

Gender Achievement Gaps in U.S. School Districts

Sean F. Reardon

Stanford University

Erin M. Fahle

St. John's University

Demetra Kalogrides

Anne Podolsky

Stanford University

Rosalía C. Zárate

Columbia University

We estimate male-female test score gaps in math and English language arts (ELA) for nearly 10,000 U.S. school districts using state accountability data from third- through eighth-grade students in the 2008–2009 through 2015–

SEAN F. REARDON, EdD, is the endowed Professor of Poverty and Inequality in Education and, by courtesy, a professor of sociology at Stanford University, 520 Galvez Mall, #526, Stanford, CA 94305; e-mail: sean.reardon@stanford.edu. He is currently an Andrew Carnegie Fellow. His research focuses on the causes, patterns, trends, and consequences of social and educational inequality, on the effects of educational policy on educational and social inequality, and on applied statistical methods for educational research. Reardon is the developer of the Stanford Education Data Archive.

ERIN M. FAHLE, PhD, is an assistant professor in the School of Education at St. John's University. Her research focuses on describing and explaining variation in U.S. educational opportunity using applied statistical methods, with an emphasis on understanding systemic gender, racial, and economic inequalities.

DEMETRA KALOGRIDES, PhD, is a research associate at the Center for Education Policy Analysis at Stanford University. She studies achievement gaps, school segregation, poverty, and inequality and works on the creation of the Stanford Education Data Archive.

ANNE PODOLSKY, JD, MA, is a doctoral student in education policy at Stanford University. Her mixed methods research focuses on improving educational opportunities and outcomes, particularly for students from underserved communities.

ROSALÍA C. ZÁRATE, PhD, is a Minority Postdoctoral Fellow in the Department of Education Policy & Social Analysis at Teachers College, Columbia University. Her mixed methods research focuses on higher education policy, retaining underrepresented students in higher education, and improving equity in science, technology, engineering, and mathematics fields.

2016 school years. We find that the average U.S. school district has no gender achievement gap in math, but there is a gap of roughly 0.23 standard deviations in ELA that favors girls. Both math and ELA gaps vary among school districts; some districts have more male-favoring gaps and some more female-favoring gaps. Math gaps tend to favor males more in socioeconomically advantaged school districts and in districts with larger gender disparities in adult income, education, and occupations; however, we do not find strong associations in ELA.

KEYWORDS: accountability testing, English language arts, gender achievement gaps, math, socioeconomic status

Introduction

In recent years, national studies of gender disparities in educational achievement show that, on average, male and female students score similarly on math tests but that female students outperform male students on reading or English language arts (ELA) tests in the United States (Chatterji, 2006; Cimpian, Lubienski, Timmer, Makowski, & Miller, 2016; Fryer & Levitt, 2010; Husain & Millimet, 2009; Lee, Moon, & Heger, 2011; Penner & Paret, 2008; Robinson & Lubienski, 2011; Sohn, 2012). Notably, these average gender achievement gaps vary among states (Hyde, Lindberg, Linn, Ellis, & Williams, 2008; Pope & Sydnor, 2010). But there is little systematic research on variation in the gaps at a smaller geographic scale. Recent studies on the relationship between socioeconomic status (SES) and gender achievement provide evidence that suggests gender achievement gaps may differ substantially among local communities. In particular, they indicate that community and family socioeconomic contexts differentially affect male and female academic achievement and educational attainment (Autor, Figlio, Karbownik, Roth, & Wasserman, 2016, 2017; Buchmann & Diprete, 2006; Chetty, Hendren, Lin, Majerovitz, & Scuderi, 2016; Diprete & Buchmann, 2013; Legewie & Diprete, 2012, 2014). Thus, the large variability in local socioeconomic contexts within the United States may produce variation in gender achievement gaps when measured locally.

In this article, we provide a high-resolution description of the patterns of gender differences in academic performance across the United States, using scores on roughly 290 million standardized tests taken by public school students. We estimate the mean math and ELA test scores for male and female students for each of nearly 10,000 U.S. school districts in Grades 3 through 8 from the 2008–2009 to 2015–2016 school years. These data enable us to estimate male-female test scores gaps, as well as changes in the gaps over grades and cohorts within districts, providing a description of gender differences in academic performance at an unprecedented level of detail. We then investigate the associations between district-level gender math and ELA

achievement gaps and (1) local socioeconomic conditions and (2) local gender disparities in adult income, educational attainment, and occupation.

Similar to the national studies, we find the average school district has no gender achievement gap in math; the average district's ELA gap is roughly -0.23 standard deviations (about three quarters of a grade level) in favor of females. However, we find significant variation in both math and ELA gender achievement gaps among school districts. District-level math and ELA gender gaps are positively correlated—districts in which females' average math scores are higher than males' average math scores tend to also be districts where females' average ELA scores are much higher than males' average ELA scores, and vice versa. Furthermore, we find that math gaps tend to favor males more in socioeconomically advantaged school districts and in districts with larger gender disparities in individual income, education, and occupational characteristics. These two variables explain about one fifth of the variation in the math gaps. In contrast, the associations between these variables and the ELA gender gap are small and inconsistent across models; average SES and socioeconomic disparity variables explain virtually none of the geographic variation in ELA gaps. Given the correlational nature of our analyses, the patterns we describe should not be interpreted as evidence of causal relationships. Rather, they elucidate population-level patterns, which may allow scholars to generate new hypotheses about the influences of gender disparities in education and spark lines of inquiry for future work.

Background

Substantial prior literature explores gender achievement gaps in math and reading during elementary and middle school. In math, national studies using the Early Childhood Longitudinal Study-Kindergarten (ECLS-K) data find mixed evidence as to whether a significant average math achievement gap in favor of males emerges by the end of kindergarten. However, these studies consistently find a significant male-favoring gap by the end of third grade that remains or grows through fifth grade to approximately 0.15 to 0.20 standard deviations (Cimpian et al., 2016; Fryer & Levitt, 2010; Husain & Millimet, 2009; Lee et al., 2011; Penner & Paret, 2008; Robinson & Lubieniski, 2011; Sohn, 2012). From fifth through eighth grade, the trend is reversed and the male-favoring math gap narrows (Robinson & Lubieniski, 2011). A 2010 meta-analysis of data from nationally representative studies of math performance in middle and high school further concluded that male and female math scores did not differ significantly on average (Lindberg, Hyde, Petersen, & Linn, 2010).

ELA gaps, in contrast, favor females by approximately 0.15 to 0.20 standard deviations in kindergarten in the ECLS-K data (Chatterji, 2006; Fryer & Levitt, 2010; Husain & Millimet, 2009; Robinson & Lubieniski, 2011). The ELA gap

narrows modestly (becomes less female favoring) through fifth grade but widens again by eighth grade (Robinson & Lubienski, 2011). In both subjects, these gaps have changed little in recent decades (Fahle & Reardon, 2018).

These national findings, however, mask significant variation in gender achievement gaps among states. Using achievement test data collected from 10 states, Hyde et al. (2008) report state-level gender gaps in performance on 2nd through 11th grade math assessments. Their results show that the male-female math gaps vary among states and grades but are generally near zero (they range from -0.13 to 0.10 standard deviations). Using the National Assessment of Educational Progress (NAEP) data, Pope and Sydnor (2010) investigate the ratios of male to female students scoring above the 95th percentile in math and reading in eighth grade, pooling data from the 2000, 2003, and 2005 assessments. Their results align with national studies, showing that math gaps generally favor males and reading gaps favor females. They also find considerable variation in these upper tail ratios in both subjects. In math, they vary from 0.81 in Hawaii (indicating that 45% of high-scoring students were male; 55% were female) to 2.07 in Kentucky (67% of high-scoring students were male; 33% were female). In reading, the upper tail male-female ratio varies from 0.57 in Massachusetts (indicating that 36% of high-scoring students were male; 64% were female) to 0.22 in Utah (18% of high-scoring students were male; 82% were female). Pope and Sydnor further note that the male-female ratio in math is strongly negatively correlated with the male-female ratio in reading. In states where males are overrepresented among high-achieving students in math, females tend to be overrepresented among high-achieving students in reading.¹

This evidence of variation raises the question “Why might gender achievement gaps differ across geographic contexts?” Prior research—discussed below—suggests that both the gender stereotypes and the availability of socioeconomic resources within a community shape gender disparities in academic interests and achievement among children. Insofar as these factors vary across local contexts, we might expect the gender achievement gaps to also vary locally.

Gender Stereotypes

Gender stereotypes encapsulate conventional beliefs about the household roles, expected behavior, and academic talents of males and females. Traditional conservative gender stereotypes in the United States² generally maintain that men should be the primary breadwinners, while women should be the primary homemakers; that males are assertive, while females are demure; and that males are talented in math and science, while females are talented in languages. When widely accepted in a community, such stereotypes may affect male and female students’ personal beliefs, interests, or actions. There is evidence that children become aware of these gender

stereotypes as early as second grade (Cvencek, Meltzoff, & Greenwald, 2011; Gunderson, Ramirez, Levine, & Beilock, 2012) and that their educational opportunities can be impeded by negative stereotypes.

In particular, stereotypes may contribute to shaping students' beliefs about their academic capability (Eccles, Jacobs, & Harold, 1990; Eccles, Wigfield, Harold, & Blumenfeld, 1993; Jacobs, Lanza, Osgood, Eccles, & Wigfield, 2002), their interest in different subjects (Cech, 2013; Charles & Bradley, 2009), and their academic performance (Ambady, Shih, Kim, & Pittinsky, 2001; Spencer, Steele, & Quinn, 1999; Tomasetto, Alparone, & Cadinu, 2011). Female students may experience stereotype threat in math, resulting in lower test scores that reinforce the negative stereotype (Ambady et al., 2001; Tomasetto et al., 2011). Stereotypes may also be reinforced by parents' or teachers' differential encouragement of male and female children to pursue subject-specific activities (Eccles et al., 1990; Upadaya & Eccles, 2015; Witt, 1997). But, interestingly, parents' rejection of these stereotypes can also moderate their negative effects: Tomasetto et al. (2011) show that the performance of female students whose mothers rejected the "male-math" stereotype did not decrease under stereotype threat.

Therefore, the extent to which community attitudes endorse these gender stereotypes may produce variation in gender performance across contexts. There is some evidence that gender stereotypes or norms differ regionally or among states in the United States, but there is no large-scale research on the extent to which they vary among local communities (Carter & Borch, 2005; Kågesten et al., 2016). At the state level, Pope and Sydnor (2010) examine the associations between stereotypical gender achievement disparities among high-performing students (what they term the gender gap stereotype index) and adults' and children's gender stereotypes. They show that adults' stereotypes about gender roles, as measured by the General Social Survey, explain up to 40% of the regional variation in the gender disparities among high-performing students—census divisions with more traditional stereotypes about gender roles had larger, more stereotypical achievement gaps. They further replicate these findings using student survey questions on NAEP, finding a positive association between students' self-reported agreement with the statement "math is for boys" and the gender gap stereotype index.

