Gender Gap Trends on Mathematics Exams Position Girls and Young Women for STEM Careers

John A. Beekman
Ball State University

David Ober
Ball State University

Nine years of results on 4.2 million of Indiana's Indiana Statewide Testing for Educational Progress (ISTEP) mathematics (math) exams (grades 3–10) taken after the implementation of No Child Left Behind have been used to determine gender gaps and their associated trends. Sociocultural factors were investigated by comparing math gender gaps and gap trends for (a) state public schools, (b) state nonpublic schools, (c) a low-performing metropolitan school, and (d) a high-performing suburban school. To control for changing sociocultural factors, multiregression analyses were conducted to predict grade-level (3–10) gender gaps and math scale scores using socioeconomic and ethnicity variables. The underrepresentation of females in earning advanced STEM degrees was investigated by determining the gender of the highest performer on the ISTEP math exams in grades 3–10 for each of state's 292 school corporations. Boys' percentages were higher across all grades by about a 2:1 ratio, similar to high-end results on Scholastic Aptitude Test (SAT) math exams. Simulations of distributions for d = .27 and variance ratio = 1.13 fitted 2013 college-bound SAT math empirical data. Results of the analyses of the state's ISTEP math exam data and the 2013 SAT math scores of college-bound seniors support the arguments that girls and young women possess the abilities to pursue STEM careers that require advanced mathematical skills.

The national interest in increasing the number of science, technology, engineering, and mathematics (STEM) graduates that enter the work force is demonstrated by two relatively recent reports. The first report we cite is from the President's Council of Advisors on Science and Technology (Olson & Riordan, 2012) that targets post-secondary STEM education strategies to increase the number of STEM graduates by 1 million students over the next decade. The second report targets primary and secondary schools through a report from the pharmaceutical industry (Battelle Technology Partnership Practice, 2014) that cites current STEM job openings for up to 600,000 unfilled positions and projected STEM job growth by the U.S. Bureau of Labor.

Both reports espouse the tenet that all students—particularly the "underrepresented majority" of women and minorities—will benefit from hands-on, inquiry-based learning opportunities and earlier STEM-related recruitment strategies. For example, over the past five years members of the Pharmaceutical Research and Manufacturers of America have impacted over more than 1.6 million students and 17,500 teachers through grant competitions for STEM-related education initiatives. About 85% of their funding has been for K-12 education.

The underrepresentation of young women in STEM-related fields and its relationships with mathematics (math) can be documented in numerous ways. One measure of underrepresentation is the annual number of doctorates awarded to females. According to the National Science Foundation's National Center for Science and Engineering Statistics (2013), out of nearly 51,000 doctorates awarded in 2012, 46.2% were awarded to females. In the social sciences, education, and humanities, the percentages of degrees awarded to females were 58.2, 68.7, and 51.8%, respectively; however, the percentages of females earning doctorates in the life sciences, physical sciences (including math), and engineering were 55.7, 28.5, and 22.4%, respectively. When one does a further breakdown of the physical sciences, one obtains percentages of 37.1, 20.6, 41.0, 20.9, and 28.3% for doctoral degrees earned by females in chemistry, computer science, geosciences, physics-astronomy, and math, respectively, during 2012. If one is to improve the career opportunities for females in math, engineering, and the physical sciences, the primary and secondary school pipelines for college-bound STEM students are worthy of careful study. That is the purpose of this investigation.

Since 1995, Indiana has been administering at grades 3, 6, 8, and 10 the Indiana Statewide Testing for Educational Progress (ISTEP) exams in English/language arts (EnLA) and math. Beginning in 2005, the ISTEP exams were expanded to grades 4, 5, 7, and 9 in both subjects to provide accountability exams in all schools for grades 3 through 10. Previous studies (Grissmer, Beekman, & Ober, 2014) have shown that inherent statistical uncertainties in annual achievement scores and small sample sizes at the school level lead to fluctuations in the short term that mask policy changes designed to close achievement gaps. In addition, changing demographics can influence scores in significant ways. Therefore, these sources of uncertainty must be addressed when investigating and identifying effects associated with gender gaps in math at school, corporation/district, and state grade levels.

Indiana's accountability data provide rich aggregated data sets at the state and corporation (district) levels for grades 3 through 10 for the time period 2002 through 2010. This will permit the study of gender gap trends during this period as well as the opportunity to detect gender gap changes between grade levels that can be missed if one only investigates students at a single grade level or at grade levels that are four or more years apart. Indiana has 75,000–80,000 students per grade level and provides about 4.2 million students for the nine-year period studied.

Research Questions

We will investigate questions that relate to the trends of distribution means for girls and boys and the associated gender gaps on Indiana's ISTEP math exams; gaps will also be studied to determine the extent that family variables and changing demographics influence gender effects. Gender differences on math exams were determined using the effect size d, which is the difference in the means of the two groups divided by the pooled within-gender standard deviation. The performances of boys and girls at the high ends of their distributions will be compared as will their variance ratios (VRs), the variance of boys' scores divided by the variance of girls' scores. From these analyses, we will provide evidence to assist in answering questions relating to why girls and young women should be encouraged by parents (caregivers) and teachers, and advised by high school (college) counselors (advisors) to pursue STEM careers.

We will answer five research questions; the first three questions are related to the effect sizes of distribution means of girls' and boys' performance on ISTEP math exams, and the final two questions will entail VRs and performances at the high ends of the distributions. First, for the time period 2002–2010 and across grades 3 through 10, what are the state-level magnitudes of gender gaps and their associated trends? Second, what are the differences in gender gaps and trends when comparing (a) the state's public schools, (b) the state's nonpublic schools, (c) a low-performing metropolitan school corporation, and (d) a high-performing suburban school corporation? Third, how do gender gaps vary over time and across grade levels for the state's 292 school corporations/districts when one controls for changing family and socioeconomic (SES) demographics? Fourth, how do boy–girl VRs change with time and across grade levels? And lastly, what are the percentages of girls and boys (grades 3–10) in corporations who are the highest-performing students, and are these percentages related to the percentages of highest-performing college-bound seniors that take the SAT math exam?

Review of the Literature

Researchers have used numerous exams to quantify the differences in the mathematical performances of girls and young women compared with those of boys and young men. With the passage of the No Child Left Behind (NCLB) Act in 2001, testing data in EnLA and math continue to become available at expanded grade levels and at annual intervals. Additional subject areas and end-of-course assessments that include the sciences are providing teachers and researchers with information to improve the teaching and learning. The increases in state-level data have made it possible to fill in grade-level and trend information not provided by the National Assessment of Educational Progress (NAEP) testing.

Studies from the past decade of gender differences on state, national, and international math exams have focused on gender effect sizes and VRs that provide information about the centers and the high ends of the two distributions. The paper by Hyde, Lindberg, Linn, Ellis, and Williams (2008) used state NCLB math assessment data from 10 states to compare gender effect sizes and VRs for one year. They reported that although they requested data from all 50 states, they received responses with adequate statistical information from only 10 states—California, Connecticut, Indiana, Kentucky, Minnesota, Missouri, New Jersey, New Mexico, West Virginia, and Wyoming. The NAEP scores for these states matched very closely the NAEP national average, so the authors believed that these states were a very representative U.S. sample. From these states, they obtained effect sizes for grades 2 through 11 from a sample size of about 7.2 million students. The values were minor, with values of d and a ± standard error (SE) ranging from $.06 \pm .003$ to $-.02 \pm .002$.

One explanation for the underrepresentation of women at the highest levels in STEM careers uses VRs. The VR is the ratio of the male variance to the female variance. For grades 2 through 11 for the 10 states noted above, these ratios were all greater than 1.0, varying from 1.11 to 1.21. The authors assess such values to not be large. Boys' and girls' scores on the difficult questions of NAEP exams were compared, and it was found that even for more difficult questions requiring substantial depth of knowledge, gender differences were quite small.

In a follow-up study by Hyde and Mertz (2009), these researchers examined gender differences in the general population for the above 10-state study (grades 2 through 11) by ethnicity; high-end distributions were also analyzed to determine the percentages of females who possessed mathematical talent and profound mathematical talent. First, they found that the gender gaps by ethnicity had the same patterns of effect sizes as for all students that Hyde et al. (2008) had reported. Another significant finding of Hyde and Mertz (2009) pertained to complex problem solving. They analyzed 12th-grade NAEP exams for gender differences that required complex problem solving, and they found effect sizes that averaged .07, a very small difference. Thus, Hyde and Mertz (2009) provided further evidence that U.S. girls have reached parity with boys, even in high school, and even for measures requiring complex problem solving.

A later publication by Hyde, Lindberg, Peterson, and Linn (2010) also reported parity for girls and boys in mean math performance in the United States, even in high school. This 2010 study was a meta-analysis using data from 242 studies conducted between 1990 and 2007; they reported an effect size of $d = .05$ and VR = 1.08.

