equity in South Africa's schools. Perspectives in Education 2005;23:139–148.

Kaiser J. The role of family configuration, income, and gender in the academic achievement of young self-care children. Early Child Development and Care 1994;97:91–105.

Kass RG, Fish JM. Positive reframing and the test performance of test anxious children. Psychology in the Schools 1991;28:43–52.

Kee DW, Gottfried A, Bathurst K. Consistency of hand preference: Predictions to intelligence and school achievement. Brain and Cognition 1991;16:1–10. [PubMed: 1854465]

Keller J. Blatant stereotype threat and women's math performance: Self-handicapping as a strategic means to cope with obtrusive negative performance expectations. Sex Roles 2002;47:193–198.

Kelly-Vance L, Caster A, Ruane A. Non-graded versus graded elementary schools: An analysis of achievement and social skills. Alberta Journal of Educational Research 2000;46:372–390.

Kenney-Benson GA, Pomerantz EM, Ryan AM, Patrick H. Sex differences in math performance: The role of children's approach to schoolwork. Developmental Psychology 2006;42:11–26. [PubMed: 16420115]

Kiger DM. The effect of group test-taking environment on standardized achievement test scores: A randomized block field trial. American Secondary Education 2005;33:63–72.

Kimura D. Body asymmetry and intellectual pattern. Personality and Individual Differences 1994;17:53–60.

Kloosterman P. Beliefs and achievement in seventh-grade mathematics. Focus on Learning Problems in Mathematics 1991;v13:3.

Koizumi R. The relationship between perceived attainment and optimism, and academic achievement and motivation. Japanese Psychological Research 1992;34:1–9.

Kontrová J, Palkovicová E, Árochová O. Load and stress in the teaching process. Studia Psychologica 1991;33:129–137.

Kumar S, Harizuka S. Cooperative learning-based approach and development of learning awareness and achievement in mathematics in elementary school. Psychological Reports 1998;82:587–591.

Kwok DC, Lytton H. Perceptions of mathematics ability versus actual mathematics performance: Canadian and Hong Kong Chinese children. British Journal of Educational Psychology 1996;66:209–222. [PubMed: 8652357]

Lachance JA, Mazzocco MMM. A longitudinal analysis of sex differences in math and spatial skills in primary school age children. Learning and Individual Differences 2006;16:195–216. [PubMed: 20463851]

Lagace DC, Kutcher SP, Robertson HA. Mathematics deficits in adolescents with bipolar I disorder. American Journal of Psychiatry 2003;160:100–104. [PubMed: 12505807]

Lakes KD, Hoyt WI. Promoting self-regulation through school-based martial arts training. Journal of Applied Developmental Psychology 2004;25:283–302.

Landgren M, Kjellman B, Gillberg C. “A school for all kinds of minds” - The impact of neuropsychiatric disorders, gender and ethnicity on school-related tasks administered to 9–10-year-old children. European Child & Adolescent Psychiatry 2003;12:162–171. [PubMed: 14505066]

Lau S, Leung K. Relations with parents and school and Chinese adolescents' self-concept, delinquency, and academic performance. British Journal of Educational Psychology 1992;62:193–202. [PubMed: 1637685]

LeFevre J, Kulak AG, Heymans SL. Factors influencing the selection of university majors varying in mathematical content. Canadian Journal of Behavioural Science 1992;24:276–289.

Leonard J. How group composition influenced the achievement of sixth-grade mathematics students. Mathematical Thinking and Learning 2001;3:175–200.

Lesko AC, Corpus JH. Discounting the difficult: How high math-identified women respond to stereotype threat. Sex Roles 2006;54:113–125.

Lim TK. Gender-related differences in intelligence: Application of confirmatory factor analysis. Intelligence 1994;19:179.

Lindblad F, Lindahl M, Theorell T, von Scheele B. Physiological stress reactions in 6th and 9th graders during test performance. Stress and Health 2006;22:189–195.