Parental Resources

Some research suggests that parental SES and education may influence the development of gender differences in performance among children through parental spending. Although there is not strong evidence that parents spend more money on male or female children (Hao & Yeung, 2015), there is evidence that parents invest their resources (time and money) in their children in gendered ways (Raley & Bianchi, 2006). For example, parents engage in more reading, storytelling, and verbal activities with their

female children as early as 9 months of age (Baker & Milligan, 2016) but believe that their sons are more talented in science and math (Raley & Bianchi, 2006). These gendered patterns of investment may arise from parents' own gender stereotypes or because broader social norms lead children to develop gendered interests, which parents then respond to and reinforce. Either way, parents' investment in and support of gendered activities may create or reinforce children's gender-stereotypical interests, identities, or skills.^{3,4}

However, the variability of parental resources and of spending on children may lead to variability in the extent to which gender stereotypes are reinforced. Affluent, highly educated parents spend more money and more time with their children than their peers (Dotti Sani & Treas, 2016; Duncan & Murnane, 2011; Guryan, Hurst, & Kearney, 2008; Hao & Yeung, 2015; Kornrich, 2016; Ramey & Ramey, 2010). Therefore, if investments are gendered and thus exacerbate children's gendered interests/skills, then greater investments of rich families may lead to greater gender differences in children's interests/skills. As a result, gender achievement gaps may be larger and more stereotypically patterned in higher SES communities.

Empirical evidence to some extent supports this hypothesis. Pope and Sydnor (2010) find that states with higher median income have more stereotypical upper tail gender achievement gaps in math and ELA; however, they do not find significant associations for parental education. Penner and Paret (2008) find that the achievement gap between the highest achieving males and females in math is greatest for students from families with high parental education. Lubienski, Robinson, Crane, and Ganley (2013) find that gender gaps in math performance on the ECLS-K are larger among high-SES students than low-SES students beginning in third grade. Together, these findings suggest that higher SES exacerbates gender achievement gaps, leading to more male-favoring gaps in math and, to some extent, to more female-favoring gaps in ELA.

A competing hypothesis about the influence of SES suggests a different possible pattern. Trivers and Willard (1973) contend that in poorer conditions, including lower socioeconomic conditions, parents will invest more in their daughters because in such contexts, daughters will have higher returns to education (and higher likelihood of finding a high-status spouse) compared with sons. In contrast, in better conditions, parents will invest more in their sons because they have higher potential for economic success than their daughters (Trivers & Willard, 1973).

Again, there is some empirical evidence in support of this hypothesis. Sons of higher status fathers are more likely to attend private school than daughters (Hopcroft & Martin, 2016) and tend to achieve higher degrees of educational attainment than daughters (Hopcroft, 2005; Hopcroft & Martin, 2016). Conversely, daughters from lower income families are more likely to attend private school than sons (Hopcroft & Martin, 2016) and more likely to have higher educational attainment than sons (Cox, 2003).⁵

Among children from low-income families and those raised by a single parent, or a working mother, male students have lower average academic, behavioral, and economic outcomes than females, relative to the gender differences among children from more advantaged families (Autor, Figlio, Karbownik, Roth, & Wasserman, 2016, 2017; Buchmann & Diprete, 2006; Chetty et al., 2016; Diprete & Buchmann, 2013; Legewie & Diprete, 2012, 2014). Entwisle, Alexander, and Olson (2007) find that males receiving meal subsidies perform lower on reading tests than similar females, in part because of parents' lower expectations for males' school achievement. In another study, the same authors find that the math reasoning skills of male students were more strongly influenced by the education level and median household income in the neighborhood than were the skills of their female peers (Entwisle, Alexander, & Olson, 1994). Moreover, recent evidence suggests that living in high-poverty and high-crime communities more negatively affects males' achievement than females' achievement (Chetty et al., 2016; Chetty & Hendren, 2017).

Research Aims and Framework

The primary goal of this article is to provide detailed information about the geographic variation of male-female test score gaps. First, we provide a description of the patterns of gender differences in academic performance among nearly 10,000 U.S. school districts; our data span six grades and 8 years, covering 13 unique student cohorts. Of particular interest is the joint distribution—the variances and the covariance—of gender achievement gaps in math and ELA.

Figure 1 presents a stylized illustration of the dimensions of this joint distribution. School districts will fall into one of the four quadrants of the figure, each of which represents a different stereotypical or gender-favoring average pattern. Districts in the upper left quadrant have stereotypical gender gap patterns—males outperform females in math, on average (positive math gap), and females outperform males in ELA (negative ELA gap). In contrast, districts in the lower right quadrant have gender achievement patterns that are opposite in direction to common stereotypes. Districts in the lower left quadrant have average gender gaps favoring female students in both subjects, while those in the upper right quadrant are places where male students outperform female students on average in both subjects.

We can plot the estimates of math and ELA gaps on a figure similar to Figure 1, examining how the locations of districts in the figure vary by grade, cohort, and local socioeconomic characteristics. The correlation between the math and the ELA gender gaps in this figure will illustrate the extent to which districts vary primarily along the stereotype dimension (the northwest-southeast dimension) or along the gender-favoring dimension (the southwest-northeast dimension). Variation or change along the stereotype dimension

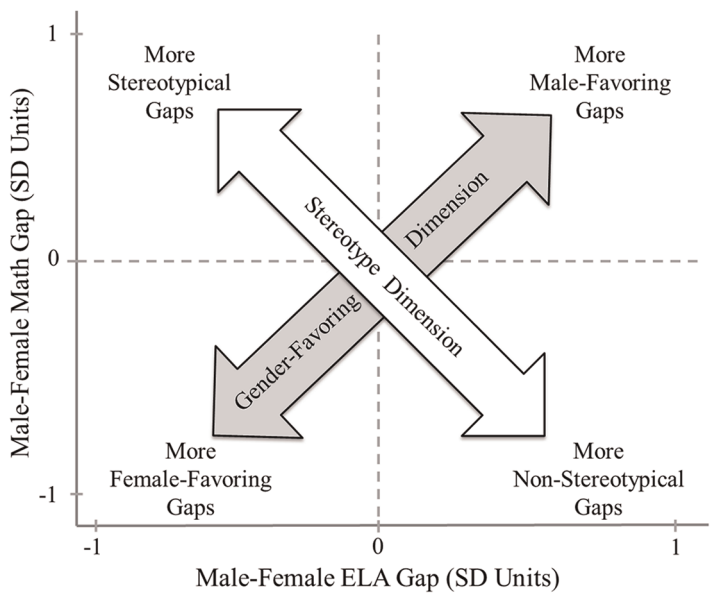


Figure 1. Dimensions of gender achievement.

indicates that districts differ in the extent to which gender achievement gaps conform to the conventional stereotype that males outperform females in math and females outperform males in ELA. Variation or change along the gender-favoring dimension indicates the extent to which gender achievement patterns in both subjects are more male- or female-favoring.

Second, we investigate the associations between district-level gender achievement gaps and two aspects of local communities: average adult SES and gender disparities in individuals' income, educational attainment, and occupations. The former is a measure of both the home and neighborhood conditions of the children as well as of their school lives (peers, school quality, etc.). The latter serves as a proxy measure of local gender role models, norms, stereotypes, and expectations.⁶ Our goal is to provide descriptive evidence of whether gender achievement gaps vary systematically with these district characteristics and to classify that variation along the dimensions in Figure 1.

Data

Achievement Data

The student achievement data used in this study come from the *EDFacts* database, a federal database that includes aggregated state accountability test

score data for every school in the United States. The *EDFacts* data include counts of students scoring at each state-defined proficiency level (e.g., “Below Basic,” “Basic,” “Proficient,” and “Advanced”) on state accountability tests in Grades 3 through 8 for both math and ELA. The counts are disaggregated by school, grade, year, test subject, and gender. These data are available for the 2008–2009 through 2015–2016 school years. In our analysis, we include all public schools serving any students in Grades 3 through 8, regardless of whether they are part of an elementary (K–8) or unified (K–12) school district. We aggregate data from all the schools in a school district and use these aggregated data to measure gender achievement gaps in each school district. We focus on districts rather than schools for several reasons: school districts more closely correspond to local communities than schools, detailed socioeconomic data from the American Community Survey are available at the district but not at the school level, and our estimated gender achievement gaps are much more precise for districts than for individual schools. In aggregating school data within school districts, we assign charter schools to either (1) the public school district chartering them or—if they are not chartered by a traditional public school district—(2) the public school district in which they are geographically located. As a result, a “school district” in our analysis is a geographic unit, rather than strictly an administrative unit, and so corresponds to the population of public school students living in a geographic region.

Using the counts of male and female students in each proficiency category within each geographic school district, we estimate the means and standard deviations of the underlying male and female test score distributions in each district using the heteroskedastic ordered probit (HETOP) model introduced by Reardon, Shear, Castellano, and Ho (2017).^{7,8} We link these estimates to a common scale and standardize them relative to the student-level national distribution of scores within their respective subject, grade, and year, using the methods described by Reardon, Kalogrides, and Ho (2017).

There are roughly 12,000 school districts serving Grades 3 through 8 in the United States; the data allow us to estimate both male and female mean achievement in at least one grade-year-subject for 9,679 school districts.⁹ On average, we have 150 separate grade-year-subject-gender estimates of mean achievement per district in the analytic sample, a total of almost 1.45 million observations.¹⁰ These 1.45 million observations are based on roughly 290 million test score records across the subjects, grades, and years in our sample (an average of roughly 200 test scores per district-grade-year-subject-gender observation in our analytic sample). We denote the estimated mean test score for a given gender subgroup s in district d , subject b , grade g , and year y as $\hat{\mu}_{sbgyd}$.

Covariate Data

We include two primary covariates in our analyses: (1) the average socioeconomic characteristics of parents of children in public school and

Table 1
**Socioeconomic Composite Factor Loadings for the
 Male-Female Socioeconomic Difference Measure**

Male-Female Socioeconomic Difference Measure	Factor Loadings
Median household income (in 10,000s)	0.186
Proportion of adults with BA+	0.182
Poverty rate, 5- to 17-year-olds	-0.176
Percent of 25- to 64-year-olds in the labor force and unemployed	-0.142
Percent of 25- to 64-year-olds not in the labor market	-0.154
Proportion in management, business, and financial occupations	0.200
Proportion in computer, engineering, and science occupations	0.176

Note. The factors were generated in the overall sample of 12,954 districts. The same factor loadings were applied to the male- and the female-specific versions of the socioeconomic composite.

(2) the difference in individual income, education, and occupational characteristics between adult males and females living in the district. For the first, we use a measure of the average SES provided in the Stanford Education Data Archive V2.1 (SEDA, <https://seda.stanford.edu>; Reardon, Ho, et al., 2018). This measure is a composite of the median household income, the proportion of adults with a bachelor's degree or higher, the poverty rate of 5- to 17-year-olds, the unemployment rate, the proportion of households receiving food stamps or in the Supplemental Nutrition Assistance Program (SNAP), and the proportion of single mother-headed households. See the SEDA technical documentation for more information on this composite (Fahle et al., 2018).