Characteristics of the high end of the distribution of female achievement and reflected in Ph.D. production are reported by Burrelli (2008) using National Science Foundation data that provide historical trends of female employment in academia. This paper examines the proportions of science and engineering (S&E) doctoral degrees earned by women over the past 50 years; this is then compared with the number of women with S&E doctorates employed in colleges and universities over the years 1973 through 2006. The proportions of S&E doctoral degrees earned by women rose from 8% in 1958 to 40% in 2006. These gains varied by field. From 1958 to 2006, the percentages of young women in math, physical sciences, and engineering rose from 6, 4, and less than 1% to 30, 29, and 20%, respectively. Women constituted 33% of all academic S&E doctoral employment and 30% of full-time faculty in colleges and universities in 2006, versus 9 and 7%, respectively, in 1973. The impressive growth in the percentage of Ph.D.s in the mathematical sciences awarded to U.S. citizens who are women is also recorded by Daverman (2011).

Kane and Mertz (2012) made international comparisons of gender differences using measures of the performances by U.S. and international students on international math exams. Their analyses included the Trends in International Mathematics and Science Study (TIMSS), the Program for International Student Assessment (PISA), and the International Mathematical Olympiad (IMO).

The TIMSS data were taken from fourth and eighth grade students. In 2003, approximately 138,000 fourth graders from 26 countries and 256,000 eighth graders from 48 countries participated in the TIMSS, and in 2007 approximately 183,000 fourth graders from 38 countries and 242,000 eighth graders from 52 countries participated. The PISA study included data of 15-year-old school children's scholastic performances and measures reading, math, and science literacy; in 2009, over 475,000 students from 65 countries participated in PISA exams in reading and math. The IMO is "an extremely difficult, proof-based, essay-style examination in mathematical problem solving" (Kane & Mertz, 2012, p. 11). In the 2001–2010 period, over 80 countries per year have sent six-member teams of precollegiate students to participate in the IMO.

Kane and Mertz (2012) found that effect sizes had a wide range of variation. For countries in their study, d ranged from −.027, −.38, and .31 for the Czech Republic, Bahrain, and Tunisia, respectively. For 17 single- and mixed-gender schools, d ranges were from −.71 for Qatar to .30 for Australia (single gender), and from −.54 for Oman to .17 for Egypt (mixed gender). For the eighth grade, TIMSS data VRs ranged from .91 to 1.52; the authors concluded that this was quite similar to a 1.08 value they had obtained for U.S. student data by Hyde et al. (2010). Even though a math achievement gap does exist, Kane and Mertz conclude that gender is not the reason for this gap. They conclude instead that sociocultural factors and gender equity keep both girls and boys from performing higher in math.

The studies just described reach similar conclusions that effect sizes of large data sets for states and nations are small to negligible; when there are statistically significant differences, the sociocultural factors and gender equity just described by Kane and Mertz (2012) are plausible. The studies just described also conclude that U.S. VRs are present at all grade levels. In these reviews, researchers have been unable to explain the underrepresentation of young women in STEM careers on the basis of effect sizes and VRs.

Given these findings, it is quite reasonable to ask what evidence does the current study add to the literature that can be used as evidence for parents (caregivers) and teachers (advisors) to encourage girls and young women to pursue careers in math and other STEM areas during their K-16 education. To answer this question, the following issues are presented in this investigation: (a) effect sizes and VRs are provided for math in grades 3–10 for 2002–

2009, a near decade-long data set following implementation of NCLB; (b) effect sizes and VRs are provided for math and EnLA by gender, ethnicity, and SES to provide perspective for grades 3–10; (c) multiregression models are developed for predicting math pass rates and math gender gaps by grade level over the nine-year period using family variables (ethnicity and SES); (d) gender percentages are presented for the highest-performing math students in 292 school corporations/districts in grades 3–10; (e) the 2013 SAT math distribution (200–800) is presented with percentages of females among the college-bound seniors by intended college STEM major; and (f) time trends of effect sizes for ISTEP math exams in grades 3–10 for the state's lowest- and highest-performing school corporations.

Data Sources

ISTEP Background

Indiana's ISTEP exams were first administered in the spring of 1988 for grades 1, 2, 3, 6, 8, and 9 for the purpose of assessing and improving student achievement. In 1995 testing was moved to the fall, and testing was limited to grades 3, 6, 8, and 10. In 2000, state academic standards were developed; beginning in 2002, ISTEP testing was aligned with the new standards. At each grade level between 70,000 and 80,000, students were tested. Beginning in 2005, grades 3 through 10 were tested with state enrollments between 76,000 and nearly 81,000 students.

Test data studied in this investigation represented more than 2,000 schools (elementary, middle, and high schools) distributed among 292 school corporations (districts). The average annual percentages of girls over all grade levels ranged from 48.3 to 49.5%. Data for the current study included the following grade levels and years: grades 3, 6, and 8 (2002–2009); grades 4, 5, and 7 (2005–2009); grade 9 (2005–2007), and grade 10 (2002–2008). Disaggregated state-level data were reported for the state's public and nonpublic schools, individual schools, and school corporations/districts. Scale scores were reported by ethnicity, gender, and SES status (free/reduced lunch). The highest scale scores, standard deviations, pass rates, and enrollments for each subgroup were also reported.

Analyses

Effect Sizes and VRs

In this study, we will express differences in gender gap means using the effect size, d. This calculation is made using the mean score for boys M_B, minus the mean score for girls, M_G, divided by the pooled within-gender standard deviation $(SE)/N^{1/2}$ where SE and N are the standard error and number of boy or girl test takers, respectively. Note that the standard deviation of $(M_B - M_G)$ is $(SE)/N^{1/2}$. Thus, $d = (M_B - M_G)/(SE)/N^{1/2}$. One can set up the null hypothesis H_0: $d = 0$, versus the alternate hypothesis H_1: $d \neq 0$, and use a critical region $|M_B - M_G| \geq z_{\alpha/2} \cdot (SE)/N^{1/2}$ where N is the number of boy or girl test takers, and $z_{\alpha/2}$ is a critical value from the t-distribution. We assume that the probability distributions for the boys' test scores and the girls' test scores are each normally distributed. We use the values of the t-distribution to determine the critical values for $(M_B - M_G)/(SE)/N^{1/2}$. At the 95% confidence level, with $N \geq 31$, $z_{.025} = 1.96$. If $|M_B - M_G| \cdot (SE)/N^{1/2} < 1.96$, we can say that the effect size is close to zero at the 95% confidence level. Many of our references come to that conclusion, as we do many times.

VRs will be used in this study to examine the differences in the high ends of the boys' and girls' math exam score distributions. The calculation is made by dividing the boys' variance by the girls' variance for a group. Ratios greater than unity indicate greater variability in the scores of boys. When the gender gap scores are small, this means that larger numbers of the boys than girls will be in the high end of the distribution. The cause of VRs greater than unity is of high research interest.

Regression Analyses Predict Exam Scores (and Pass Rates) and Gender Gaps While Controlling for Changing Demographics

Publicly available disaggregated pass-rate data from the ISTEP were analyzed for grades 3, 6, 8, and 10 for exams that were administered between the fall of 2002 through the spring of 2010. The pass rates for the EnLA and math subject areas were investigated. Schools that had 10 or more students in classes at each grade level during the nine-year period for grades 3, 6, and 8, and during the seven-year period (fall 2002–fall 2008) for grade 10 were included in the current study. For grades 3, 6, 8, and 10, there were 907, 476, 390, and 333 schools, respectively.

The regression analyses carried out in this investigation followed the methodology used by Grissmer, Flanagan, Kawata, and Williamson (2000) and Grissmer and Flanagan (2006) on state NAEP data. In the current study, estimations and/or predictions have been made using pass rates and scale scores in math; we then did similar analyses to predict the yearly school-level normalized gender gaps using the same variables as for pass rates and scale scores, namely each school's percentages of girls and boys, free/reduced lunches, special education students, English as a Second Language/Limited English Proficiency (ESL/LEP) students, and ethnicities. We made separate estimates by grade.

Statewide scale score gains in math from a base year. The equation used for estimating math pass rates and scale scores is as follows:

$$y_{ij} = a + \sum f_k F_{ijk} + g_m d_{2002+m} + e_{ij} \qquad (1)$$

where y_{ij} is the percentage of math pass rates or scale scores on a z-scale that has been normalized to the fall 2002 pass rates for the i-th school ($i = 1$, N schools) in the j-th year ($j = 1,9$); F_{ijk} is the k-th family variable for the i-th school in the j-th year; d_{2002+m} is the m-th dummy state gain variable ($m = 1,8$) measured from the fall 2002 baseline year to year m to spring 2010 with values ranging from 0, 1, through 8; e_{ij} is the error term for the i-th school in the j-th year; and a, f_k, and g_m are coefficients of the regression analysis.

Statewide normalized gender gap changes in math from a base year. Equation 1 was used to compute school-level normalized gender gap estimates using the same variables as just described for computing estimates for the scale scores also using Equation 1. The gender gap analyses were also computed for nine years of data for grades 3 (903 schools), grade 6 (470 schools), grade 8 (390 schools), and seven years of data for grade 10 (333 schools).

In these analyses, the variable d_{2002+m} is the m-th dummy state change in the gender gap for each year with the gender gap change variable ($m = 1,8$) measured from the fall 2002 baseline year to year m to spring 2010 with values ranging from 0, 1, . . . through 8.