Lindsay G, Desforges M. The use of the infant Index/Baseline-PLUS as a baseline assessment measure of literacy. Journal of Research in Reading 1999;22:55–66.

Lloyd JEV, Walsh J, Yailagh MS. Sex differences in performance attributions, self-efficacy, and achievement in mathematics: If I'm so smart, why don't I know it? Canadian Journal of Education 2005;28:384–408.

Lopez CL, Sullivan HJ. Effect of personalization of instructional context on the achievement and attitudes of Hispanic students. Educational Technology Research and Development 1992;40:5–13.

López CL, Sullivan HJ. Effects of personalized math instruction for Hispanic students. Contemporary Educational Psychology 1991;16:95–100.

Lopez-Sobaler AM, Ortega RM, Quintas ME, Navia B, Requejo AM. Relationship between habitual breakfast and intellectual performance (logical reasoning) in well-nourished schoolchildren of Madrid (Spain). European Journal of Clinical Nutrition 2003;57:49–53.

Low R, Over R. Gender differences in solution of algebraic word problems containing irrelevant information. Journal of Educational Psychology 1993;85:331–339.

Lowrie T, Kay R. Relationship between visual and non-visual solution methods and difficulty in elementary mathematics. Journal of Educational Research 2001;94:248–255.

Lubinski D, Humphreys LG. A broadly based analysis of mathematical giftedness. Intelligence 1990;14:327–355.

Lummis M, Stevenson HW. Gender differences in beliefs and achievement: A cross-cultural study. Developmental Psychology 1990;26:254–263.

Lyons JB, Schneider TR. The influence of emotional intelligence on performance. Personality and Individual Differences 2005;39:693–703.

Manger T, Eikeland OJ. Relationship between boys' and girls' nonverbal ability and mathematical achievement. School Psychology International 1996;17:71–80.

Manger T, Eikeland OJ. The effect of mathematics self-concept on girls' and boys' mathematical achievement. School Psychology International 1998;19:5–18.

Manger T, Eikeland OJ. The effects of spatial visualization and students' sex on mathematical achievement. British Journal of Psychology 1998;89:17–25. [PubMed: 9532723]

Manger T. Gender differences in mathematical achievement at the Norwegian elementary-school level. Scandinavian Journal of Educational Research 1995;39:257–269.

Manger T, Eikeland O. Gender differences in mathematical sub-skills. Research in Education 1998;59:59–68.

Manger T, Gjestad R. Gender differences in mathematical achievement related to the ratio of girls to boys in school classes. International Review of Education 1997;43:193.

Maqsud M. Effects of metacognitive skills and nonverbal ability on academic achievement of high school pupils. Educational Psychology 1997;17:387–397.

Maqsud M, Khalique CM. Relationships of some socio-personal factors to mathematics achievement of secondary school and university students in Bophuthatswana. Educational Studies in Mathematics 1991;22:377–390.

Maqsud M, Rouhani S. Relationships between socioeconomic status, locus of control, self-concept, and academic achievement of Batswana adolescents. Journal of Youth and Adolescence 1991;20:107–114.

Maree JG, Erasmus CP. Mathematics skills of tswana-speaking learners in the north west province of South Africa. International Journal of Adolescence and Youth 2006;13:71–97.

Matthews DJ. Diversity in domains of development: Research findings and their implications for gifted identification and programming. Roeper Review 1997;19:172.

Mboya MM. Self-concept of academic ability as a function of sex, age, and academic achievement among African adolescents. Perceptual and Motor Skills 1998;87:155–161. [PubMed: 9760641]

McCoy LP. Effect of demographic and personal variables on achievement in eighth-grade algebra. Journal of Educational Research 2005;98:131–135.

McKenzie B, Bull R, Gray C. The effects of phonological and visual-spatial interference on children's arithmetical performance. Educational and Child Psychology 2003;20:93–108.