Second, we construct a measure of the income, educational, and occupational differences between adult men and women living in a district from the 2006–2010 Education Demographic and Geographic Estimate (EDGE) detailed tables.¹¹ Note that this variable is constructed using data for all adults, not only parents of relevant children as above.¹² EDGE tabulates the demographic and socioeconomic characteristics of adults, by gender, who live in each school district in the United States using American Community Survey (ACS) data. Using principal components analysis, we construct a gender-specific composite of median income, educational attainment, occupation, poverty rates, unemployment rates, labor force participation rates, proportion in business/management occupations, and proportion in science occupations. The factor loadings for each variable are shown in Table 1. Although data on other occupation categories are available in EDGE, we use only measures of participation in business/management and science occupations, as those are stereotypically male-dominated sectors (and the inclusion of other occupational categories did not improve

Table 2
Means and Standard Deviations of Covariates

Covariate	Mean	Standard Deviation	Source
Percent Black in public schools	0.09	0.18	CCD
Percent Hispanic in public schools	0.14	0.21	CCD
Percent other race in public schools	0.04	0.10	CCD
Percent White in public schools	0.73	0.28	CCD
Percent of students in private school	0.09	0.07	ACS, 2006–2010
Socioeconomic status composite, parents	0.10	1.00	ACS, 2006–2010
Socioeconomic composite, male adults	0.34	1.22	ACS, 2006–2010
Socioeconomic composite, female adults	−0.68	0.90	ACS, 2006–2010
Male-female socioeconomic difference	1.02	0.65	ACS, 2006–2010

Note. CCD = Common Core of Data; ACS = American Community Survey. In total, the summary includes 9,679 school districts. Means and standard deviations are unweighted. The socioeconomic status includes income, education, poverty, unemployment, proportion receiving food assistance, and proportion of single mother-headed households. The socioeconomic composite includes income, education, employment, poverty, proportion in business occupations, and proportion in science occupations. The male and female specific socioeconomic composites use the factor loadings shown in Table 1.

the reliability of our measures). The difference between the male and the female composites is then our measure of the individual income, education, and occupational differences between adult males and females in a school district (called the “male-female socioeconomic difference”).

We also include controls for student demographics using data from SEDA and the ACS. SEDA provides district-level measures of the percentage of Black, Hispanic, other race, and White students in public school districts, averaged over grades and years.¹³ These variables are constructed from the Common Core of Data, an annual survey of all public elementary and secondary schools and school districts in the United States. From the ACS data, we create a district-level measure of the percentage of students who attend private school. Table 2 provides a summary of the measures of average SES, male-female socioeconomic difference, and student characteristics from the various data sources. The average district in our sample has a male-female socioeconomic difference of 1 standard deviation and is 9% Black and 14% Hispanic.

Methods

Our aim is to provide a description of gender achievement gaps across U.S. school districts and to generate unbiased estimates of the association between district covariates and subject-specific gender achievement gaps and growth rates of achievement gaps. Complicating these aims is the issue

that measurement of gender achievement gaps may be confounded by differences among the standardized tests used in different states, grades, and years. Specifically, gender achievement gaps measured using tests with more multiple-choice items (vs. constructed-response items) are more male-favoring than tests with fewer multiple-choice items (Beller & Gafni, 2000; Bielinski & Davison, 2001; DeMars, 1998; Garner & Engelhard, 1999; Lindberg et al., 2010; Reardon, Kalogrides, Fahle, Podolsky, & Zárate, 2018).

Because state accountability tests vary in format (as well as other factors such as content that may also influence the measurement of gaps), this poses an issue for generating district-level average gap estimates that are comparable across states and possibly across grades and years. Any gap comparisons will be biased by differences in test format across states, and it is not clear how biased they will be given that information on many tests' item composition is not readily available. Moreover, if the item format of state accountability tests is related to the average SES levels within the state, it will complicate the estimation of unbiased coefficients. For example, if states that have higher average SES also have tests with more multiple-choice questions, this will bias the estimated association between SES and gender achievement gaps upward, leading to a possibly erroneous conclusion that the gender gaps are more male favoring in higher SES school districts.

To address this issue, we adopt the following procedure to purge our district gap estimates of systematic differences that might arise because of differences across states in the content or format of their tests (see the appendix for more detail). We first residualize all the test score means within state, subject, grade, year, and gender—that is, we subtract the statewide average score within a gender, subject, grade, and year from each district's corresponding estimate. Note, however, that the resulting residualized gender-specific district means do not contain any information about the average magnitude of the gender mean or achievement gap within each state (because they constrain average male and female scores within each state to be zero). This will limit our ability to provide an accurate description of the variation in gender gaps across states and will also lead to an underestimation of the true variance of test score means and gaps in the United States. To remedy this, we add the average NAEP scores in the corresponding state-subject-grade-year-gender to the residualized state means. That is, if $\hat{\mu}_{sbgyf}$ and $\hat{\mu}_{sbgyf}^{naep}$ denote the average standardized test scores on the state test and the NAEP test, respectively, in state f for gender s in subject-grade-year bgy , then we compute,

$$\hat{\mu}'_{sbgyd} = \hat{\mu}_{sbgyd} - \hat{\mu}_{sbgyf} + \hat{\mu}_{sbgyf}^{naep} \quad (1)$$

The NAEP-adjusted residualized estimate of the mean, $\hat{\mu}'_{sbgyd}$, is purged of between-state differences in the tests used and therefore of bias due to differences in the content or the format of those tests. Note that Equation

1 implies that all between-state, -subject, -grade, and -cohort variations in the NAEP-adjusted residualized estimates come from the information provided by the NAEP assessments and all within-state-subject-grade-cohort information comes from the state accountability tests. We use these NAEP-adjusted residualized means in all the analyses that follow.¹⁴

A summary of the NAEP-adjusted mean residuals is shown in Table 3. On average, across districts, female students have higher mean test scores than male students in ELA (difference of -0.23 standard deviations, averaged across grades/years). In math, the differences in the district average test score means are close to zero but consistently favor males in nearly every grade and year. Both ELA and math male-female differences tend to favor females more in later grades and in later years.

Models

We fit the following model to construct estimates of the average math and ELA gaps within each district:

$$\begin{aligned}\hat{\mu}'_{sbgyd} = & [\beta_{00d} + \beta_{01d}(\text{grade} - 5.5) + \beta_{02d}(\text{cohort} - 2007)] \cdot \text{math} \\ & + [\beta_{10d} + \beta_{11d}(\text{grade} - 5.5) + \beta_{12d}(\text{cohort} - 2007)] \cdot \text{math} \cdot (\text{male} - 0.5) \\ & + [\beta_{20d} + \beta_{21d}(\text{grade} - 5.5) + \beta_{22d}(\text{cohort} - 2007)] \cdot \text{ela} \\ & + [\beta_{30d} + \beta_{31d}(\text{grade} - 5.5) + \beta_{32d}(\text{cohort} - 2007)] \cdot \text{ela} \cdot (\text{male} - 0.5) \\ & + e_{sbgyd} + r_{sbgyd}\end{aligned}$$

$$\begin{aligned}\beta_{00d} &= \gamma_{000} + \mathbf{X}_d \boldsymbol{\Gamma}_{000} + u_{00d} \\ \beta_{01d} &= \gamma_{010} + \mathbf{X}_d \boldsymbol{\Gamma}_{010} + u_{01d} \\ \beta_{02d} &= \gamma_{020} + \mathbf{X}_d \boldsymbol{\Gamma}_{020} + u_{02d} \\ \beta_{10d} &= \gamma_{100} + \mathbf{X}_d \boldsymbol{\Gamma}_{100} + u_{10d} \\ \beta_{11d} &= \gamma_{110} + \mathbf{X}_d \boldsymbol{\Gamma}_{110} \\ \beta_{12d} &= \gamma_{120} + \mathbf{X}_d \boldsymbol{\Gamma}_{120} + u_{12d} \\ \beta_{20d} &= \gamma_{200} + \mathbf{X}_d \boldsymbol{\Gamma}_{200} + u_{20d} \\ \beta_{21d} &= \gamma_{210} + \mathbf{X}_d \boldsymbol{\Gamma}_{210} + u_{21d} \\ \beta_{22d} &= \gamma_{220} + \mathbf{X}_d \boldsymbol{\Gamma}_{220} + u_{22d} \\ \beta_{30d} &= \gamma_{300} + \mathbf{X}_d \boldsymbol{\Gamma}_{300} + u_{30d} \\ \beta_{31d} &= \gamma_{310} + \mathbf{X}_d \boldsymbol{\Gamma}_{310} \\ \beta_{32d} &= \gamma_{320} + \mathbf{X}_d \boldsymbol{\Gamma}_{320} + u_{32d}\end{aligned}$$

$$e_{sbgyd} \sim N(0, \hat{\omega}_{sbgyd}^2); r_{sbgyd} \sim N(0, \sigma^2); \mathbf{U}_d \sim MVN(0, \boldsymbol{\tau}^2), \quad (2)$$

where $\hat{\mu}'_{sbgyd}$ is the NAEP-adjusted residualized estimated mean test score for subgroup s (male or female), in subject b (math or ELA), district d , grade g ,

Table 3
NAEP-Adjusted Mean Achievement Estimates by Gender, Subject, Grade, and Year

Grade	English Language Arts								Math							
	2009	2010	2011	2012	2013	2014	2015	2016	2009	2010	2011	2012	2013	2014	2015	2016
Male																
3	-0.05	-0.06	-0.06	-0.05	-0.06	-0.06	-0.06	-0.06	0.07	0.07	0.06	0.06	0.05	0.05	0.05	0.05
4	-0.06	-0.07	-0.06	-0.06	-0.07	-0.08	-0.07	-0.07	0.07	0.07	0.06	0.06	0.05	0.05	0.05	0.04
5	-0.07	-0.08	-0.07	-0.07	-0.08	-0.08	-0.08	-0.08	0.07	0.06	0.06	0.06	0.04	0.05	0.04	0.04
6	-0.08	-0.08	-0.08	-0.08	-0.09	-0.09	-0.09	-0.09	0.07	0.06	0.06	0.05	0.04	0.05	0.04	0.03
7	-0.09	-0.09	-0.09	-0.09	-0.10	-0.10	-0.10	-0.10	0.07	0.07	0.06	0.04	0.03	0.03	0.02	0.03
8	-0.09	-0.10	-0.09	-0.09	-0.10	-0.11	-0.11	-0.11	0.07	0.06	0.05	0.04	0.03	0.03	0.02	0.03
Female																
3	0.13	0.13	0.12	0.13	0.12	0.11	0.11	0.11	0.01	0.01	0.01	0.01	0.01	0.02	0.02	0.02
4	0.14	0.14	0.14	0.14	0.13	0.12	0.12	0.12	0.01	0.02	0.02	0.02	0.01	0.02	0.02	0.02
5	0.15	0.15	0.15	0.15	0.14	0.13	0.14	0.13	0.02	0.02	0.02	0.02	0.02	0.03	0.02	0.02
6	0.17	0.17	0.17	0.17	0.16	0.14	0.15	0.14	0.03	0.03	0.03	0.03	0.02	0.04	0.03	0.03
7	0.18	0.18	0.18	0.18	0.17	0.16	0.16	0.16	0.04	0.04	0.04	0.02	0.02	0.03	0.02	0.03
8	0.20	0.20	0.20	0.20	0.18	0.17	0.17	0.17	0.04	0.04	0.04	0.03	0.03	0.03	0.02	0.03
Male-Female																
3	-0.18	-0.18	-0.18	-0.18	-0.18	-0.17	-0.18	-0.17	0.06	0.06	0.05	0.05	0.04	0.04	0.03	0.03
4	-0.21	-0.20	-0.20	-0.20	-0.20	-0.20	-0.20	-0.19	0.06	0.05	0.05	0.04	0.04	0.03	0.02	0.02
5	-0.23	-0.22	-0.22	-0.22	-0.22	-0.22	-0.22	-0.21	0.05	0.04	0.04	0.03	0.03	0.02	0.02	0.01
6	-0.25	-0.25	-0.25	-0.25	-0.24	-0.24	-0.24	-0.24	0.04	0.04	0.03	0.03	0.02	0.01	0.01	0.01
7	-0.27	-0.27	-0.27	-0.27	-0.26	-0.26	-0.26	-0.26	0.03	0.03	0.02	0.02	0.01	0.01	0.00	0.00
8	-0.29	-0.29	-0.29	-0.29	-0.29	-0.28	-0.28	-0.28	0.03	0.02	0.01	0.01	0.00	0.00	0.00	-0.01