Highest ISTEP math scale score in corporations by gender and grade level. When reporting disaggregated ISTEP data for Indiana schools, the highest (and lowest) scale scores achieved by students in each disaggregation category (including gender) are reported by school and by corporation/district grade levels. Because STEM subjects are underrepresented in advanced degrees by a 2:1 ratio or more, a determination of the trends of the number of highest achieving math students by gender beginning in grade 3 is key to any intentions designed to improve female representation in STEM. Therefore, grade-level disaggregated data for 2008 were examined for each corporation to determine whether the highest math scale scores were achieved by a girl or boy student, or whether there were ties.

Highest ISTEP math scale score in corporations by gender, SES, and grade level. The highest-performing boys and girls in each grade for the corporations were further examined to determine the effects of SES. Because disaggregated data for the highest scale scores were reported for students with paid lunches and with free/reduced lunches, it was possible to match highest scores for both gender and SES. Grade-level disaggregated data for 2008 were again examined for each corporation to determine the SES (paid or free/reduced lunch) and gender for the highest-performing math student.

Results and Discussion

Time Trends in ISTEP Math Exams

The first research question concerning the magnitudes of gender gaps and their associated trends for Indiana's boys and girls taking the state's ISTEP math exams was answered by determining the effect sizes of the state-level means of the distributions for the two groups. This was done across grade levels for the public schools (grades 3–10) and nonpublic schools (grades 3, 6, 8, and 10) for the nine-year time period 2002 through 2010. These results for both the public and nonpublic schools are presented in Figures 1 and 2, respectively; gender gaps are expressed in standard deviation units, and error bars for the normalized gender gaps are shown above the bars to indicate plus and minus one standard deviation uncertainties.

Indiana's nonpublic schools have annual enrollments at each grade that are typically between 5,500 and 6,000 students while public schools have corresponding enrollments that range between 75,000 and 80,000 students; the total number of students associated with the public and nonpublic school analyses were approximately 4.2 million and 180,000 students, respectively. Indiana's nonpublic schools have pass-rate performance levels and school demographics that are quite similar at all grade levels to those of Indiana's high-performing public schools; the average nonpublic school performs at the level of a public school that performs among the top 10% of schools on ISTEP exams in both EnLA and math.

The public school effect sizes are largest in grades 3, 4, and 10 with values less than .10 and .05 or less for other grade levels. The average effect sizes of .01, .02, and .02 at grades 7, 6, and 9, respectively, suggest a strong junior high/middle school minimum of gender effects in math; this provides evidence that school counselors, parents, and caregivers of girls and young women should encourage them to pursue STEM careers. A similar effect size decline to negative values (grade 4 through grade 9) was found in the study by Hyde et al. (2008) in their study of 7.2 million students.

Research question two is designed to explore the gender gaps of two schools that have widely differing SES and ethnic makeup. In previous studies (Grissmer et al., 2014), it has been demonstrated that across grade levels, high-performing schools are typically associated with low percentages of students on free/reduced lunches (as in

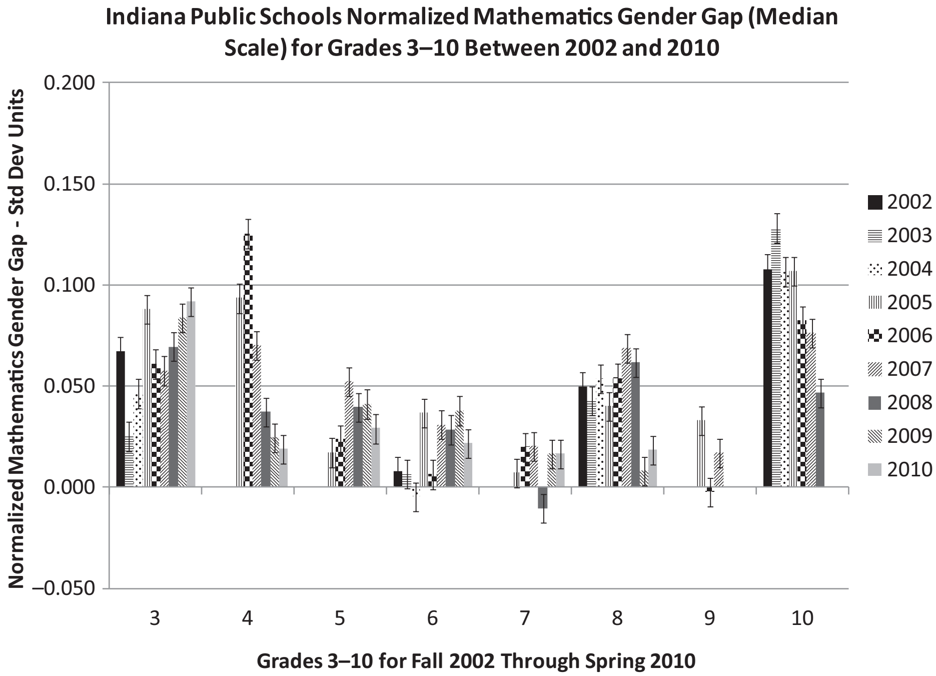


Figure 1. Normalized gender gaps (boys'–girls' scores) in ISTEP math exams for Indiana public schools (grades 3–10) for fall 2002–spring 2010 (grades 3–8) and fall 2002–2008 (grades 9–10). Error bars represent plus and minus one standard deviation of uncertainty. The 2007 state public school percentages for ethnicities and SES, and ISTEP pass rates (averaged across grades 3, 6, 8, and 10) for 315,769 students are as follows: 1.3, 12.0, 6.1, 3.5, 76.2, and 37.5% for Asian, Black, Hispanic, multiracial, White, and free/reduced lunch students, respectively. The corresponding pass rate percentages for all students, Asian, Black, Hispanic, multiracial, White, and free/reduced lunch students are 72.5, 84.0, 48.8, 57.5, 67.8, 77.8, and 58.8%, respectively.

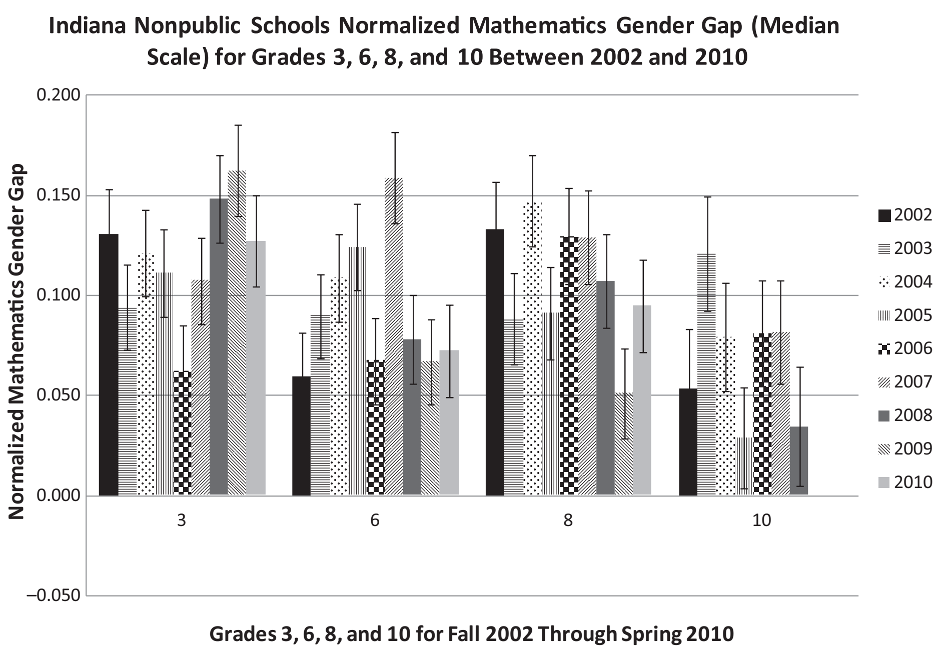


Figure 2. Normalized gender gaps (boys'–girls' scores) in ISTEP math exams for Indiana nonpublic schools (grades 3, 6, 8, and 10) for fall 2002–spring 2010 (grades 3–8) and fall 2002–2008 (grades 9–10). Error bars represent plus and minus one standard deviation of uncertainty. The 2007 state nonpublic school percentages for ethnicities and SES, and ISTEP pass rates (averaged across grades 3, 6, 8, and 10) for 22,256 students are as follows: 2.0, 4.0, 4.2, 2.9, 85.0, and 9.4% for Asian, Black, Hispanic, multiracial, White, and free/reduced lunch students, respectively. The corresponding pass rate percentages for all students, Asian, Black, Hispanic, multiracial, White, and free/reduced lunch students are 86.3, 92.1, 62.0, 74.0, 78.1, 88.8, and 63.3%, respectively.

