# Gender Similarities Characterize Math Performance


**Janet Shibley Hyde, Sara M. Lindberg, Marcia C. Linn[1], Amy B. Ellis, Caroline C. Williams**

University of Wisconsin – Madison

[1]University of California, Berkeley


## Background

Gender differences in mathematics performance and ability remain a concern as scientists seek to address the underrepresentation of women at the highest levels of mathematics, the physical sciences, and engineering. Stereotypes that girls and women lack mathematical ability persist and are widely held by parents and teachers.

Meta-analytic findings from 1990 (6) indicated that gender differences in mathematics performance in the general population were trivial, d = -.05.

$$d = \frac{M_{males} - M_{females}}{s_w}$$

However, measurable differences existed for complex problem solving beginning in the high school years (d = +0.29 favoring males), which might forecast the underrepresentation of women in science, technology, engineering, and mathematics (STEM) careers.

Several crucial cultural shifts have occurred since then. In previous decades, girls took fewer advanced mathematics and science courses in high school than boys did, and girls' deficit in course taking was one of the major explanations for superior male performance on standardized tests in high school. By 2000 high school girls were taking calculus at the same rate as boys, although they still lag boys in taking physics. Today women earn 48% of the undergraduate degrees in mathematics, although gender gaps in physics and engineering remain large.

## Methods

**State assessments** of cognitive performance, mandated by **No Child Left Behind** legislation, provide a contemporary source of data on these questions. We contacted the state departments of education of all 50 states, requesting statistical information on gender differences, by grade level and by ethnicity. Responses with adequate statistical information were received from 10 states: California, Connecticut, Indiana, Kentucky, Minnesota, Missouri, New Jersey, New Mexico, West Virginia, and Wyoming. In all cases, the data represent the testing of all students attending school in that grade.

## Results

### Gender and Average Performance

Effect sizes for gender differences, representing the testing of over 7 million students in state assessments, are uniformly < .10, representing trivial differences (Table 1). Of these effect sizes, 21 were positive, indicating better performance by males, 36 were negative, indicating better performance by females, and 9 were exactly 0. This distribution of effect sizes is consistent with a sampling distribution around an effect size of 0, indicating no gender difference (Figure 1). In contrast to earlier findings, these very current data provide no evidence of a gender difference favoring males emerging in the high school years. Effect sizes for the magnitude of gender differences are similarly small across all ethnic groups (Table 2).

### Gender and Variance

Another explanation for the underrepresentation of women at the highest levels in STEM careers is the Greater Male Variability Hypothesis, which states that males have a larger variance in test scores than females do, creating a preponderance of males among very high scorers.

The variance ratio (VR), the ratio of the male variance to the female variance, assesses these differences in variance. Greater male variance is indicated by VR > 1.0. All VRs, by state and grade, are > 1.0 (range 1.11 and 1.21, Table 1). Thus our analyses show greater male variability, although the discrepancy in variances is not large. Analyses by ethnicity show a similar pattern (Table 2).

Does this greater variability translate into gender differences at the upper tail of the distribution?

Table 1. Effect sizes across grades. US math test results are similar across grades 2 through 11.

| Grade | *d* | VR | *N* |
|---|---|---|---|
| Grade 2 | .06 | 1.11 | 460,980 |
| Grade 3 | .04 | 1.11 | 754,894 |
| Grade 4 | -.01 | 1.11 | 763,155 |
| Grade 5 | -.01 | 1.14 | 929,155 |
| Grade 6 | -.01 | 1.14 | 886,354 |
| Grade 7 | -.02 | 1.16 | 898,125 |
| Grade 8 | -.02 | 1.21 | 837,979 |
| Grade 9 | -.01 | 1.14 | 608,229 |
| Grade 10 | .04 | 1.18 | 619,591 |
| Grade 11 | .06 | 1.17 | 446,381 |

## Results Continued

Data from the state assessments provide information on the percentage of boys and girls scoring above a selective cut point based on actual distributions Results vary by ethnic group. Table 3 shows data for grade 11 for the state of Minnesota. For Whites, the ratio of boys:girls scoring above the 95th percentile and 99th percentile are 1.45 and 2.06 respectively, and are similar to predictions from theoretical models. For Asian Americans, ratios are 1.09 and 0.91, respectively. Even at the 99th percentile, the gender ratio favoring males is small for Whites and is reversed for Asian Americans.