Note. NAEP = National Assessment of Educational Progress. The table is based on NAEP-adjusted mean residual estimates. Gender-subject-grade-year averages are estimated using a district fixed-effects model with subject-grade-year-gender fixed effects; coefficients from this model are reported in the table. No adjustments are made for precision.

and year y ; *math* is an indicator variable equal to 1 if the tested subject is math; *ela* is an indicator variable equal to 1 if the tested subject is ELA; *male* is an indicator variable equal to 1 if the tested subgroup is male; *grade* is a continuous variable indicating the tested grade; *cohort* is a continuous variable indicating the tested cohort (where cohort is defined as the tested $year - grade$, and so indicates the year in which a cohort was in their spring kindergarten semester); and \mathbf{X}_d is a vector of (time and grade invariant) district-level covariates. The $\mathbf{u}_{..d}$ are multivariate normally distributed mean-zero residuals with variance-covariance matrix $\boldsymbol{\tau}^2$ to be estimated, r_{sbgyl} is a normally distributed mean-zero residual with variance σ^2 to be estimated, and e_{sbgyl} is a normally distributed mean-zero sampling error term with known variance $\hat{\omega}_{sbgyl}^2$ equal to the sampling variance of $\hat{\mu}_{sbgyl}^*$.¹⁵ Model estimation is performed using the hierarchical linear modeling software.

In other words, this model uses up to 192 (2 genders, 6 grades, 8 years, 2 subjects) estimates of gender-grade-year-subject NAEP-adjusted means in each district to estimate each districts' average performance in math, male-female gender achievement gap in math, average performance in ELA, male-female gender achievement gap in ELA, growth over grades and cohorts of each of those terms, and residual error term. The average performance and gaps ($\beta_{.0d}$) are then modeled as a function of a vector of district covariates and a residual error term indicating the difference between the true average/gap that is predicted by the covariates and the national average/gap. Similarly, the grade ($\beta_{.1d}$) and cohort slopes ($\beta_{.2d}$) of the average performance and gaps are modeled as functions of district covariates and district-specific residual error terms. We exclude the district-level error terms on the grade slopes of the gender achievement gaps in math and ELA (β_{11d}, β_{31d}) because our initial models including them indicated that their variance was not statistically distinguishable from zero. That is, we cannot reject the null hypothesis that the gender gaps change at the same rate from third to eighth grade in all districts.

In our null model (Model 1), we do not include any district-level covariates, \mathbf{X}_d . From this model, we recover an estimate of the true variance in the gender achievement gaps among U.S. school districts. We then formally test whether average parental SES and adult male-female socioeconomic differences are associated with overall achievement, gender achievement gaps, and the growth across grades or cohorts in both these measures by adding these measures, in \mathbf{X}_d , to the null model (Model 2). Next, we add racial composition variables to test whether the average SES and male-female socioeconomic difference associations hold after controlling for student demographics (Model 3). Finally, we estimate a variant of Model 3 that includes state-level random effects on the four $\beta_{.0d}$ terms (Model 4). In this model, we state-mean center the district covariate vector \mathbf{X}_d ; as a result, the coefficient estimates from Model 4 are the same as we would obtain had we added state fixed effects in each of the $\beta_{..d}$ equations (but the model is

more computationally efficient than the fixed-effects model). Note that centering the vector \mathbf{X}_d changes the interpretation of the estimated coefficient vector $\hat{\Gamma}_{..0}$. It now represents average within-state associations. This has the advantage of ensuring that the associations are not biased by between-state differences in the standardized tests.

Results

Table 4 reports selected coefficients from the fitted models. Model 1 (the null model) shows that the average district male-female math gap is approximately 0.03 standard deviations and the average ELA gap is -0.23 standard deviations. In other words, in the average school district, there is essentially no gender achievement gap in math, but two thirds of a grade-level difference in favor of females in ELA (one grade level in Grades 3–8 is roughly equal to 0.33 standard deviations). Because we center grade and cohort at the midpoints of the grades and cohorts contained in our data (grade 5.5 and cohort 2007), these can be interpreted as the average gaps halfway through fifth grade for the middle cohort in our sample. On average, both the math and the ELA gaps change in favor of females from Grades 3 through 8 (by roughly -0.06 standard deviations in math and -0.10 standard deviations in ELA in five grades). Across cohorts, the gaps change relatively little per year; the models imply that the average math gap changed by roughly -0.05 standard deviations (in favor of females) and by roughly 0.02 standard deviations (in favor of males) from the earliest to the latest cohorts in our sample. These results are largely consistent with the less parametric patterns evident in Table 3.

The gender gaps vary significantly among districts in both ELA and math. From Model 1, the estimated distribution of the ELA gaps (mean = -0.23 ; $SD = 0.049$), assuming normality, implies that 95% of school districts have ELA gaps between -0.13 and -0.33 standard deviations (i.e., favoring females by between one third and one grade level). In no district is males' average performance higher than that of females. A district's ELA gap would have to be almost 5 standard deviations from the mean ELA gap in this case. The distribution of math gaps (mean = 0.027; $SD = 0.049$) implies that 95% of districts have math gaps that are between -0.07 and 0.13 standard deviations, favoring males in 72% of school districts and females in 28% of school districts. In Figure 2, we map estimates of the ELA and math achievement gaps across U.S. school districts to provide a visual representation of this variation.¹⁶ In the maps, orange indicates that female students outperform male students on average, and blue the opposite; white indicates missing data. Darker shades signify larger average gaps. In both subjects, the maps confirm that there is a clear variation in the gaps among and within states, although some states have less between-district variation in the gaps than others.

Table 4
Selected Multivariate Regression Results

	Math				English Language Arts			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Male-female gap (intercept)	0.0267*** (0.0006)	0.0269*** (0.0006)	0.0275*** (0.0006)	0.0190*** (0.0026)	-0.2295*** (0.0006)	-0.2295*** (0.0006)	-0.2298*** (0.0006)	-0.2382*** (0.0039)
Socioeconomic status composite		0.0197*** (0.0007)	0.0136*** (0.0008)	0.0145*** (0.0009)		0.0011 (0.0007)	0.0014 (0.0009)	-0.0037*** (0.0009)
Male-female socioeconomic difference		0.0119*** (0.0011)	0.0125*** (0.0011)	0.0125*** (0.0011)		0.0047*** (0.0012)	0.0030*** (0.0012)	0.0015 (0.0011)
Percent Black			-0.0608*** (0.0038)	-0.0567*** (0.0043)			-0.0283*** (0.0041)	-0.0363*** (0.0043)
Percent Hispanic			-0.0048 (0.0029)	-0.0097* (0.0040)			0.0403*** (0.0031)	0.0170*** (0.0040)
Percent other race			0.0228*** (0.0065)	0.0110 (0.0076)			-0.0138* (0.0070)	-0.0131† (0.0076)
Percent private			0.0182* (0.0088)	0.0133 (0.0092)			-0.0167† (0.0096)	-0.0021 (0.0092)
Grade	-0.0117*** (0.0003)	-0.0118*** (0.0003)	-0.0120*** (0.0003)	-0.0161*** (0.0007)	-0.0193*** (0.0003)	-0.0194*** (0.0003)	-0.0195*** (0.0003)	-0.0227*** (0.0007)
Socioeconomic status composite		-0.0020*** (0.0003)	-0.0004 (0.0004)	-0.0005 (0.0004)		-0.0008*** (0.0003)	-0.0003 (0.0004)	-0.0001 (0.0004)
Male-female socioeconomic difference		0.0036*** (0.0005)	0.0033*** (0.0005)	0.0032*** (0.0005)		0.0027*** (0.0005)	0.0024*** (0.0005)	0.0023*** (0.0005)
Percent Black			0.0128*** (0.0017)	0.0118*** (0.0017)			0.0038* (0.0017)	0.0052*** (0.0017)
Percent Hispanic			0.0034* (0.0013)	0.0031* (0.0013)			0.0025* (0.0013)	0.0019 (0.0013)
Percent other race			0.0114*** (0.0032)	0.0122*** (0.0032)			0.0111*** (0.0031)	0.0105** (0.0031)
Percent private			-0.0171*** (0.0032)	-0.0160*** (0.0032)			-0.0025 (0.0031)	-0.0030 (0.0031)

(continued)

Table 4 (continued)

	Math				English Language Arts			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Cohort	-0.0042*** (0.0002)	-0.0043*** (0.0002)	(0.0043) -0.0045*** (0.0002)	(0.0043) -0.0066*** (0.0005)	0.0019*** (0.0002)	0.0019*** (0.0002)	(0.0042) 0.0021*** (0.0002)	(0.0042) 0.0027*** (0.0005)
Socioeconomic status composite		-0.0006** (0.0002)	0.0008** (0.0002)	0.0007* (0.0003)		-0.0002 (0.0002)	-0.0007** (0.0003)	-0.0005† (0.0003)
Male-female socioeconomic difference		0.0018*** (0.0004)	0.0014*** (0.0004)	0.0014*** (0.0004)		0.0000 (0.0004)	0.0002 (0.0004)	0.0001 (0.0004)
Percent Black			0.0103*** (0.0012)	0.0095*** (0.0012)			-0.0037** (0.0013)	-0.0024† (0.0013)
Percent Hispanic			0.0047*** (0.0009)	0.0042*** (0.0009)			-0.0021* (0.0009)	-0.0025** (0.0009)
Percent other race			0.0066** (0.0022)	0.0064** (0.0022)			0.0033 (0.0022)	0.0026 (0.0022)
Percent private			-0.0151*** (0.0030)	-0.0147*** (0.0030)			-0.0035 (0.0030)	-0.0038 (0.0030)
Within-district standard deviation	0.097	0.097	0.097	0.097	0.097	0.097	0.097	0.097
District-level gap standard deviation	0.049	0.042	0.041	0.041	0.049	0.048	0.047	0.041
District-level grade slope standard deviation								
District-level cohort slope standard deviation		0.009	0.009	0.009	0.010	0.010	0.010	0.010
State-level gap standard deviation				0.01396				0.025
District-level correlation between math and ELA gaps	0.83	0.92	0.94	0.98	0.83	0.92	0.94	0.98
District-level gap reliability	0.60			0.52	0.60			0.53
District-level grade slope reliability								
District-level cohort slope reliability				0.32				0.37
State-level gap reliability	0.33			0.82	0.37			0.91
District-level relative R ²		0.26	0.30	0.29		0.00	0.05	0.28

Note. Selected coefficients are shown. Standard errors are in parentheses; The number of observations in all the models is 1,447,540, and the number of districts in all models is 9,679. All models use the NAEF-adjusted within state-subject-grade-year-gender mean residuals.
† $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