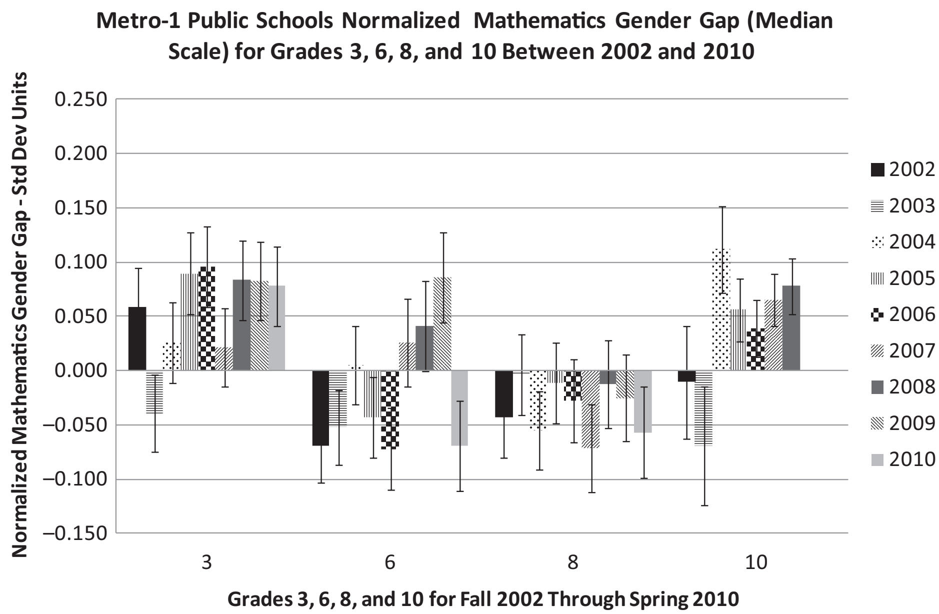


Figure 3. Normalized gender gaps (boys'–girls' scores) in ISTEP math exams for a low-performing metropolitan school, Metro-1 (grades 3, 6, 8, and 10) for fall 2002–spring 2010. Error bars represent plus and minus one standard deviation of uncertainty. The 2007 low-performing Metro-1 school percentages for ethnicities and SES, and ISTEP pass rates (averaged across grades 3, 6, 8, and 10) for 10,280 students are as follows: .3, 56.7, 12.5, 3.7, 23.1, and 56.8% for Asian, Black, Hispanic, multiracial, White, and free/reduced lunch students, respectively. The corresponding pass rate percentages for all students, Asian, Black, Hispanic, multiracial, White, and free/reduced lunch students are 51.3, NR, (not reported for less than 10 students per grade corporation wide), 46.5, 52.3, 60.3, 61.3, and 50.5%, respectively.

suburban schools) and low-performing schools have high percentages of students on free/reduced lunches (as in metropolitan schools).

Presented in Figure 3 are the 2002–2010 time trends of the gender gap values at grades 3, 6, 8, and 10 on math ISTEP exams for the Metro-1 public school system. The Metro-1 school district is the largest corporation in the state, and it has an annual K-12 enrollment of about 33,000 students. The Indiana Department of Education's school new accountability system rated Metro-1 with the lowest overall 2011 score. The system uses performance and growth in EnLA and math achievement to determine ratings.

One of the largest and highest-performing suburban schools in the state (the Suburb-1 school corporation) is located in the Metro-1 suburbs. Suburb-1 school corporation is the fifth largest school corporation in the state with 18,000 students; the state's new school accountability system ranks this school at the 99th percentile. The demographics of this school are quite similar to those of the state's nonpublic schools. (Earlier it was noted that Indiana's average nonpublic schools have student characteristics quite different from the average state public school, and their pass rates on ISTEP exams are at about the 90th percentile of the public school distribution.)

Presented in Figure 4 are 2002–2010 time trends at grades 3, 6, 8, and 10 of the gender gap values on math ISTEP exams for the Suburb-1 school corporation. The normalized gender gaps for this high-performing school corporation are very similar to the trends and gap sizes for the nonpublic school data shown in Figure 2. However, the grade 6 and grade 8 effect sizes for both the high-performing Suburb-1 school (Figure 4) and nonpublic schools (Figure 2) have gender effect values that are above .20 and .10, respectively. These results support a sociocultural cause (concluded by Kane & Mertz, 2012) taking place at a time when girls and young women should be encouraged to develop mathematical skills needed in STEM careers.

Normalized gender gap trends for nonpublic and a high-performing suburban school. Indiana's nonpublic school math gender gap trends in Figure 2 indicate gender gaps exist at each grade level with the nonpublic school gaps for grades 3, 6, and 8 all being relatively flat during the entire time period (2002–2010) at about .10 standard deviation units. The grade 10 nonpublic data trend is from about .09–.05 standard deviation units for the period studied (2002–2008), which is quite similar to the public school trend data for grade 10.

As noted earlier, the performance and demographics of Indiana's average nonpublic schools are both quite similar to public schools with performance levels at the

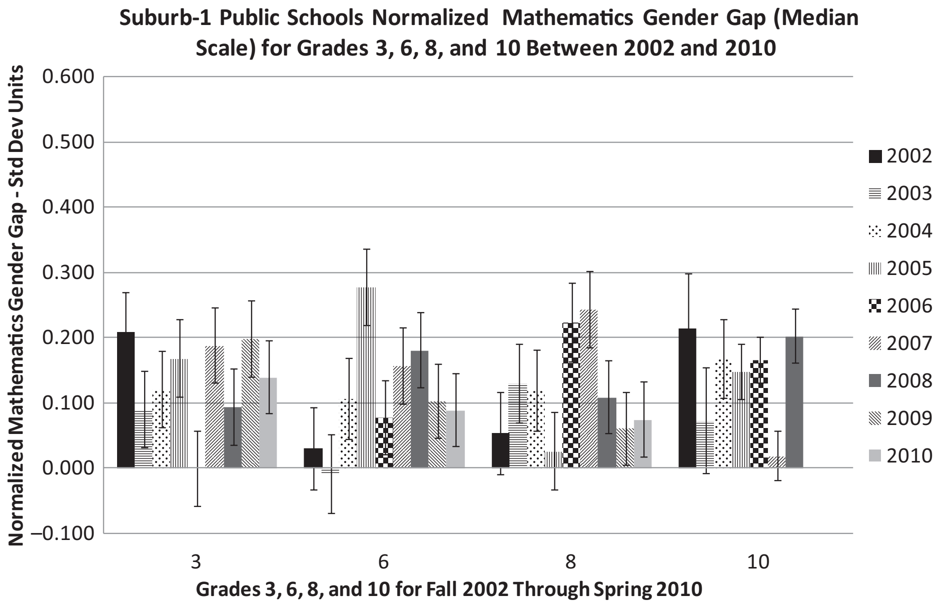


Figure 4. Normalized gender gaps (boys'–girls' scores) in ISTEP math exams for a high-performing suburban school, Suburb-1 (grades 3, 6, 8, and 10) for fall 2002–spring 2010. Error bars represent plus and minus one standard deviation of uncertainty. The 2007 high-performing Suburb-1 school percentages for ethnicities and SES, and ISTEP pass rates (averaged across grades 3, 6, 8, and 10) for 4,607 students are as follows: 8.5, 3.0, 1.9, 4.0, 82.4, and 6.2% for Asian, Black, Hispanic, multiracial, White, and free/reduced lunch students, respectively. The corresponding pass-rate percentages for all students, Asian, Black, Hispanic, multiracial, White, and free/reduced lunch students are 94.0, 97.3, 82.0, 79.5, 88.8, 95.0, and 77.3%, respectively.

90th percentile. Presented in Figure 4 are the normalized gender gap data for one of the highest-performing Indiana public school corporations, the suburban corporation Suburb-1. For grades 3, 6, and 8, there is a relatively flat trend with gaps of .11–.13 standard deviation units. At grade 10, Suburb-1 has trends similar to the public and nonpublic data that begin at .15–.20 standard deviations units; Suburb-1 has a gap of .20 standard deviation units in 2008.

Normalized gender gap trends for a low-performing metropolitan school. Presented in Figure 3 are the normalized gender gap data for a low-performing Indiana public school corporation, the metropolitan corporation Metro-1. The grade 3 trend data for Metro-1 follow the state trend of increasing from .05 to .10 standard deviation units between 2002 and 2010. The grade 10 Metro-1 data rise from −.07 to nearly .08 standard deviation units between 2002 and 2008. However, except for 2008–2010 the Metro-1 trend data for grades 6 and 8 both show slightly higher performance by girl students.

ISTEP math and EnLA gender gaps compared with SES and ethnicity gaps. In order to provide a better context for the effect size of the ISTEP math gender gaps studied in this investigation (see Figure 1) compared with other gaps associated with SES and ethnicity, further analyses were performed on the 2008 data, the final year that grade 10 ISTEP data were available. These results are presented in Table 1.

As noted earlier (see Figure 1), the state average grade-level gender gap effect sizes for ISTEP math are between .01 and .09. It is seen, however, that SES and ethnicity gaps are much larger than math gender gaps by factors of five or more. It should also be noted that a girl–boy gender gap exists on the ISTEP EnLA exams that ranges from .15 to .30, favoring girls. The SES and ethnicity gaps at all grade levels for the ISTEP EnLA are quite similar to the ISTEP math exam gaps.