### Gender and Item Complexity

Earlier studies indicated that, although girls equaled or surpassed boys in basic computation and understanding of mathematical concepts, boys exceeded girls in complex problem solving beginning in the high school years, d = 0.29. Complex problem solving is crucial for success in STEM careers. We coded test items from all states using a 4-level Depth of Knowledge framework. The results were disappointing. For most states and most grade levels, none of the items were at Levels 3 or 4. Therefore, it was impossible to determine whether there was a gender difference in performance at Levels 3 and 4.

The dearth of Level 3 or 4 items in state assessments has an additional serious consequence. With the increased emphasis on testing associated with NCLB, more teachers are gearing their instruction to the test. If the tests do not assess the sorts of reasoning that are crucial to careers in STEM disciplines, then these skills may be neglected in instruction, putting American students at a disadvantage relative to those in other countries where tests and curricula emphasize more challenging content.

To address this limitation in the state assessments, we turned to the **NAEP** data. NAEP categorizes items as easy, medium, or hard. We coded hard sample items for depth of knowledge. Although no items were at Level 4, many were at Level 3. For hard items that were at Level 3, at grade 12, effect sizes ranged between 0 and 0.15 (average d = 0.07). Thus, even for difficult items requiring substantial depth of knowledge, gender differences were still quite small in these contemporary studies.

## References and Notes


Hyde, J. S., Lindberg, S. M., Linn, M. C., Ellis, A. B., & Williams, C.C. (2008). Gender similarities characterize math performance. *Science, 321,* 494-495. http://www.sciencemag.org/cgi/reprint/321/5888/494.pdf



We thank personnel in each of the 10 responding states for providing data. Special thanks to Margaret Biggerstaff of the Minnesota Department of Education for additional analyses. This research was funded through grant REC 0635444 from NSF. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

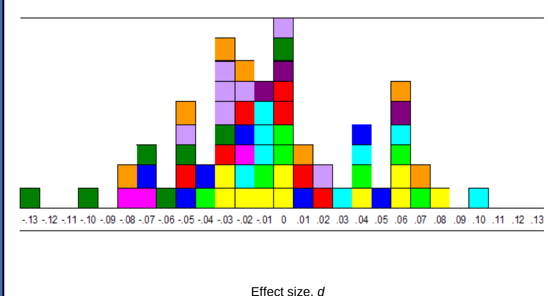

Table 2. Effect sizes across ethnicity. Gender differences in US math test results are similar for different ethnicities.

| Ethnic Group | *d* | VR | N |
|---|---|---|---|
| American Indian/ Alaskan Native | -.08 | 1.07 | 33,661 |
| Asian/Pacific Islander | .01 | 1.06 | 36,727 |
| Black | -.09 | 1.13 | 79,639 |
| Hispanic | -.03 | 1.11 | 159,689 |
| Caucasian | .01 | 1.10 | 588,735 |
| Not Reported | .01 | 1.15 | 6,306,392 |

Table 3. The upper tail. Percentage of Minnesota children scoring above the 95th and 99th percentiles in 11th grade mathematics testing, by gender and ethnicity.

| Ethnic group | Girls above 95th %ile | Boys above 95th %ile | Male: Female ratio | Girls above 99th %ile | Boys above 99th %ile | Male: Female ratio |
|---|---|---|---|---|---|---|
| Asian/Pacific Islander (n = 219) | 5.71% | 6.27% | 1.09 | 1.37% | 1.25% | 0.91 |
| White (n = 3,473) | 5.38% | 7.80% | 1.45 | 0.90% | 1.85% | 2.06 |

Figure 1. Effect sizes across grades and US states. The distribution centers on 0, or no gender difference. Each square represents the effect size for one grade within one state. New Mexico (olive), Kentucky (fuchsia), Wyoming (tan), Minnesota (blue), Missouri (red), West Virginia (lavender), Connecticut (green), California (yellow), Indiana (aqua), New Jersey (purple).




## Conclusions

Our analysis shows that, for grades 2-11, the general population no longer shows a gender difference in math skills. There is evidence of slightly greater male variability in scores, although the causes remain unexplained. Gender differences in mathematics performance, even among high scorers, are insufficient to explain lopsided gender patterns in participation in some STEM fields. An unexpected finding was that state assessments designed to meet NCLB requirements fail to test complex problem solving of the kind needed for success in STEM careers, a lacuna that should be fixed.