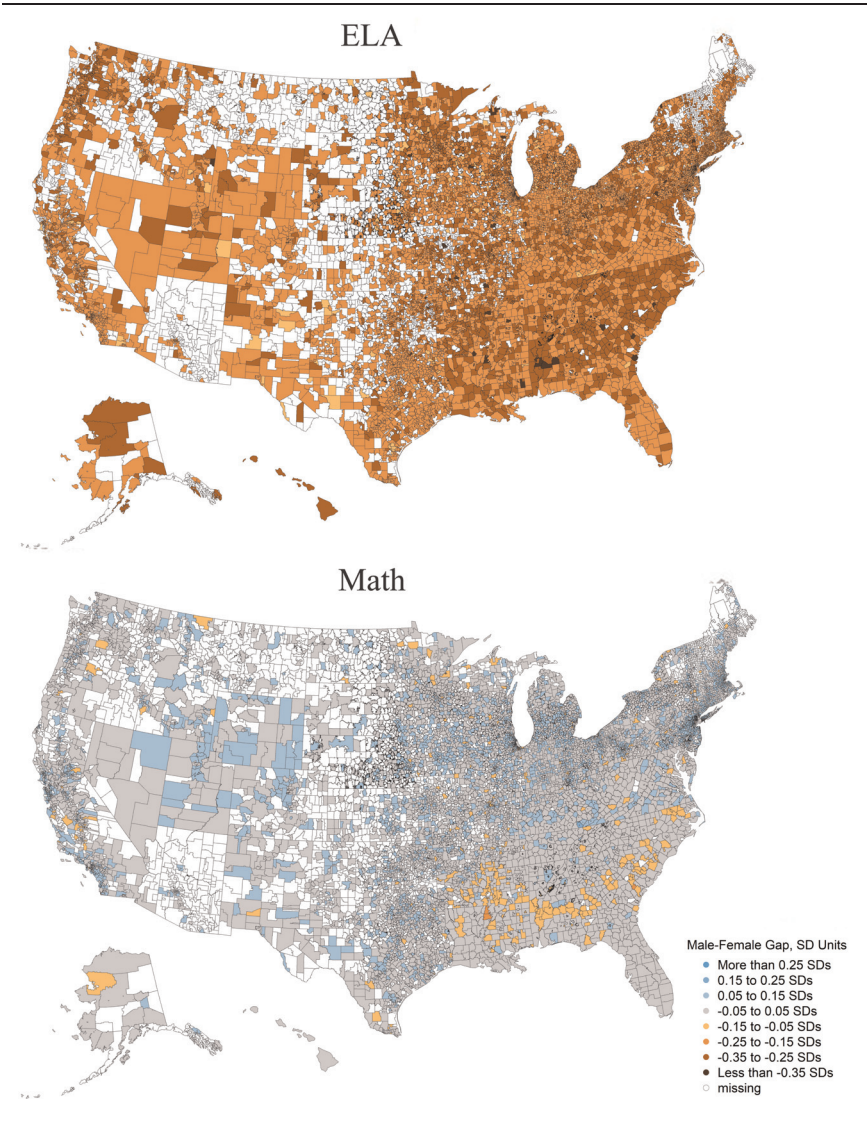


Figure 2. Average gender achievement gaps in U.S. school districts.
Note. ELA = English language arts. The empirical Bayes estimates from Model 3 are displayed.

Figure 3 shows the empirical Bayes estimate of the male-female math achievement gap plotted against the empirical Bayes estimate of the ELA achievement gap for each district in our sample. Note that the gaps fall predominantly in the upper left quadrant indicating that in most school districts,

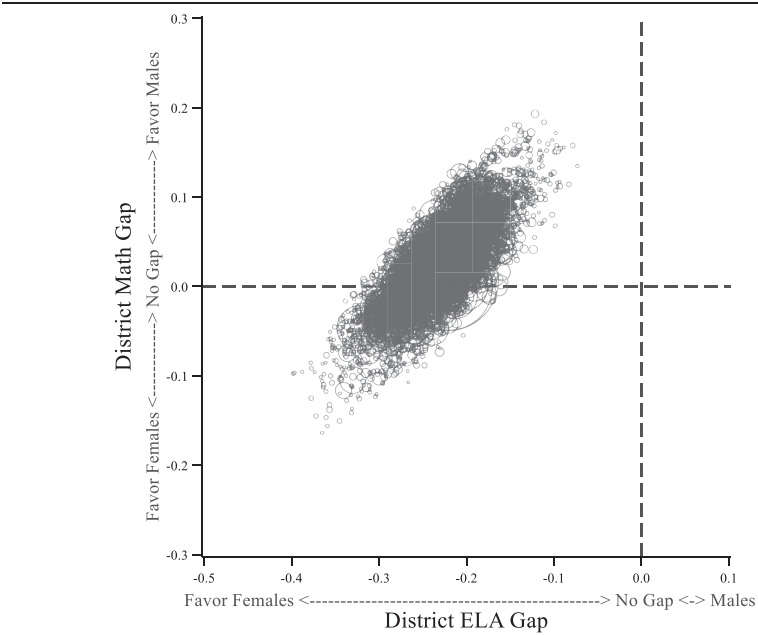


Figure 3. The relationship between math and ELA gender achievement gaps, across U.S. school districts.

Note. ELA = English language arts. Observations are empirical Bayes test score gap estimates from Model 3, weighted by the average number of students in a district-subject-grade-year. They are in standard deviation units, standardized to the NAEP national distribution within each subject, grade, and year.

gender achievement gaps align with subject-specific gender stereotypes. In contrast, the math and ELA gaps are positively correlated: Districts with more male-favoring math gaps tend to also have more male-favoring (less female-favoring) ELA gaps. This suggests that variation among districts is gender favoring (as described in Figure 1), despite the fact that gaps on average are stereotypical. Indeed, Table 5 shows that the estimated correlation between the math and the ELA gaps is 0.83.

Table 5 further shows that there is a moderate correlation between the average performance and the male-female gap in math ($r = 0.48$) but that there is almost no association in ELA ($r = 0.04$). Districts with higher math performance tend to have more male-favoring math gaps, but the average performance of students in ELA is unrelated to the size of the ELA gap.

Model 2 in Table 4 shows that both district SES and male-female socioeconomic differences are positively associated with the male-female math gap. In wealthier school districts and in school districts with more

Table 5
Correlation Matrix

	Average Math Score	Male-Female Math Gap	Average ELA Score	Male-Female ELA Gap	SES Composite	Male-Female SES Composite Difference	Proportion of Black Students
Average math score	1.00						
Male-female math gap	0.48	1.00					
Average ELA score	0.93	0.52	1.00				
Male-female ELA gap	0.08	0.83	0.04	1.00			
SES composite	0.73	0.48	0.77	0.00	1.00		
Male-female socioeconomic difference	0.44	0.34	0.44	0.07	0.45	1.00	
Proportion of Black students	-0.40	-0.41	-0.38	-0.15	-0.47	-0.19	1.00

Note. ELA = English language arts; SES = socioeconomic status. The correlations between the average scores and gaps, and between these parameters and the covariates are disattenuated to account for measurement error in the gaps and average scores. The correlations among the covariates are based on the observed data, with no disattenuation or weighting. Correlations among the average test score and test score gap measures are estimated from Model 1. Correlations between the average performance/gaps and the covariates (SES composite, the male-female socioeconomic difference, and the percentage of Black students) are derived from the explained variance between a model that includes each covariate separately (models not shown in article) and a model with no covariates (Model 1).

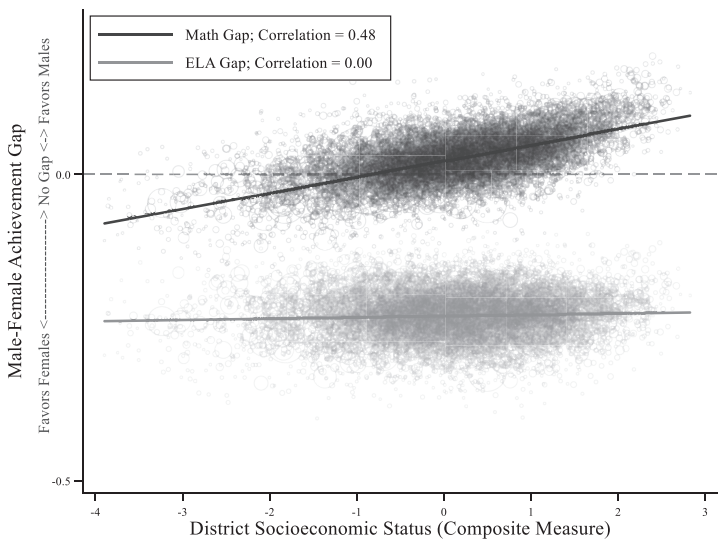


Figure 4. Male and female achievement in English language arts (ELA) and math versus socioeconomic status (SES) in U.S. school districts.

Note. The empirical Bayes estimates shown underestimate the true variance in the male-female achievement gaps.

socioeconomic gender inequality, math gaps favor males more, on average. In ELA, there is no significant relationship between overall SES and the gender gap in Model 2. There is a small positive association between local gender income, educational and occupational disparities, and the ELA gender gap: ELA gaps favor males more in communities with large socioeconomic disparities between adult males and females. Overall, these results suggest meaningful associations between math achievement gaps and both local SES and local gender socioeconomic disparities but no or very small associations for ELA. This is evident in the proportion of variance explained by the two SES variables (Model 2): Together they explain 26% of the variance in math gaps but none of the variance in the ELA gaps.

We illustrate these results in Figures 4 and 5. Figure 4 shows the male-female achievement gaps in math plotted against the SES composite. For the ELA gaps, the slope of the line is nearly flat; the gaps are uncorrelated ($r = 0.00$) with SES (as is evident in the multivariate models). However, in math, the slope of the line is positive ($r = 0.48$), indicating that the gap is more male favoring in high-SES places compared with low-SES places. Figure 5 shows positive relationships in both math and ELA between achievement gaps and our measure of socioeconomic gender inequality

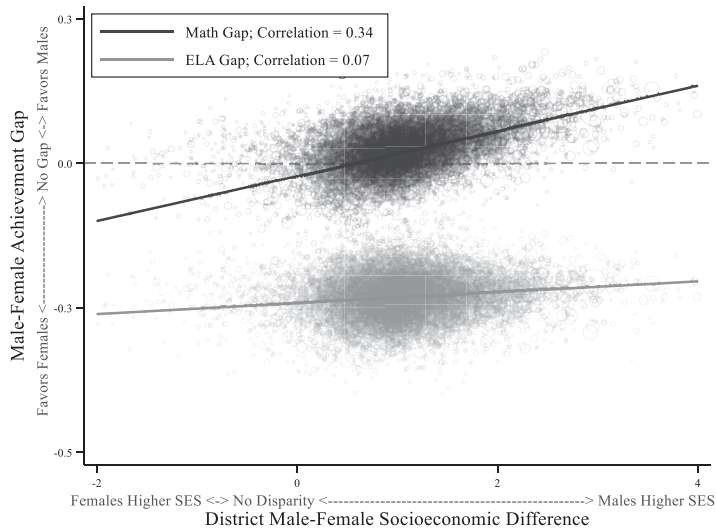


Figure 5. Male and female achievement in English language arts (ELA) and math versus the male-female socioeconomic difference in U.S. school districts.

Note. The empirical Bayes estimates shown underestimate the true variance in the male-female achievement gaps.

(income, education, and occupation composite difference). However, the relationship is much steeper in math (as seen in Model 3), indicating that in places with larger male-female socioeconomic disparities among adults math and, to a much smaller extent, ELA gaps tend to favor male students more (relative to the average district).