Regression Analysis Results

A comparison of Indiana's public school math gender gap trends in Figure 1 between 2002 and 2010 for grades 3, 6, and 8, and between 2002 and 2008 for grade 10 indicates that an overall gender gap exists at each grade level. At grade 3, there is an increasing gender gap trend during the nine-year period 2002–2010 that is centered at about .06 standard deviation units and has a value of .09 standard deviation units in 2010. The grade 10 gender gap trend during the seven-year period 2002–2008 is centered at about .10 standard deviation units and decreases to a value of .05 standard deviation units in 2008.

The grade 6 and grade 8 gender gaps are both smaller than those at grades 3 and 10; however, their trend changes are opposite one another and smaller than those of grades

Table 1

Gender, Ethnicity, and SES Effect-Size Gaps on 2008 ISTEP Math and EnLA Exams Were Computed by Grade Level; Average Sample Sizes Were 78,745 Students per Grade, and Standard Errors of the Effect Sizes Are Less than .01

Gap Category	Math Scale Score Effect Size by Grade—Fall 2008						
	3	4	5	6	7	8	10
Male–female	.07	.04	.04	.03	−.01	.06	.05
Asian-White	.24	.40	.43	.46	.55	.48	.39
White-Black	.63	.54	.61	.61	.64	.74	.73
White-Hispanic	.54	.34	.38	.45	.42	.48	.47
Paid-FR lunch	.52	.48	.53	.55	.57	.61	.59
Gap Category	English Language Arts Scale Score Effect Size by Grade—Fall 2008						
	3	4	5	6	7	8	10
Male–female	−.15	−.27	−.25	−.23	−.30	−.30	−.24
Asian-White	.12	.28	.26	.28	.33	.30	.11
White-Black	.57	.58	.58	.61	.62	.62	.72
White-Hispanic	.57	.51	.52	.51	.52	.52	.60
Paid-FR lunch	.52	.57	.59	.59	.59	.62	.64

3 and 10. Specifically, the grade 6 gender gap trend has an increasing gender gap trend during the nine-year period 2002–2010 that is centered at about .02 standard deviation units and has a value of .04 standard deviation units in 2010. However, the grade 8 gender gap trend decreases slowly during the nine-year period 2002–2010 and is centered at about .05 standard deviation units and decreases to a value of .04 standard deviation units in 2010.

Addressing the third research question as to how gender gaps may be affected by changing SES and family variables begins by observing the slow transition (see Figure 1) from an increasing gap at grade 3 that transitions to a decreasing gap at grade 10. Was this slow change due to changing student demographics such as those associated with the increasing percentages of free/reduced lunches (1.7–2.0% per year for grades 3–10, respectively) and decreasing percentages of white students (−.8 to −.9% per year for grades 3–10, respectively) in Indiana during this period? Or, were these trends due to the implementation of Indiana's PL 221 accountability system (1999) and the national NCLB Act (2001), which both had increasing emphasis on pass rates with benchmarks beginning in 2005?

To answer these questions as to whether changing school demographics were responsible for the state gender gap trends, multiregression gap analyses were also performed. School-level gap trend data described previously for grades 3, 6, 8, and 10 were examined from 2002 to 2010 by controlling for each school's percentages of girls and boys, free/reduced lunches, special education students, ESL/LEP students, and ethnicities.

Presented in Table 2 are the grade-level regression coefficients for predicting school-level scale scores using Equation 1. (These analyses yield the state annualized gains d_{2002+m} in scale scores after controlling for demographics.) In grades 3, 6, and 8, eight annualized gain variables d_{2002+m} ($m = 2, 9$) and in grade 10, six annualized gain variables ($m = 2, 7$) were obtained from predicting the corresponding estimates for scale scores from 2106 schools.

Also using Equation 1 as described earlier, school-level normalized gender gap estimates were estimated using the same variables as just described for computing estimates for the scale scores. Presented in Table 2 are the grade-level regression coefficients for computing school-level gender gaps for each school over the seven- and nine-year periods.

The predicted Indiana public schools' normalized gender gaps (boys'–girls' scores) using regression analysis for grade 3 have values ranging from .09 to .17; without regression analysis the values range from .03 to .09 (see Figure 5). The comparable values with regression for grades 6, 8, and 10 are −.01 to .05, .13 to .18, and .14 to .18, respectively. Without controls, respective results for grades 6, 8, and 10 are −.01 to .04, .01 to .07, and .05 to .13.

In summary, controlling for the percentages of girls and boys, free/reduced lunches, special education students, ESL/LEP students, and ethnicities increased the math gender gaps (boys'–girls' scores) on average by .06–.11 standard deviation units for grades 3, 8, and 10. However, for grade 6, there are no systematic differences in gaps

Table 2
ISTEP Math Scale Score and Gender Gap Regression Coefficients for Estimating Annualized 2002–2010 (Grades 3, 6, and 8) and 2002–2008 (Grade 10) Scale Score Gains (d_{20xx}) and Gender Gap Changes (d_{20xx}) Using Family Variables in Equation 1

Scale Score Variable Grade	Beta—Standardized Coefficients 3	6	8	10
d_{2003}	.108***	.132***	.120***	.000
d_{2004}	.143***	.198***	.168***	.748***
d_{2005}	.160***	.283***	.202***	.755***
d_{2006}	.158***	.353***	.191***	.776***
d_{2007}	.129***	.346***	.263***	.797***
d_{2008}	.134***	.342***	.296***	.763***
d_{2009}	.506***	.551***	.347***	
d_{2010}	.645***	.655***	.509***	
Free/reduced	−.375***	−.359***	−.487***	−.179***
Black	−.069***	−.125***	−.229***	−.152***
Special ed	−.061***	−.105***	−.142***	−.096***
Asian	.084***	.118***	.175***	.090***
Hispanic	−.013	−.075***	−.063***	−.001***
Multiracial	−.042***	−.026**	.021*	.005**
ESL/LEP	−.068***	.037***	.011	.005***
Adjusted R^2	.533	.514	.651	.878

Gender Gap Variable Grade	Beta—Standardized Coefficients 3	6	8	10
d_{2003}	−.049**	.013	.018	.025
d_{2004}	−.013	.000	.009	.078**
d_{2005}	.029**	.039**	−.012	.059**
d_{2006}	.012**	−.003	.026	.055
d_{2007}	.001**	.061***	.008	.024
d_{2008}	.012***	.050***	−.026	−.027
d_{2009}	.048***	.088***	−.047**	
d_{2010}	.067***	.062***	−.077***	
Free/reduced	−.029**	−.083***	−.118***	−.055**
Black	−.066***	.024*	−.036**	−.046**
Special ed	−.031***	−.056***	.065***	−.040
Asian	.014*	.013	.196***	.050***
Hispanic	.018	.013	−.144***	.036*
Multiracial	.020**	.015	.136***	.035*
ESL/LEP	−.002	.008	.116***	.013
Adjusted R^2	.016	.015	.030	.015

Note. Significance levels: $*p < .10$, $**p < .05$, $***p < .01$.

between regression predictions and actual gaps for the time period 2002 through 2010. The disappearance of the gender gap at grade 6 (with or without controls) as compared with the other grades studied in this investigation is unclear at this time. Because it is near zero or very small over the nine years studied, we believe it is a real effect and not a test aberration. In a later part of this report, corporation/district gaps were investigated by grade level and deciles for the 2008 exams. For 2008, the gaps remained large for the two highest-performing deciles, and gaps were negative for the schools in the middle two deciles; gaps were effectively zero for the lowest decile.

It is difficult to answer the question about whether the passing of state law PL 221 and the national NCLB Act impacted gender gaps on ISTEP exams. However, based on the results when controlling for family variables (see Figure 5), one would conclude that these two policies have not had an effect on math gender gaps.

VRs Across Grades and Gender Percentages at High Ends of Distributions

VRs across grade levels and their time trends. The fourth research question relates to the work of Hyde and Mertz (2009) who demonstrated how VRs with and without an effect size influence the percentages of females at the high end of math distributions. Therefore, the wide range of consecutive grades (3–10) and extended time period studied in this investigation (2002–2009) permit one to carefully examine VRs from elementary grades into mid high school. We are also able to determine whether mid high school ISTEP math VRs mirror national math exam results such as the SAT math exams. Presented in Table 3 are the ISTEP grades 3 through 10 VRs and grade-level averages for 2002 through 2009. It is seen that the grade 3 and grade 10 VRs begin at 1.14 and end at 1.18; the lowest values of 1.10 at grades 3 and 4 are followed by increases to 1.21 and 1.23 at grades 7 and 8, respectively. No apparent time trends are observed. The 2013 SAT math VR of 1.13 for college-bound seniors (College Board, 2013a) is quite similar to the ISTEP grade 10 values that range from 1.12 to 1.23 between 2002 and 2008; the grade 10 overall average VR for this seven-year period is 1.18. In a later section, female SAT math representation will be predicted with d = .27 and VR = 1.13.

Gender representations at high ends of ISTEP distributions. The fifth and final research question entailed determining at corporation (district) grade levels the gender of the highest-performing student on the state math exams. This was performed for grades 3 through 8 and 10 on the 2008 exam for 292 corporations as a whole as well as for the two SES cases—the highest performer was a paid lunch or a free/reduced lunch student. Presented in Tables 4 and 5 are the results of these comparisons for grades 3 through 8, and for grade 10.