Mclntyre RB, Lord CG, Gresky DM, Ten Eyck LL, Jay Frye GD, Bond CFJ. A social impact trend in the effects of role models on alleviating women's mathematics stereotype threat. Current Research in Social Psychology 2005;10

McNiece R, Jolliffe F. An investigation into regional differences in educational performance in the national child development study. Educational Research 1998;40:17–30.

Medina M Jr. Spanish achievement in a maintenance bilingual education program: Language proficiency, grade and gender comparisons. Bilingual Research Journal 1993;17:57.

Miller CJ, Crouch JG. Gender difference in problem-solving-expectancy and problem context. Journal of Psychology 1991;125:327–336.

Mills CJ, Ablard KE, Gustin WC. Academically talented students' achievement in a flexibly paced mathematics program. Journal for Research in Mathematics Education 1994;25:495–511.

Mills CJ, Ablard KE, Stumpf H. Gender differences in academically talented young students' mathematical reasoning: Patterns across age and sub-skills. Journal of Educational Psychology 1993;85:340–346.

Mohsin M, Nath SR, Chowdhury AMR. Influence of socioeconomic factors on basic competencies of children in Bangladesh. Journal of Biosocial Science 1996;28:15–24. [PubMed: 8690740]

Moodaley RR, Grobler AA, Lens W. Study orientation and causal attribution in mathematics achievement. South African Journal of Psychology 2006;36:634–655.

Morrison FJ, Griffith EM, Alberts DM. Nature-nurture in the classroom: Entrance age, school readiness, and learning in children. Developmental Psychology 1997;33:254–262. [PubMed: 9147834]

Murphy LO, Ross SM. Protagonist gender as a design variable in adapting mathematics story problems to learner interests. Educational Technology Research and Development 1990;38:27–37.

Mwamwenda TS. Sex differences in mathematics performance among african university students. Psychological Reports 2002;90:1101–1104. [PubMed: 12150391]

Narciss S, Huth K. Fostering achievement and motivation with bug-related tutoring feedback in a computer-based training for written subtraction. Learning and Instruction 2006;16:310–322.

Nasser F, Birenbaum M. Modeling mathematics achievement of Jewish and Arab eighth graders in Israel: The effects of learner-related variables. Educational Research and Evaluation 2005;11:277–302.

Nelson JR, Benner GJ, Lane K, Smith BW. Academic achievement of K-12 students with emotional and behavioral disorders. Exceptional Children 2004;71:59–73.

Nyangeni NP, Glencross MJ. Sex differences in mathematics achievement and attitude toward mathematics. Psychological Reports 1997;80:603–608.

Olszewski-Kubilius P, Turner D. Gender differences among elementary school-aged gifted students in achievement. Journal for the Education of the Gifted 2002;25:233–268.

Onatsu-Arvilommi T, Nurmi JE. The role of task-avoidant and task-focused behaviors in the development of reading and mathematical skills during the first school year: A cross-lagged longitudinal study. Journal of Educational Psychology 2000;92:478–491.

O'Neil HF, Abedi J, Miyoshi J. Monetary incentives for low-stakes tests. Educational Assessment 2005;10:185–208.

Ong W, Allison J, Haladyna TM. Student achievement of 3rd-graders in comparable single-age and multiage classrooms. Journal of Research in Childhood Education 2000;14(2):205–215.

Opyene-Eluk P, Opolot-Okurut C. Gender and school-type differences in mathematics achievement of senior three pupils in central Uganda: An exploratory study. International Journal of Mathematical Education in Science and Technology 1995;26:871–886.

Pajares F, Miller MD. Role of self-efficacy and self-concept beliefs in mathematical problem-solving – A path-analysis. Journal of Educational Psychology 1994;86:193–203.

Pajares F. Mathematics self-efficacy and mathematical problem solving: Implications of using different forms of assessment. Journal of Experimental Education 1997;65:213–228.

Pajares F. Self-efficacy beliefs and mathematical problem-solving of gifted students. Contemporary Educational Psychology 1996;21:325–344. [PubMed: 8979868]

Panchon A. The effects of white-noise and gender on mental task. Psychologia 1994;37:234–240.