After controlling for racial composition and the percentage of students in private school, (Model 3) or estimating the association within states (Model 4), the associations between the math gaps and SES and gender socioeconomic disparities remain statistically significant (see Table 4). In ELA, the results for SES are inconsistent in sign and significance across models, and the coefficients are very small and of little practical significance. The association between gender socioeconomic disparities and ELA gender gaps is smaller and no longer statistically significant after controlling for racial composition and the percentage of students in private school.

Although it is not the focus of our analysis here, Model 3 indicates that racial composition is also associated with the gender gaps. For example, both math and ELA gaps are more female favoring in districts with a larger proportion of Black students. Although this suggests that within-district gender gaps are smaller (more female favoring) among Black students than

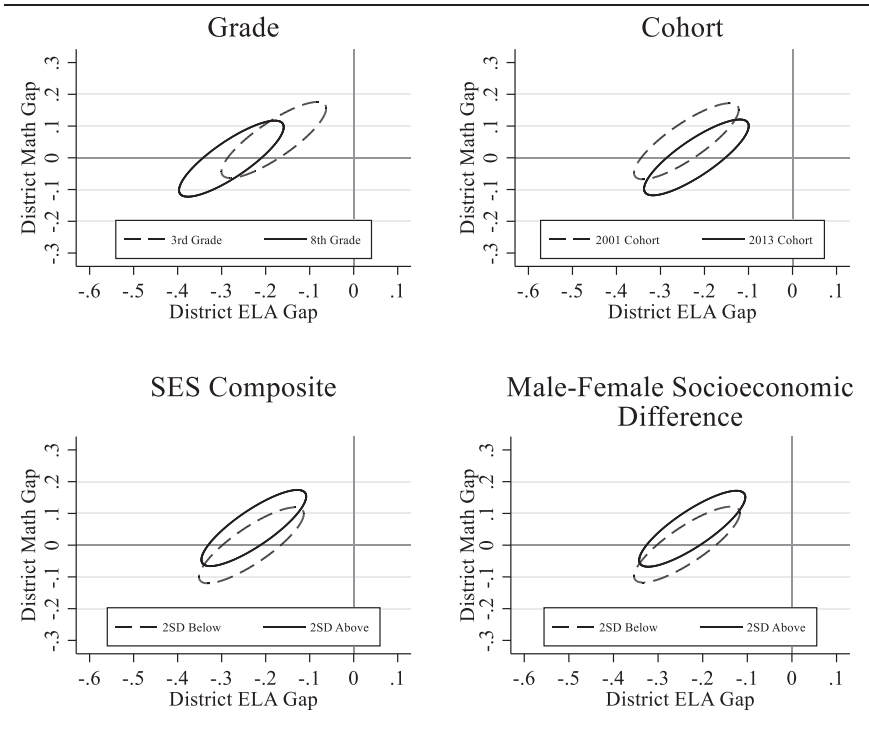


Figure 6. Stylized model results.

Note. ELA = English language arts; SES = socioeconomic status.

among White students, the pattern does not prove this. A similar pattern would result if White and Black gender gaps were similar within any given district but both White and Black gender gaps were negatively correlated with the proportion of Black students in a district. Nonetheless, the result does indicate that Black children are, on average, in school districts where gender gaps are more female favoring, while White children are disproportionately in school districts where gender gaps favor males more. These associations are not large, but they persist even after controlling for district socioeconomic characteristics (see Model 3 of Table 4).

Using the framework described in Figure 1, we can characterize where (in what kinds of districts) and when (in what grades/cohorts) gaps are more stereotypical versus gender favoring. In Figure 6, we plot the stylized results from Model 3, illustrating the joint distribution of math and ELA gaps in the same two-dimensional space as in Figures 1 and 3. These stylized results are derived from the estimates in Models 1 and 3 of Table 4. In each panel of the figure, we plot the estimated 95% coverage ellipses (the ellipses in which

95% of districts lie) for two sets of estimated gaps. These illustrate the relative direction and magnitude of the differences in gaps associated with different grades, cohorts, or types of districts.

In the first panel of the figure (upper left), we show that gender achievement gaps favor females more in later grades than earlier ones: The ellipse representing the joint distribution of the gaps shifts down and to the left (in the female-favoring direction) between third and eighth grade. A comparison of the implied distribution of gaps across cohorts (upper right panel) reveals that gender gaps are smaller and less stereotypical (math gaps are less male favoring, ELA gaps are less female favoring) in more recent cohorts, evidenced by the shift down and to the right of the ellipse. Finally, in districts with high average SES (lower left), or large adult male-female socioeconomic disparities (lower right), the distribution of gaps is shifted upward in the figure relative to poorer districts and those with smaller adult gender socioeconomic disparities, indicating that math gaps are more male favoring in more socioeconomically advantaged and unequal districts. Interestingly, the magnitudes of the differences in achievement gap patterns related to these two variables are much smaller than the magnitudes of the average differences in gaps across grades or cohorts.

In fitting Model 1 in Table 4, we found no statistically significant variation in the gap grade slopes among districts (motivating the removal of the random coefficients on the math and the ELA gap grade slopes from Model 1). This is somewhat surprising; we might anticipate that factors that produce variation in the gaps among districts would also have cumulative effects as student's age, which would be reflected as differences between districts in the gap grade slopes in our model. However, the test of between-district variance on this interaction term is a low-power test; there is likely a modest degree of variation among district that our models have insufficient power to detect. We do find that the district characteristics predict changes in the gender achievement gaps from third to eighth grade. In high-SES districts, math achievement gaps grow more female favoring than they do in lower SES districts (Model 2); however, this loses significance when controlling for student characteristics (Model 3) or estimating the association within states (Model 4). Furthermore, in communities with larger male-female socioeconomic disparities, the growth in the math and the ELA gaps is more male favoring (Models 2, 3, and 4).

The cohort slopes, in contrast, show that math gaps have changed in favor of females and ELA gaps in favor of males over the cohorts in our sample—in opposition to the common stereotypes—as illustrated in Figure 6. There is significant variance in the gap cohort slopes among districts and the change in the math is more male favoring, and ELA gaps over cohorts is more female favoring in high-SES communities (Model 3). In math, we also find that the change in gaps over cohorts is more male favoring in communities with high male-female socioeconomic disparities

but find no significant results in ELA (Model 3). These results hold when looking within states (Model 4).

Discussion

No prior study has examined gender achievement gaps at the local level and with the level of detail we have here. Given this, our primary goal in this article is to establish a set of stylized facts regarding the size, variation, and correlates of gender achievement gaps in math and ELA. Five particular patterns—and their implications—are worth noting.

First, in virtually every school district in the United States, female students outperformed male students on ELA tests in Grades 3 through 8 during the 2008–2009 to 2015–2016 school years. In the average district, the gap is roughly one quarter of a standard deviation, though the gaps vary among districts. A quarter of a standard deviation is a substantial gap; it corresponds to roughly two thirds of a grade level and is larger than the effects of most large-scale educational interventions. On math tests, in contrast, the gender gap in the average district is quite small—roughly 0.03 standard deviations in favor of males. Again, this varies among districts. Female students modestly outperform males in a quarter of districts; males modestly outperform females in the others. But in only a small number of districts are the gaps larger than a third of a grade level. The math and ELA gaps in the average district align with those estimated at the state or national level in other studies (Fryer & Levitt, 2010; Husain & Millimet, 2009; Lee et al., 2011; Penner & Paret, 2008; Pope & Sydnor, 2010; Robinson & Lubinski, 2011; Sohn, 2012).

Second, districts' math and ELA gender gaps are strongly positively correlated: School districts vary largely on the gender-favoring dimension and very little on the subject-specific stereotype dimension (as shown in Figures 1 and 3). The fact that gender gaps vary among school districts suggests that local conditions and processes—in addition to larger societal forces—play a role in shaping them. Moreover, the results imply that if the variation among districts is driven by, say, local norms, the content of these norms must be primarily about relative general academic expectations for male and female students, rather than about subject-specific differential expectations.

Third, the data are relatively silent with regard to what local processes produce these gaps—most of the variation among districts in gender achievement gaps is unaccounted for by socioeconomic and demographic district characteristics. In our models, gender gaps in math appear to be related to local socioeconomic conditions; many of the communities with the largest math achievement gaps are affluent, predominantly White, suburban communities in which adult gender employment and income disparities tend to be particularly large. This same pattern is not true of gender gaps in ELA performance, however.

Fourth, on average, gender achievement gaps become more female favoring from Grades 3 through 8 in both math and ELA. In third grade, male students outperform female students in math by roughly a sixth of a grade level and female students outperform male students by roughly half a grade level on ELA tests. By eighth grade, in the average district, male and female students score equally well on math tests, but females are nearly a grade level ahead of their male classmates in ELA. It is important to note that other studies, using national data, find that this pattern reverses in high school. On NAEP, for example, male students outperform female students in math in high school, and the ELA gap in favor of females is smaller in high school than in eighth grade (Fahle & Reardon, 2018). This suggests that the forces that shape gender achievement gap patterns vary during the child and the adolescent developmental period as well as among local communities.

Fifth, gender achievement gaps in Grades 3 through 8 have been trending toward gender parity over recent cohorts of students, though this trend is more pronounced in math than in ELA. Our estimates indicate that the math gap has declined by roughly 0.05 standard deviations (about a sixth of a grade level) from the 2001 to the 2013 spring kindergarten cohort. In the most recent cohorts, there is no gender gap in math in the average school district. The trend in ELA gaps has been much slower: The ELA gap declined, on average, by roughly 0.02 standard deviations across the same set of cohorts. The combination of these trends indicates that gender gaps in middle school achievement have become somewhat less pronounced and less aligned with subject-specific gender stereotypes in recent years.

It is not clear, however, what has driven these changes over time in the gender gaps. One possibility is that gender norms have changed in recent years, but there is little evidence to suggest that there has been a significant change in gender norms in the past decade. Another possibility is that the recession played a role. The recession generally lowered family incomes, and affected male workers more than female workers, thereby reducing the male-female income, educational, and occupational disparity in many communities. Given that lower community SES and smaller gender occupational disparities are both associated with less male-favoring math achievement gaps, it is possible that the recession led to reduced gender disparities in math. One might test this hypothesis, by examining whether gender gaps changed more in communities hardest hit by the recession, but such an analysis is beyond the scope of this article.

This article has several limitations. One is that the tests used to measure achievement vary in format and content among states, grades, and years, and these differences may lead to differences in measured gender achievement gaps. We address this issue by using the NAEP assessments to adjust the gender-specific scores on each states' tests. This method is not perfect, however, and may not fully correct for the differences in content and format of

states' tests. Another limitation is that we do not have good measures of local norms, expectations, stereotypes, or of how boys and girls are treated in school, at home, or in their community. Because of this, we cannot rule out the potentially important influence that these factors may have on gender achievement gaps that we are unable to observe with our coarse proxy measure. Third, the patterns described here apply to Grades 3 through 8. We cannot speak to the existence of gaps or trends in later grades and how they may differ from what we see here. Finally, we find some suggestive relationships between race and gender achievement gaps but are unable to estimate gender gaps by race. Prior evidence shows that gender achievement gaps are characteristically different among students from different racial groups (e.g., Penner & Paret, 2008), and we are limited in our ability to explore that using our data.