It is seen in Table 4 that boys have a 13.1 percentage point lead out of the 292 corporations studied for grade 3, and this trend widens as one progresses to the higher grades. Between grades 8 and 10, this initial gap widens further to 28.3 percentage points.

A comparison of grade 10 students in Tables 4 and 5 indicates that boys and girls are the top students (or tie for the top score) in 65.7 and 37.4% of the corporations, respectively. However, when the SES values of each

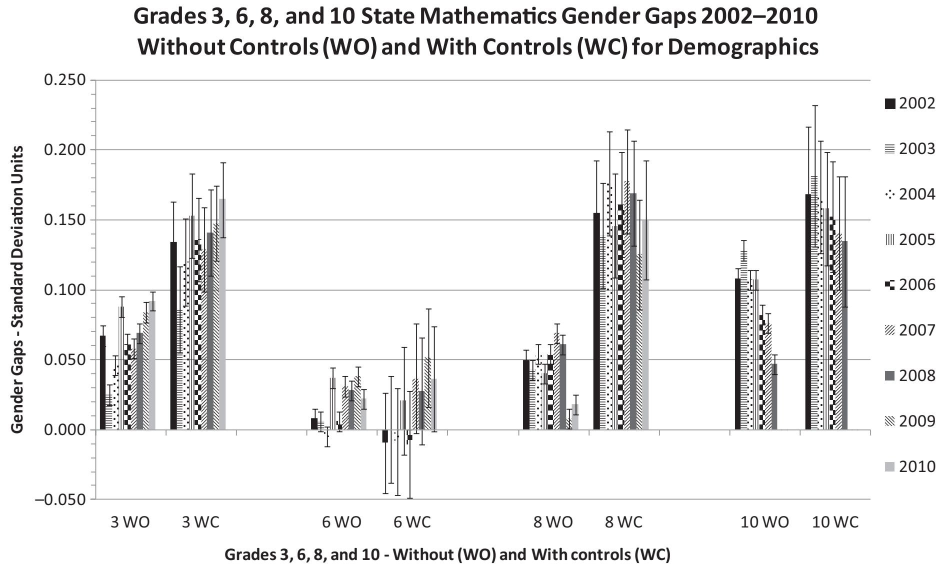


Figure 5. Comparisons of state normalized gender gaps (boys'–girls' scores) in ISTEP math exams for grades 3, 6, and 8 (fall 2002–spring 2010) and grade 10 (fall 2002–fall 2008) controlling for annual family demographics by regression and not controlling for demographics. Error bars represent plus and minus one standard deviation of uncertainty.

Table 3
VRs Have Been Computed by Grade and Year From State Math Scale Scores Where the Average Sample Size Was 78,745 Students per Grade, and Standard Errors of the VRs Are Approximately .01; VRs Have Also Been Computed at Each Grade Level on 2008 English Language Arts and Math State Data for the Gender, Ethnic, and SES Subgroups

Year	VR by Grade—ISTEP Math							
	3	4	5	6	7	8	9	10
2002	1.14			1.18		1.27		1.21
2003	1.12			1.17		1.30		1.19
2004	1.13			1.12		1.27		1.12
2005	1.17	1.12	1.14	1.17	1.23	1.22	1.21	1.16
2006	1.12	1.08	1.11	1.15	1.17	1.19	1.16	1.14
2007	1.13	1.07	1.11	1.18	1.18	1.24	1.18	1.19
2008	1.14	1.14	1.07	1.18	1.24	1.17		1.23
2009	1.15	1.11	1.08	1.10	1.20	1.19		
Average	1.14	1.10	1.10	1.16	1.21	1.23	1.19	1.18

Gap Category	VRs for Math Scale Score by Grade—Fall 2008						
	3	4	5	6	7	8	10
Male–female	1.14	1.14	1.07	1.18	1.24	1.17	1.23
Asian-White	1.61	1.82	1.78	2.42	2.17	1.74	2.69
White-Black	.95	.94	.95	.99	.97	.94	.84
White-Hispanic	.88	.87	.89	.94	.88	.93	.80
Paid-FR lunch	.88	.90	.88	.90	.89	.87	.81

Gap Category	VRs for English Language Arts Scale Score by Grade—Fall 2008						
	3	4	5	6	7	8	10
Male–female	.97	.99	1.11	1.05	1.14	1.12	1.13
Asian-White	1.17	1.44	1.43	1.46	1.57	1.62	2.23
White-Black	1.25	1.05	.97	1.01	1.00	.94	.93
White-Hispanic	1.30	1.11	1.11	1.11	1.08	1.13	1.04
Paid-FR lunch	1.10	.98	.93	.92	.92	.92	.90

Table 4

Percentages of School Corporation in Which the Highest Math Scale Scores Are Achieved by a Boy, Girl, or There Were Ties

Percentages of Highest-Achieving Math Student by Gender in State's 292 Corporations at Each Grade Level—Fall 2008			
Grade	Boy	Girl	Ties
3	48.4	35.3	16.3
4	57.4	35.3	7.3
5	52.6	45.3	2.1
6	58.3	38.5	3.1
7	59.7	37.9	2.4
8	58.6	38.6	2.8
9	—	—	—
10	62.6	34.3	3.1

Note. These data were compiled in 2008 for 292 school corporations, which were classified as metropolitan, suburban, town, and rural; the percents of each classification were 7.5, 26.7, 15.1, and 50.7%, respectively.

Table 5

Percentages of School Corporations in Which the Highest Math Scale Scores Are Achieved by Paid Lunch and Free/Reduced Lunch Students; These Percentages Are Further Sorted to Investigate the Percentages of Boys and Girls That Achieve the Highest Math Scale Score

Percentages of SES and Gender of Highest-Achieving Student in Corporations by Grade Level—Fall 2008				
Grade	Paid Lunch		Free/Reduced Lunch	
	Boy	Girl	Boy	Girl
3	55.1	42.8	20.2	17.8
4	52.7	36.3	16.1	9.6
5	43.2	41.4	13.7	8.2
6	53.8	35.3	8.6	7.9
7	54.1	38.0	9.6	3.8
8	50.0	38.4	13.0	4.8
9	—	—	—	—
10	57.5	34.6	10.6	4.8

Notes. Percent of grade-level totals is greater than 100% since in some schools a boy and girl both achieved the same highest score. These data were compiled in 2008 for 292 school corporations, which were classified as metropolitan, suburban, town, and rural; the percents of each classification were 7.5, 26.7, 15.1, and 50.7%, respectively.

gender of the highest-achieving grade 10 students are compared (ratios of the percentages of free/reduced to paid lunches), boy and girl ratios are .184 (10.6/57.5) and .139 (4.8/34.6), respectively (see Table 5). Another way to express the above .184 and .139 gender difference for the top free and reduced lunch students is to note that of this subgroup of top free and reduced lunch students at grade 10 in Table 5, 57 and 43% are boys and girls, respectively.

Comparisons of gender percentages at high ends of ISTEP and SAT math distributions. Presented in Table 6 are SAT percentages of students by gender and subject area (math and critical reading) for the class of 2013 college-bound seniors (College Board, 2013a) who recorded perfect SAT scores (800), scored in the top 1%, and scored in the top 5% of the two exams just noted. These three respective 2013 SAT math exam percentages for college-bound young women (perfect 800, top 1%, and top 5%) are 34.7, 35.4, and 37.4% (see Table 6), which are quite similar to the 2008 ISTEP percentages of grade 10 young women where 34.3% of the highest corporation (district) test scores were young women (see Table 4). These percentages are also quite similar to the percentages of young women predicted for a normal distribution by Hyde and Mertz (2009) using effect sizes (.05) and VRs (1.2) that are quite similar to those of this investigation.

Average SAT math percentiles of intended college STEM majors for 2013 college-bound seniors and associated percentages of female representation. The U.S. Department of Labor's Bureau of Labor Statistics (2014) reported that there were approximately 3.0 million high school graduates between January and October of 2013, and approximately 2.0 million were enrolled in college in October of 2013. According to the College Board (2013a) there were 1.66 million college-bound seniors that took the math SAT, and these results are reported at ten-point increments between 200 and 800 by percentiles and numbers of total students, males, and females.

Using the means and standard deviations for this group of 2013 college-bound seniors, the female and male SAT math distributions had an effect size of .27 and a VR of 1.13. From these two distributions, the percentage of females at each score increment (200–800) was computed for the empirical data. With the above information, normal distributions were computed for the number of females and number of males taking the exam; as was done with the empirical data, the percentage of females at each score increment was computed for the simulation data ($d = .27$ and $VR = 1.13$). A comparison of the empirical and simulation results for the percentage of females at each scale score increment is presented in Figure 6. The agreement between the empirical and simulation percentages for scale scores between 300 and 700 is excellent; at the 95th percentile and above (720–800), the empirical female representation percentages average 3 percentage points higher than the simulation estimates.