Park HS. Computational mathematical abilities of African American girls. Journal of Black Studies 1999;30:204–215.

Park H, Bauer SC. Gender differences among top performing elementary school students in mathematical ability. Journal of Research and Development in Education 1998;31:133–141.

Pascarella ET, Bohr L, Nora A. Intercollegiate athletic participation and freshman-year cognitive outcomes. Journal of Higher Education 1995;66:369–387.

Pehkonen E. Learning results from the viewpoint of equity: Boys, girls and mathematics. Teaching Mathematics and its Applications 1997;16:58.

Pigg AE, Waliczek TM. Effects of a gardening program on the academic progress of third, fourth, and fifth grade math and science students. Horttechnology 2006;16:262–264.

Pomplun M. Gender differences for constructed-response mathematics items. Educational and Psychological Measurement 1999;59:597–614.

Pope GA, Wentzel C, Braden B. Relationships between gender and Alberta Achievement Test Scores during a four-year period. Alberta Journal of Educational Research 2006;52:4–15.

Quinn DM. The interference of stereotype threat with women's generation of mathematical problem-solving strategies. Journal of Social Issues 2001;57:55–71.

Rammstedt B. Self-estimated intelligence - gender differences, relationship to psychometric intelligence and moderating effects of level of education. European Psychologist 2002;7:275–284.

Randhawa BS. Validity of performance assessment in mathematics for early adolescents. Canadian Journal of Behavioural Science-Revue Canadienne Des Sciences Du Comportement 2001;33:14–24.

Randhawa BS. Self-efficacy in mathematics, attitudes, and achievement of boys and girls from restricted samples in two countries. Perceptual and Motor Skills 1994;79:1011–1018.

Randhawa BS. Understanding sex differences in the components of mathematics achievement. Psychological Reports 1993;73:435–444.

Reynolds AJ, Mehana M. Does preschool intervention affect childrens perceived competence. Journal of Applied Developmental Psychology 1995;16:211–230.

Reys RE, Reys B. Computational estimation performance and strategies used by fifth- and eighth-grade Japanese students. Journal for Research in Mathematics Education 1991;22:39–58.

Reys RE, Reys B. Mental computation performance and strategy use of japanese students in grades 2, 4, 6, and 8. Journal for Research in Mathematics Education 1995;26:304–326.

Robinson NM, Abbott RD. The structure of abilities in math-precocious young children: Gender similarities and differences. Journal of Educational Psychology 1996;88:341–352.

Rosselli M, Ardila A, Bateman JR. Neuropsychological test scores, academic performance., and developmental disorders in Spanish-speaking children. Developmental Neuropsychology 2001;20:355–373. [PubMed: 11827093]

Rouxel G. Cognitive-affective determinants of performance in mathematics and verbal domains - gender differences. Learning and Individual Differences 2001;12:287–310.

Rouxel G. Cognitive-affective determinants of performance in mathematics and verbal domains gender differences. Learning and Individual Differences 2000;12:287–310.

Rudnitsky A, Etheredge S, Freeman SJM. Learning to solve addition and subtraction word problems through a structure-plus-writing approach. Journal for Research in Mathematics Education 1995;26:467–486.

Ruthven K. The influence of graphic calculator use on translation from graphic to symbolic forms. Educational Studies in Mathematics 1990;21:431.

Saigal S, Lambert M, Russ C. Self-esteem of adolescents who were born prematurely. Pediatrics 2002;109:429–433. [PubMed: 11875137]

Salawu AA. Relationship between adolescents' perception of parents' behaviour and their academic achievement. IFE Psychologia: An International Journal 1993;1:153–165.

Salerno CA. The effect of time on computer-assisted instruction for at-risk students. Journal of Research on Computing in Education 1995;28:85–97.