We have demonstrated a set of stylized patterns of gender inequality in U.S. school districts; however, there is clearly more work to do to understand the processes that shape these gender inequalities. The patterns that we observe are likely the cumulative effect of different types of phenomena operating at different geographic levels—large-scale social processes, local norms and beliefs, and personal interactions, applied simultaneously. Within the United States, subject-specific gender stereotypes are ubiquitous and persistent social constructs that, to some extent, shape individuals' gendered interests and structure individuals' actions and interactions (Ridgeway, 2011). At the same time, individuals' actions are also influenced by their local school and neighborhood contexts, as well as by their personal interactions (with family and peers). Understanding the different processes that generate gender inequality, the scales at which they operate, and how they interact with one another is an important area of future research.

Appendix

Constructing NAEP-Adjusted Residualized Achievement Measures

We can write the estimated mean test score for gender s in subject b , grade g , year y , and district d in state f as

$$\hat{\mu}_{sbgyd} = \mu_{sbgyd}^{naep} + v_{sbgyf} + \hat{\epsilon}_{sbgyd},$$

where μ_{sbgyd}^{naep} is the true mean score for that population on the NAEP test, v_{sbgyf} is an error term specific to gender s on the test in state f in subject-grade-year bgy (i.e., it is the difference in gender gaps as measured by the state test and the NAEP test in state f ; we assume here that this difference is constant across districts within state f , but not common across states); and $\hat{\epsilon}_{sbgyd}$ is sampling error specific to the estimated test score mean of gender s in subject b , grade g , year y , and district d . The estimated male-female gap in district d will then be

$$\begin{aligned}\hat{g}_{bgvd} &= [u_{[male]bgvd} - u_{[female]bgvd}] + [v_{[male]bgvf} - v_{[female]bgvf}] + [\hat{e}_{[male]bgvd} - \hat{e}_{[female]bgvd}] \\ &= [g_{bgvd}^{naep}] + [\Delta v_{bgvf}] + [\Delta \hat{e}_{bgvd}],\end{aligned}$$

where g_{bgvd}^{naep} is the true male-female gap that we would observe if the NAEP test in subject b , grade g , and year y were administered in district d ; Δv_{bgvf} is the difference between the gender gap on NAEP and the gender gap on the state f test in subject b , grade g , and year y ; and $\Delta \hat{e}_{bgvd}$ is measurement error in the gap.

If Δv_{bgvf} is not constant across the tests used in different states, grades, years, and subjects, the measured gaps will not be comparable across states, grades, years, and subjects. Prior research suggests that Δv_{bgvf} varies considerably among states (Reardon, Kalogrodes, et al., 2018). To address this, we residualize $\hat{\mu}_{sbgyd}$ by subtracting the average score of gender s in subject b , grade g , year y , and state f . We first estimate μ_{sbgyf} , the mean test score for students of gender s in subject b , grade g , year y , and state f , by taking a precision-weighted average of the $\hat{\mu}_{sbgyd}$'s. Then, we residualize $\hat{\mu}_{sbgyd}$ to obtain

$$\hat{\mu}_{sbgyd}^* = \hat{\mu}_{sbgyd} - \hat{\mu}_{sbgyf}.$$

Finally, we add the average NAEP score for gender s in subject b , grade g , year y , and state f to the corresponding residualized district means

$$\hat{\mu}_{sbgyd}' = \hat{\mu}_{sbgyd}^* + \hat{\mu}_{sbgyf}^{naep}.$$

NAEP does not report means in every grade and year, so we use interpolation to recover estimates of $\hat{\mu}_{sbgyf}^{naep}$ in nontested grades and years. We use NAEP data from 2009, 2011, 2013, and 2015 NAEP tests in Grades 4 and 8. We first standardize the gender-subject-grade-year estimates within subject, grades, and years using the standard deviation of achievement for all students in the same subject, grade, and year. We then interpolate to obtain estimated gender- and subject-specific means for the nontested grades and years.

Note that we can rewrite this as

$$\begin{aligned}\hat{\mu}_{sbgyd}' &= \hat{\mu}_{sbgyd}^* + \hat{\mu}_{sbgyf}^{naep} \\ &= \left[\mu_{sbgyd}^{naep} + v_{sbgyf} + \hat{e}_{sbgyd} \right] - E \left[\mu_{sbgyd}^{naep} + v_{sbgyf} + \hat{e}_{sbgyd} | s, b, g, y, f \right] + \hat{\mu}_{sbgyf}^{naep} \\ &= \left[\mu_{sbgyd}^{naep} + v_{sbgyf} + \hat{e}_{sbgyd} \right] - \mu_{sbgyf}^{naep} - v_{sbgyf} + \hat{\mu}_{sbgyf}^{naep} \\ &= \mu_{sbgyd}^{naep} + \hat{e}_{sbgyd}.\end{aligned}$$

Thus, $\hat{\mu}_{sbgyd}'$ is an estimate of the mean test scores of gender s in district d that has the bias due to the difference between the state's test and the NAEP removed. The $\hat{\mu}_{sbgyd}'$'s, and the resulting gap estimates, are therefore comparable across states within a grade, year, and subject.

Notes

The research described here was supported by grants from the Institute of Education Sciences (R305D110018 and R305B090016), the Spencer Foundation (Award 201500058), the William T. Grant Foundation (Award 186173), the Carnegie Corporation of New York (Grant G-F-17-54651), the Bill and Melinda Gates Foundation, the Overdeck Family Foundation, and the Russell Sage Foundation. We thank Richard DiSalvo, Jenny Buontempo, and Joseph Van Matre for excellent research assistance. Some of the data used in this article were provided by the National Center for Education Statistics (NCES). The opinions expressed here are ours and do not represent views of the Institute of Education Sciences or the U.S. Department of Education. The first two authors contributed equally to this article.

¹The negative correlation that they observe between the gender gaps in representation in the upper tails of states' math and reading distributions, however, is not evident in other measures of gender achievement gaps. We compute gender gaps between the means of the male and the female NAEP test score distributions in each state; the correlations between mean math and reading state gender gaps are generally positive and small (analyses not shown here). NAEP data can be retrieved here: <https://www.nationsreportcard.gov/ndecore/landing>.

²These stereotypes have remained relatively stable over the past 30 years despite the large cultural shifts in women's roles (Haines, Deaux, & Lofaro, 2016).

³Alternatively, sociologists have hypothesized that children in affluent families have the opportunity to indulge in their gendered interests, which may exacerbate gender differences (Charles, Harr, Cech, & Hendley, 2014). In other words, affluent students are able to pursue academic activities that align with their gendered interests (whereas poor students must make decisions more economically), such that gender differences in achievement will be magnified among affluent students.

⁴Gender differences in parental spending, however, may not all be stereotypical—they invest in different aspects of their children's education (unrelated to field). In particular, parents tend to have higher educational expectations for their daughters and invest more time in educational activities with their daughters but are more involved in school activities for their sons and save more for their sons' college education (Raley & Bianchi, 2006).

⁵The types of differential investments parents make in male and female children may also vary by social class (Hao & Yeung, 2015). Hao and Yeung find that although parents invest more in girls than in boys at all levels of SES, lower SES parents invest more in school-related (tuition, school supplies, tutoring) and socio-cultural expenditures (drawing, music, sports, community activities, toy/presents, vacations), whereas high-SES parents invest more in status-signaling expenditures (clothes, shoes, cars) for female children.

⁶Ideally, we would have assessed whether gender achievement gaps are related to a direct measure of stereotypes, such as a survey; however, these type of data are unavailable at the local level. As a proxy, researchers have used composites of differences between males and females in their economic participation and opportunity, educational attainment, health and survival, and political participation as a measure of collective attitudes. This is most often seen in the international literature (e.g., Guiso, Monte, & Sapienza, 2008; Hyde & Mertz, 2009). In so much as adult socioeconomic gender disparities signal that a community has more traditional gender stereotypes about occupational roles, we might expect them to also indicate a community that has academic gender stereotypes. However, the measure may also reflect a gender advantage. For example, large male-female disparities may indicate that males in a community have more opportunity, both economically and educationally.

⁷Note that we exclude state-grade-year-subject cases where (1) not all students in a state take a common test, (2) data were incomplete due to pilot testing, (3) the number of tests reported was more than 10% higher than the enrollment (typically because some students took multiple tests in a subject, such as in eighth grade math in some states), and (4) state testing participation rates were lower than 95%. In each of these state-grade-year-subject cases, we cannot meaningfully compare the test score distributions across gender and districts. For a list of the omitted cases, see Fahle et al. (2018; Table A1).

⁸For a complete description of the methodology used to estimate the district-subgroup-subject grade-year test score distributions, see Fahle et al. (2018).

⁹In addition to excluding a small number of state-grade-year-subject cases where state-level student participation rates were less than 95%, we exclude individual districts with lower than 95% participation rates. Such exclusions were rare except in 2015 and 2016, when some states and districts experienced high nonparticipation rates. We also do not report estimates for district-subject-grade-year-gender cells in which there are fewer than 20 students because estimated means are very imprecise in such cases. For a detailed description of these issues see Fahle et al. (2018). Finally, we restrict our analytic sample to district-grade-year-subject cases where both male and female test score mean estimates are available.

¹⁰Given that we have 8 years, 6 grades, 2 subjects, and 2 genders, the maximum possible number of estimates per district is 192.

¹¹We use the 2006–2010 EDGE data to construct the different measures in order to be consistent with the average SES measure taken from SEDA.

¹²The EDGE data from the ACS do not include separate tabulations of enrolled students' mothers' and fathers' socioeconomic characteristics (except for educational attainment and unemployment status). Therefore, we use data for all adults to construct the gender-specific socioeconomic composites.

¹³The demographic data provided in SEDA include imputed data and will not match the raw CCD. The multiple imputation process is described in detail in the SEDA technical documentation (Fahle et al., 2018).

¹⁴We further checked whether test format had implications for the variance of achievement by comparing the state-level standard deviation ratios (male/female) between the ED*Facts* data and the NAEP data by subject, grade, and year. The results were somewhat inconclusive because the state-level ratios of male/female standard deviations are imprecisely estimated in NAEP, due to modest sample sizes, and vary little among states. We found that the average male/female standard deviation ratio is very similar in size on NAEP and state tests (about 1.05). When we estimated the correlation between the NAEP and the state test standard deviation ratios (taking measurement error into account), the estimated correlation is high (0.87).

¹⁵Note that we use the sampling variance of $\hat{\mu}_{shgyl}^*$ (the residualized means) as an estimate of $\hat{\omega}_{shgyl}^2$ (the sampling variance of the NAEP-adjusted residualized means). We do not add in the sampling variance of the NAEP state mean because the state-level error is common to all districts in a state (and is similar across states, since the NAEP sample sizes are the same across states).

¹⁶The maps are shown to illustrate the variation; however, note that they understate the true variance of the gaps in ELA and math as a result of shrinkage. The reliabilities of the gaps are 0.60 in both subjects from Model 1.

References

- Ambady, N., Shih, M., Kim, A., & Pittinsky, T. L. (2001). Stereotype susceptibility in children: Effects of identity activation on quantitative performance. *Psychological Science*, 12, 385–390.
- Autor, D., Figlio, D., Karbownik, K., Roth, J., & Wasserman, M. (2016). School quality and the gender gap in educational achievement. *American Economic Review*, 106, 289–295.
- Autor, D., Figlio, D., Karbownik, K., Roth, J., & Wasserman, M. (2017). *Family disadvantage and the gender gap in behavioral and educational outcomes* (NBER Working Paper No. 22267). Retrieved from <https://www.nber.org/papers/w22267>
- Baker, M., & Milligan, K. (2016). Boy-girl differences in parental time investments: Evidence from three countries. *Journal of Human Capital*, 10, 399–441.