Figure 6 can be used for determining the potential STEM career opportunities for girls and young women. For example, the College Board (2013b) reported that

Table 6

Numbers and Percentages of College-Bound Females, Males, and Total Students Who Took the 2013 Math and Critical Reading Portions of the SAT and Achieved Perfect Scores, Ranked in Top 1%, and Ranked in Top 5%; Approximately 1.66 Million Students Took the Exam

College-Bound Students Fall 2013 Achieving Rankings of Perfect (800), Top 1%, and Top 5%						
Subject		Math			Critical Reading	
			Numbers of Students			
Score/Rank	800	Top 1%	Top 5%	800	Top 1%	Top 5%
Female	4,627	8,508	31,922	5,434	11,014	38,898
Male	8,689	15,536	53,419	5,704	11,397	40,151
Total	13,316	24,044	85,341	11,138	22,411	79,049
Subject		Math			Critical Reading	
			Percentages of Students			
Score/Rank	800	Top 1%	Top 5%	800	Top 1%	Top 5%
Female	34.7	35.4	37.4	52.2	49.1	49.2
Male	65.3	64.6	62.6	47.8	50.9	50.8
Total	100.0	100.0	100.0	100.0	100.0	100.0

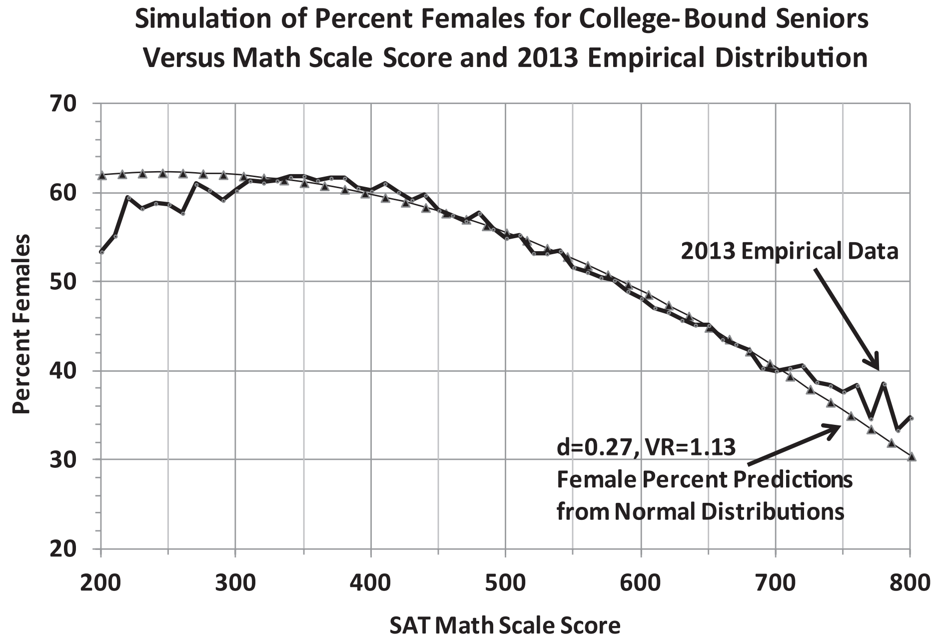
11,361 of the college-bound seniors identified math and statistics as their intended college major, and the average SAT math scale score for this major was 604 (the 76th percentile). In Figure 6, approximately 47% of the students at this percentile are females. Presented in Table 7 are similar data for additional STEM areas for 2013 college-bound seniors.

The total numbers of bachelors, masters, and doctoral STEM degrees awarded each year come from the upper and top portions of the math SAT distributions just cited; these distributions are formed from math (and science) pipelines that originate in public schools like those in grades 3–10 studied in this investigation. When one merges public school math demographics and the relative numbers of intended STEM college major choices made by female college-bound seniors (see Table 7), the upper bounds of the total numbers of bachelors, masters, and doctoral STEM degrees awarded each year can be estimated. Presented in Table 8 are the 2011–2012 number of students and percentages of STEM degrees at the bachelors, masters, and doctoral levels (National Center for Educational Statistics, Digest of Education Statistics, 2012). Following the percentages of high-achieving female math students through grades 3–10, college-bound seniors, and degrees awarded are clearly related; these data indicate that girls and young females have the ability to pursue careers in STEM disciplines.

Where are the future STEM Ph.D.s in the college-bound seniors' 2013 SAT math distribution? Figure 6 and Tables 7 and 8 provide a context for discussing where future STEM doctoral graduates may be in the 2013 SAT math distribution. Table 8 indicates that 8.3% of the bio and med science graduates earn doctoral degrees and 53.3% of them are females, the highest female percentage among the STEM areas. The engineering, math, chemistry, and physics percentages of doctoral degrees to bachelor's degrees earned are quite different with percentages that are 9.0, 8.9, 18.7, and 30.8%, respectively; however, their respective percentages of females receiving doctoral degrees are only 22.8, 28.2, 39.2, and 19.5%. Although it is not always true, it is likely that the majority of the doctoral degree recipients were above the SAT math averages of the college-bound seniors with intended college majors in the STEM areas under discussion. This is verified by two studies. First, the average SAT math score of all entering freshmen for the top 25 institutions ranked by SAT scores was 728 (Rosenberg, 2013). Similarly in a study conducted by the National Science Foundation (National Center for Educational Statistics, Digest of Education Statistics, 2012), the top 50 schools were identified which produced the largest number of S&E doctorates between 1997 and 2006; the average SAT math scores for entering freshmen at these schools for 2013 was 697. These two studies suggest that future STEM Ph.D. recipients are likely to be in the range of SAT math scores of college-bound seniors of between 600 and 800; about 43% of the 417,000 students with scores in this range are females (see Figure 6).

Conclusions

Analyses of nearly a decade of boys' and girls' math scale scores for Indiana's ISTEP exams show that gender gaps measured by effect sizes are present at all grade levels (3–10) at multiple times. These results confirm the



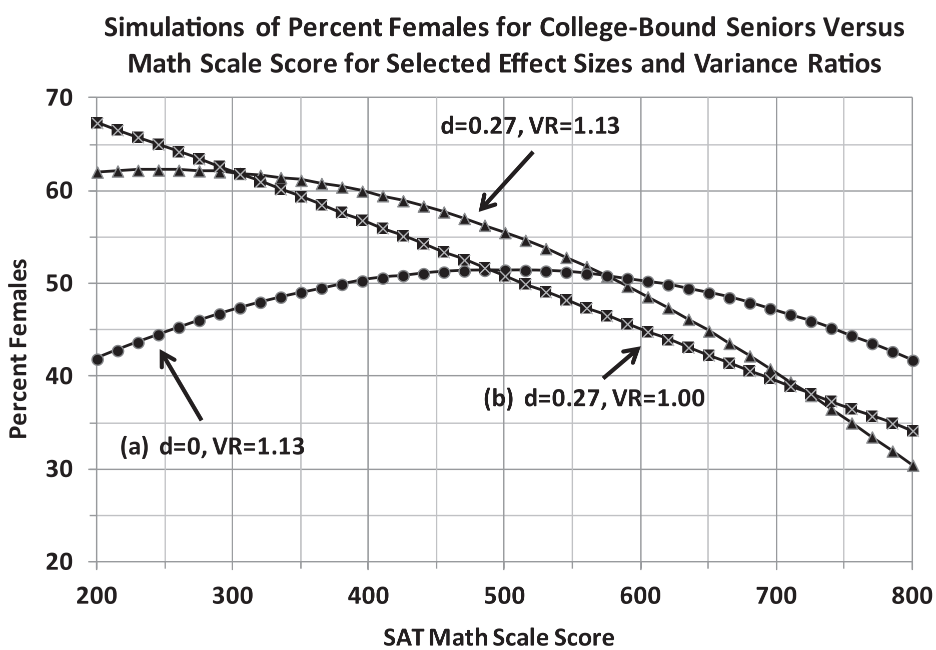


Figure 6. The upper figure compares the empirical percentages of females versus 2013 SAT math scores, and the percentages of females versus 2013 SAT math scores simulated by assuming normal distributions of college-bound females and males with an effect size of .27 and a VR of 1.13. The lower figure compares the female percent predictions of the upper figure (d = .27 and VR = 1.13) with simulations where (a) the effect size is d = 0 with a VR of 1.13, and where (b) the effect size is .27 with equal variances (VR = 1.00).

findings of other researchers (see Hyde et al., 2008, 2010; Kane & Mertz, 2012). The gender gaps are largest at grades 3 and 10, and their trends slowly increase and decrease, respectively, and the gaps are smallest for the middle schools (see Figure 1). When one controls for changing demographics, regression analyses indicate that the predicted gender gaps are larger than actual gaps at grades 3, 8, and 10 by approximately .10 standard deviation units. However, at grade 6 the actual gender gaps are effectively zero for 2002–2004 and for 2006, and approximately .03 for 2005 and 2007– 2010 but without statistical significance (see Figure 5). One also observes that for the highest-performing suburban school studied in this investigation and the nonpublic schools of the state, statistically

Table 7

Intended College STEM Major Student Demographics of 2013 College-Bound Seniors: Average Math SAT Scores, Percentiles, Percent Females, Numbers of Total Students Specifying Intended Majors, and Percentages of Total College-Bound Students in the Intended STEM Major