Sappington J, Larsen C, Martin J. Sex-differences in math problem-solving as a function of gender-specific item content. Educational and Psychological Measurement 1991;51:1041–1048.

Sarouphim KM. Discover in middle school: Identifying gifted minority students. Journal of Secondary Gifted Education 2004;15:61–69.

Schellinger T. Correlations among special-educations students WISC - RIQS and SRA scores. Perceptual and Motor Skills 1991;73:1225–1226. [PubMed: 1805176]

Seegers G. Gender-related differences in self-referenced cognitions in relation to mathematics. Journal for Research in Mathematics Education 1996;27:215–240.

Sekaquaptewa D. Solo status, stereotype threat, and performance expectancies: Their effects on women's performance. Journal of Experimental Social Psychology 2003;39:68–74.

Shannon HD, Allen TW. The effectiveness of a REBT training program in increasing the performance of high school students in mathematics. Journal of Rational-Emotive & Cognitive Behavior Therapy 1998;16:197–209.

Sheehan KR, Gray MW. Sex bias in the SAT and the DTMS. Journal of General Psychology 1992;119:5–14.

Shibley IA Jr. Milakofsky L. College chemistry and Piaget: An analysis of gender difference, cognitive abilities, and achievement measures seventeen years apart. Journal of Chemical Education 2003;80:569–573.

Shymansky JA, Yore LD, Anderson JO. Impact of a school district's science reform effort on the achievement and attitudes of third- and fourth-grade students. Journal of Research in Science Teaching 2004;41:771–790.

Skaalvik EM, Rankin RJ. Math, verbal, and general academic self-concept: The internal/external frame of reference model and gender differences in self-concept structure. Journal of Educational Psychology 1990;82:546–554.

Skaalvik EM, Rankin RJ. Gender differences in mathematics and verbal achievement, self-perception and motivation. British Journal of Educational Psychology 1994;64:419–428.

Skaggs G, Lissitz RW. The consistency of detecting item bias across different test administrations: Implications of another failure. Journal of Educational Measurement 1992;29:227–242.

Slate JR, Jones CH, Turnbough R, Bauschlicher L. Gender differences in achievement scores on the metropolitan achievement test-6 and the stanford achievement test-8. Research in the Schools 1994;1:59–62.

Smees R, Sammons P, Thomas S, Mortimore P. Examining the effect of pupil background on primary and secondary pupils' attainment: Key findings from the improving school effectiveness project. Scottish Educational Review 2002;34:6.

Smith JL, White PH. Development of the domain identification measure: A tool for investigating stereotype threat effects. Educational and Psychological Measurement 2001;61:1040–1057.

Spencer SJ, Steele CM, Quinn DM. Stereotype threat and women's math performance. Journal of Experimental Social Psychology 1999;35:4–28.

Srivastava NC. Verbal test of intelligence as a predictor of success in science and mathematics. Psycho-Lingua 1993;23:65–70.

Stage FK, Kloosterman P. Gender, beliefs, and achievement in remedial college-level mathematics. Journal of Higher Education 1995;66:294–311.

Standing LG, Sproule RA, Leung A. Can business and economics students perform elementary arithmetic? Psychological Reports 2006;98:549–555. [PubMed: 16796113]

Stevenson HW, Chen C, Booth J. Influences of schooling and urbanural residence on gender differences in cognitive abilities and academic achievement. Sex Roles 1990;23:535–551.

Stricker LJ, Ward WC. Stereotype threat, inquiring about test takers' ethnicity and gender, and standardized test performance. Journal of Applied Social Psychology 2004;34:665–693.

Stumpf H, Haldimann M. Spatial ability and academic success of sixth grade students at international schools. School Psychology International 1997;18:245–259.

Subotnik RF, Strauss SM. Gender differences in classroom participation and achievement: An experiment involving advanced placement calculus classes. Journal of Secondary Gifted Education 1995;6:77.