- Beller, M., & Gafni, N. (2000). Can item format (multiple choice vs. open-ended) account for gender differences in mathematics achievement? *Sex Roles*, 42, 1–21.
- Bertrand, M., & Pan, J. (2013). The trouble with boys: Social influences and the gender gap in disruptive behavior. *American Economic Journal: Applied Economics*, 5, 32–64.
- Bielinski, J., & Davison, M. L. (2001). A sex difference by item difficulty interaction in multiple-choice mathematics items administered to national probability samples. *Journal of Educational Measurement*, 38, 51–77.
- Buchmann, C., & DiPrete, T. A. (2006). The growing female advantage in college completion: The role of family background and academic achievement. *American Sociological Review*, 71, 515–541.
- Carter, J. S., & Borch, C. A. (2005). Assessing the effects of urbanism and regionalism on gender-role attitudes, 1974–1998. *Sociological Inquiry*, 75, 548–563.
- Cech, E. A. (2013). The self-expressive edge of occupational sex segregation. *American Journal of Sociology*, 119, 747–789.
- Charles, M., & Bradley, K. (2009). Indulging our gendered selves? Sex segregation by field of study in 44 countries. *American Journal of Sociology*, 114, 924–976.
- Charles, M., Harr, B., Cech, E., & Hendley, A. (2014). Who likes math where? Gender differences in eighth-graders' attitudes around the world. *International Studies in Sociology of Education*, 24, 85–112.
- Chatterji, M. (2006). Reading achievement gaps, correlates, and moderators of early reading achievement: Evidence from the Early Childhood Longitudinal Study (ECLS) kindergarten to first grade sample. *Journal of Educational Psychology*, 98, 489–507.
- Chetty, R., & Hendren, N. (2017). *The impacts of neighborhoods on intergenerational mobility I: Childhood exposure effects* (NBER Working Paper No. 23001). Retrieved from <https://www.nber.org/papers/w23001>
- Chetty, R., Hendren, N., Lin, F., Majerovitz, J., & Scuderi, B. (2016). Childhood environment and gender gaps in adulthood. *American Economic Review*, 106, 282–288.
- Cimpian, J. R., Lubienski, S. T., Timmer, J. D., Makowski, M. B., & Miller, E. K. (2016). Have gender gaps in math closed? Achievement, teacher perceptions, and learning behaviors across two ECLS-K cohorts. *AERA Open*, 2(4). doi:10.1177/2332858416673617
- Cox, D. (2003). Private transfers within the family: Mothers, fathers, sons and daughters. In A. H. Munnell & A. Sundren (Eds.), *Death and dollars: The role of gifts and bequests in America* (pp. 168–197). Washington, DC: Brookings Institution Press.
- Cvencek, D., Meltzoff, A. N., & Greenwald, A. G. (2011). Math-gender stereotypes in elementary school children. *Child Development*, 82, 766–779.
- DeMars, C. E. (1998). Gender differences in mathematics and science on a high school proficiency exam: The role of response format. *Applied Measurement in Education*, 11, 279–299.
- DiPrete, T. A., & Buchmann, C. (2013). *The rise of women: The growing gender gap in education and what it means for American schools*. New York, NY: Russell Sage Foundation.
- Dotti Sani, G. M., & Treas, J. (2016). Educational gradients in parents' child-care time across countries, 1965–2012. *Journal of Marriage and Family*, 78, 1083–1096.
- Duncan, G. J., & Murnane, R. J. (Eds.). (2011). *Whither opportunity? Rising inequality, schools, and children's life chances*. New York, NY: Russell Sage Foundation.

- Eccles, J., Jacobs, J. E., & Harold, R. D. (1990). Gender role stereotypes, expectancy effects, and parents' socialization of gender differences. *Journal of Social Issues*, 46, 183–201.
- Eccles, J., Wigfield, A., Harold, R. D., & Blumenfeld, P. (1993). Age and gender differences in children's self- and task perceptions during elementary school. *Child Development*, 64, 830–847.
- Entwisle, D. R., Alexander, K. L., & Olson, L. S. (1994). The gender gap in math: Its possible origins in neighborhood effects. *American Sociological Review*, 59, 822–838.
- Entwisle, D. R., Alexander, K. L., & Olson, L. S. (2007). Early schooling: The handicap of being poor and male. *Sociology of Education*, 80, 114–138.
- Fahle, E. M., & Reardon, S. F. (2018). Education. *Pathways Magazine, State of the Union: The Poverty and Inequality Report Special Issue*, 9–12.
- Fahle, E. M., Shear, B. R., Kalogrides, D., Reardon, S. F., Disalvo, R., & Ho, A. D. (2018). *Stanford education data archive: Technical documentation v2.1*. Retrieved from https://stacks.stanford.edu/file/druid:db586ns4974/SEDA_documentation_v21.pdf
- Fryer, R. G., & Levitt, S. D. (2010). An empirical analysis of the gender gap in mathematics. *American Economic Journal: Applied Economics*, 2, 210–240.
- Garner, M., & Engelhard, G. J. (1999). Gender differences in performance on multiple-choice and constructed response mathematics items. *Applied Measurement in Education*, 12, 29–51.
- Guiso, L., Monte, F., Sapienza, P., & Zingales, L. (2008). Culture, gender, and math. *Science*, 320, 1164–1165.
- Gunderson, E. A., Ramirez, G., Levine, S. C., & Beilock, S. L. (2012). The role of parents and teachers in the development of gender-related math attitudes. *Sex Roles*, 66, 153–166.
- Guryan, J., Hurst, E., & Kearney, M. (2008). Parental education and parental time with children. *Journal of Economic Perspectives*, 22, 23–46.
- Haines, E. L., Deaux, K., & Lofaro, N. (2016). The times they are a-changing . . . or are they not? A comparison of gender stereotypes, 1983–2014. *Psychology of Women Quarterly*, 40, 353–363.
- Hao, L., & Yeung, W.-J. J. (2015). Parental spending on school-age children: Structural stratification and parental expectation. *Demography*, 52, 835–860.
- Hopcroft, R. L. (2005). Parental status and differential investment in sons and daughters: Trivers-Willard revisited. *Social Forces*, 83, 1111–1136.
- Hopcroft, R. L., & Martin, D. O. (2016). Parental investments and educational outcomes: Trivers-Willard in the U.S. *Frontiers in Sociology*, 1, 1–12. Retrieved from <https://www.frontiersin.org/articles/10.3389/fsoc.2016.00003/full>
- Husain, M., & Millimet, D. L. (2009). The mythical “boy crisis”? *Economics of Education Review*, 28, 38–48.
- 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.
- 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, 8801–8807.
- Jacobs, J. E., Lanza, S., Osgood, D. W., Eccles, J. S., & Wigfield, A. (2002). Changes in children's self-competence and values: Gender and domain differences across grades one through twelve. *Child Development*, 73, 509–527.
- Kågesten, A., Gibbs, S., Blum, R. W., Moreau, C., Chandra-Mouli, V., Herbert, A., & Amin, A. (2016). Understanding factors that shape gender attitudes in early

- adolescence globally: A mixed-methods systematic review. *PLoS One*, 11(6), e0157805.
- Kornrich, S. (2016). Inequalities in parental spending on young children: 1972 to 2010. *AERA Open*, 2(2). doi:10.1177/2332858416644180
- Lee, J., Moon, S., & Hegar, R. L. (2011). Mathematics skills in early childhood: Exploring gender and ethnic patterns. *Child Indicators Research*, 4, 353–368.
- Legewie, J., & DiPrete, T. A. (2012). School context and the gender gap in educational achievement. *American Sociological Review*, 77, 463–485.
- Legewie, J., & DiPrete, T. A. (2014). The high school environment and the gender gap in science and engineering. *Sociology of Education*, 87, 259–280.
- Lindberg, S. M., Hyde, J. S., Petersen, J. L., & Linn, M. C. (2010). New trends in gender and mathematics performance: A meta-analysis. *Psychological Bulletin*, 136, 1123–1135.
- Lubienski, S., Robinson, J., Crane, C., & Ganley, C. (2013). Girls' and boys' mathematics achievement, affect, and experiences: Findings from the ECLS-K. *Journal for Research in Mathematics Education*, 44, 634–645.
- Penner, A. M., & Paret, M. (2008). Gender differences in mathematics achievement: Exploring the early grades and the extremes. *Social Science Research*, 37, 239–253.
- Pope, D. G., & Sydnor, J. R. (2010). Geographic variation in the gender differences in test scores. *Journal of Economic Perspectives*, 24, 95–108.
- Raley, S., & Bianchi, S. (2006). Sons, daughters, and family processes: Does gender of children matter? *Annual Review of Sociology*, 32, 401–421.
- Ramey, G., & Ramey, V. A. (2010). The rug rat race. *Brookings Papers on Economic Activity*, 2010, 129–176.
- Reardon, S. F., Ho, A. D., Shear, B. R., Fahle, E. M., Kalogrides, D., & Disalvo, R. (2018). *Stanford education data archive* (Version 2.1). Retrieved from <http://purl.stanford.edu/db586ns4974>
- Reardon, S. F., Kalogrides, D., Fahle, E. M., Podolsky, A., & Zárate, R. C. (2018). The relationship between test item format and gender achievement gaps on math and ELA tests in fourth and eighth grades. *Educational Researcher*, 47, 284–294.
- Reardon, S. F., Kalogrides, D., & Ho, A. D. (2017). *Linking U.S. school district test score distributions to a common scale* (CEPA Working Paper No. 16-09). Retrieved from <https://cepa.stanford.edu/sites/default/files/wp16-09-v201706.pdf>
- Reardon, S. F., Shear, B. R., Castellano, K. E., & Ho, A. D. (2017). Using heteroskedastic ordered probit models to recover moments of continuous test score distributions from coarsened data. *Journal of Educational and Behavioral Statistics*, 42, 3–45.
- Ridgeway, C. L. (2011). *Framed by gender: How gender inequality persists in the modern world*. New York, NY: Oxford University Press.
- Robinson, J. P., & Lubienski, S. T. (2011). The development of gender achievement gaps in mathematics and reading during elementary and middle school: Examining direct cognitive assessments and teacher ratings. *American Educational Research Journal*, 48, 268–302.
- Sohn, K. (2012). A new insight into the gender gap in math. *Bulletin of Economic Research*, 64, 135–155.
- Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology*, 35, 4–28.
- Tomasetto, C., Alparone, F. R., & Cadinu, M. (2011). Girls' math performance under stereotype threat: The moderating role of mothers' gender stereotypes. *Developmental Psychology*, 47, 943–949.

- Trivers, R. L., & Willard, D. E. (1973). Natural selection of parental ability to vary the sex ratio of offspring. *Science*, 179, 90–92.
- Upadyaya, K., & Eccles, J. (2015). Do teachers' perceptions of children's math and reading related ability and effort predict children's self-concept of ability in math and reading? *Educational Psychology*, 35, 110–127.
- Witt, S. D. (1997). Parental influence on children's socialization to gender roles. *Adolescence*, 32, 253–259.

Manuscript received August 9, 2018

Final revision received November 20, 2018

Accepted March 16, 2019