Intended College STEM Major	Math Scale Score	Scale Score Percentile	Percent Females	Number of Students	Percent of Total College-Bound Seniors
Math	604	75.8	47.6	11,361	.7
Physical sci	582	70.6	50.0	21,057	1.3
Engineering	580	70.0	50.2	132,275	8.0
Bio and med sci	552	62.6	51.5	93,771	5.6
Computer sci	547	60.8	52.4	36,515	2.2
Engineering tech	524	53.2	53.2	19,133	1.2
STEM means/totals	566.5	66.3	50.7	314,112	18.9
National means/totals	515	50.0	54.2	1,660,047	100

Table 8

Numbers of Total Bachelors, Masters, and Doctoral STEM Degrees by Discipline (and Division) Earned in 2011–2012, and Associated Percentages of These Degrees Earned by Females

2011–2012 STEM Degrees Awarded by Disciplines and Divisions							
Disciplines	Divisions	Numbers of Degrees Awarded			Percentages of Degrees Awarded to Females		
		Bachelors	Masters	Ph.D.s	Bachelors	Masters	Ph.D.s
Bio and medical science		95,849	12,415	7,935	58.7	56.7	53.3
Computer Science and information science		47,384	20,917	1,698	18.2	27.7	21.6
Engineering		98,540	45,097	8,856	17.5	23.1	22.8
Math		18,842	6,245	1,669	43.1	40.8	28.2
Physical science		26,663	6,910	5,370	40.1	37.8	32.8
	Astronomy	392	127	170	39.0	39.4	33.5
	Chemistry	12,941	2,318	2,426	48.4	45.4	39.2
	Geo science	4,334	1,321	381	39.1	43.2	44.9
	Physics	5,265	1,681	1,621	19.0	21.3	19.5
Totals							
STEM fields		287,278	91,584	25,528	35.2	31.0	34.6
All fields		1,791,046	754,229	170,062	57.3	59.9	51.4

significant gender gaps exist at grades 3, 6, 8, and 10 about 90% of the time (see Figures 4 and 2, respectively). However, gender gaps disappear across grade levels in the lowest-performing decile of the state's school corporations and one of the lowest-performing metropolitan schools studied in this investigation; they become negative in grades 6 and 8, before becoming positive again in grade 10 (see Figure 3).

When the gender, ethnicity, and SES variables (free/reduced lunch students) were both considered, it becomes apparent that poverty is the most significant variable studied, and it is responsible for the largest differences in the gender gaps (see Tables 1 and 2). The gender gap differences between the suburban and metropolitan schools demonstrate that looking only at distribution means of state, national, and international averages when measuring gender gaps in math provides little significant information about the high-end distribution demographics. Detecting very small signals in samples that possess noise because of larger effects such as SES requires carefully designed studies and large sample sizes.

Multiregression analyses for predicting gender gaps over the nine years studied indicate that gaps become smaller and females outperform boys in schools with increasing percentages of students on free and reduced lunches, particularly in grade 8 (see Table 2); this was also seen in the low-performing metropolitan school (see Figure 3). Conversely, gaps increase in schools with higher percentages of Asian and multiracial students particularly in grades 8 and 10; this trend is also present in nonpublic and high-performing suburban schools (see Figures 2 and 4, respectively). Figure 5 demonstrates that controlling for changing demographics increased gender gaps at grades 3, 8, and 10 while maintaining zero gaps at grade 6.

The early emergence of the underrepresentation of women with bachelors, masters, and doctoral degrees in STEM fields were investigated by studying the upper portions and high ends of the grade 3–10 math exam distributions. In order to investigate the gender gaps of the highest-performing students in Indiana schools, the gender of the highest-performing student in grades 3 through 10 for 2008 was identified by determining the highest math scale score in each of the state's 292 school corporations/districts by grade, and then matching the gender associated with that score. These results showed that in grade 3, the highest-achieving ISTEP math student percentages favored boys 48.3–35.3 with 16.3% ties; by grade 10 a large gap favored boys 62.6–34.3 (see Table 4). This grade 10 result mirrors the 2013 SAT math high-end performances by young women where the sums of their percentages at the highest possible score (800), in the top 1%, and in top 5% are in the 34.7–37.4% range (Table 6). These percentages by young women are also similar to the percentages of math Ph.D.s earned by females the past decade that leveled off to below 30% (see Table 8).

In summary, this investigation agrees with previous studies that the math gender gaps are quite small (.09 in grade 10) to nonexistent (.02 in grade 6). However, the VRs persist across all grades; they are largest at grade 8 and showed no indication of decreasing across grade levels over the past nine years. It was found that the percentages of females across grade levels that had the highest ISTEP math scores in the state's 292 school corporations is about 35% across grade levels in 2008, approximately the same percentage of females in the top 5% and 1% of college-bound students that took the 2013 SAT math. The college-bound seniors with SAT math scores between 600 and 800 are 43% females. Therefore, given the government projections of STEM job needs for the next decade (Olson & Riordan, 2012) and industry estimates of current and future STEM needs (Battelle Technology Partnership Practice, 2014), girls and young women should be encouraged to pursue careers in STEM fields. This study also confirms that any reductions to the math and EnLA gaps due to poverty will be highly beneficial to all students and increase the number of students entering STEM careers.

References

Battelle Technology Partnership Practice. (2014). *Building a 21st century workforce to develop tomorrow's new medicines*. Retrieved from http://www.phrma.org/sites/default/files/pdf/stem-education-report-2014.pdf

Burrelli, J. (2008). *Thirty-three years of women in S&E faculty positions*. Arlington, VA: National Science Foundation, Info Brief, Science Resources Statistics, NSF 08-308. Retrieved from http://www.nsf.gov/statistics/infbrief/nsf08308/nsf08308.pdf

College Board. (2013a). *SAT percentile ranks for males, females, and total group, 2013 college-bound seniors—Mathematics*. New York: College Entrance Examination Board. Retrieved from http://media.collegeboard.com/digitalServices/pdf/research/SAT-Mathematics-Percentile-Ranks-2013.pdf

College Board. (2013b). *2013 college-bound seniors, total group profile report*. New York: College Entrance Examination Board. Retrieved from http://media.collegeboard.com/digitalServices/pdf/research/2013/TotalGroup-2013.pdf

Daverman, R. J. (2011). (American Mathematical Society Secretary). Statistics on women mathematicians compiled by the American Mathematical Society Secretary. *Notices of the American Mathematical Society*, *58*, 1310.

Grissmer, D. W., & Flanagan, A., (2006). *Improving the achievement of Tennessee students: Analysis of the national assessment of educational progress*. TR-381-EDU, 2006. Retrieved from http://www.rand.org/pubs/technical_reports/TR381.html

Grissmer, D. W., Flanagan, A., Kawata, J. H., & Williamson, S. (2000). *Improving student achievement: What do state NAEP scores tell us?* Santa Monica, CA: RAND Corporation, MR-924-EDU, 2000. Retrieved from http://www.rand.org/pubs/monograph_reports/MR924

Grissmer, D. W., Beekman, J. A., & Ober, D. R. (2014). Focusing on short-term achievement gains fails to produce long-term gains. *Education Policy Analysis Archives*, *22*, 5.

Hyde, J. S., & Mertz, J. E. (2009). Gender, culture, and mathematics performance. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(22), 8801–8807.

Hyde, J. S., Lindberg, S. M., Linn, M. C., Ellis, A. B., & Williams, C. C. (2008). Gender similarities characterize math performance. *Science*, *321*(5888), 494–495.

Hyde, J. S., Lindberg, S. M., Peterson, J. L., & Linn, M. C. (2010). New trends in gender and mathematics performance: A meta-analysis. *Psychological Bulletin*, *136*(6), 1123–1135.

Kane, J. M., & Mertz, J. E. (2012). Debunking myths about gender and mathematics performance. *Notices of the American Mathematical Society*, *59*(1), 10–21.

National Center for Educational Statistics, Digest of Education Statistics. (2012). *SAT mean scores and percentage distribution of college-bound seniors, by selected student characteristics*. Retrieved from http://nces.ed.gov/programs/digest/d12/tables/dt12_173.asp

National Science Foundation, National Center for Science and Engineering Statistics (2013). *Women, minorities, and persons with disabilities in science and engineering: 2013*. Special Report NSF 13-304. Arlington, VA: National Science Foundation. Retrieved from http://www.nsf.gov/statistics/wmpd/

Olson, S., & Riordan, D. G. (2012). *Engage to excel: Producing one million additional college graduates with degrees in science, technology, engineering, and mathematics*. Washington, DC: Executive Office of the President. Report to the President.

Rosenberg, M. (2013). *The 25 colleges with the highest SAT scores,* Business Insider. Retrieved from http://www.businessinsider.com/colleges-with-the-highest-sat-scores-2013-4?op=1#ixzz3DDtb4oIN

U.S. Department of Labor's Bureau of Labor Statistics. (2014). *College enrollment and work activity of 2013 high school graduates*. Retrieved from http://www.bls.gov/news.release/hsgec.nr0.htm

Authors' Note

The authors acknowledge with much gratitude the helpful comments and suggestions made by the editor, associate editor, and the reviewers to improve this manuscript.