Swiatek MA, Lupkowski-Shoplik A, O'Donoghue CC. Gender differences in above-level EXPLORE scores of gifted third through sixth graders. Journal of Educational Psychology 2000;92:718–723.

Tartre LA, Fennema E. Mathematics achievement and gender: A longitudinal study of selected cognitive and affective variables {grades 6–12}. Educational Studies in Mathematics 1995;28:199–217.

Taylor L. An integrated learning system and its effect on examination performance in mathematics. Computers & Education 1999;32:95–107.

Thompson GW, et al. Gender differences in an experimental program on arithmetic problem solving and computation. Midwestern Educational Researcher 1992;5:20.

Tiedemann J, Faber G. Preschoolers' maternal support and cognitive competencies as predictors of elementary achievement. Journal of Educational Research 1992;85:348–354.

Travis B, Lennon E. Spatial skills and computer-enhanced instruction in calculus. The Journal of Computers in Mathematics and Science Teaching 1997;16:467–475.

Tsui M, Rich L. The only child and educational opportunity for girls in urban china. Gender & Society 2002;16:74–92.

Undheim JO, Nordvik H, Gustafsson K, Undheim AM. Academic achievements of high-ability students in egalitarian education: A study of able 16-year-old students in Norway. Scandinavian Journal of Educational Research 1995;39:157–167.

Valanides NC. Formal reasoning and science teaching. School Science and Mathematics 1996;96:99–107.

VanDerHeyden AM, Broussard C, Cooley A. Further development of measures of early math performance for preschoolers. Journal of School Psychology 2006;44:533–553.

Vermeer HJ, Boekaerts M, Seegers G. Motivational and gender differences: Sixth-grade students' mathematical problem-solving behavior. Journal of Educational Psychology 2000;92:308–315.

Walsh M, Hickey C, Duffy J. Influence of item content and stereotype situation on gender differences in mathematical problem solving. Sex Roles 1999;41:219–240.

Wang N, Lane S. Detection of gender-related differential item functioning in a mathematics performance assessment. Applied Measurement in Education 1996;9:175–199.

Wangu RS, Thomas KJ. Attitude towards and achievement in mathematics among high school students of tribal town of Aizawl. Indian Journal of Psychometry & Education 1995;26:31–36.

Warwick DP, Jatoi H. Teacher gender and student achievement in Pakistan. Comparative Education Review 1994;38:377–399.

Watt HMG. Measuring attitudinal change in mathematics and english over the 1st year of junior high school: A multidimensional analysis. Journal of Experimental Education 2000;68:331–361.

Watt HMG. The role of motivation in gendered educational and occupational trajectories related to maths. Educational Research and Evaluation 2006;12:305–322.

Watt HMG, Bornholt LJ. Social categories and student perceptions in high school mathematics. Journal of Applied Social Psychology 2000;30:1492–1503.

Werdelin I. Sex differences in performance scores and patterns of development. Interdisciplinaria Revista De Psicología y Ciencias Afines 1996;13:35–65.

Williams JE, Montgomery D. Using frame of reference theory to understand the self-concept of academically able students. Journal for the Education of the Gifted 1995;18:400–409.

Witt EA, Dunbar SB, Hoover HD. A multivariate perspective on sex differences in achievement and later performance among adolescents. Applied Measurement in Education 1994;7:241–254.

Wu M, Greenan JP. The effects of a generalizable mathematics skills instructional intervention on the mathematics achievement of learners in secondary CTE programs. Journal of Industrial Teacher Education 2003;40:23–50.

Xu J, Farrell EW. Mathematics performance of shanghai high school students: A preliminary look at gender differences in another culture. School Science and Mathematics 1992;92:442–445.

Yadrick RM, Regian JW, RobertsonSchule L, Gomez GC. Interface, instructional approach, and domain learning with a mathematics problem-solving environment. Computers in Human Behavior 1996;12:527–548.

Zervas Y. Effect of a physical exercise session on verbal, visuospatial, and numerical ability. Perceptual and Motor Skills 1990;71:379–383.