FEATURE ARTICLES

Science Achievement Gaps by Gender and Race/Ethnicity in Elementary and Middle School: Trends and Predictors

David M. Quinn[1] and North Cooc[2]

Research on science achievement disparities by gender and race/ethnicity often neglects the beginning of the pipeline in the early grades. We address this limitation using nationally representative data following students from Grades 3 to 8. We find that the Black–White science test score gap (–1.07 *SD* in Grade 3) remains stable over these years, the Hispanic–White gap narrows (–.85 to –.65 *SD*), and the Asian–White Grade 3 gap (–.31 *SD*) closes by Grade 8. The female–male Grade 3 gap (–.23 *SD*) may narrow slightly by eighth grade. Accounting for prior math and reading achievement, socioeconomic status, and classroom fixed effects, Grade 8 racial/ethnic gaps are not statistically significant. The Grade 8 science gender gap disappears after controlling for prior math achievement.

Keywords: achievement gap; descriptive analysis; disparities; econometric analysis; elementary schools; middle schools; regression analyses; science education

Since the 1950s, leaders in education, science, politics, and business have stressed the need for "scientific literacy" among the U.S. general public (DeBoer, 2000). Today, concern over scientific literacy is growing due to the increasing demand for graduates entering careers in science, technology, engineering, and math (STEM) (Association of American Universities, 2006). In the coming decades, science occupations are predicted to grow faster than the average rate for all fields (Lacey & Wright, 2009), and a significant amount of science and math training will be required for 9 of the 10 fastest growing occupations requiring a bachelor's degree or higher (Wang, 2013). Statistics such as these led the President's Council of Advisors on Science and Technology (2012) to call for a 33% increase in the number of STEM bachelor's degrees annually.

Developing a more scientifically literate citizenry will require confronting the persistent gender and racial/ethnic gaps in science proficiency (National Center for Education Statistics [NCES], 2012a). These gaps have important implications for economic and technological advancement, as well as for social equity. Economists point out that meeting the increasing demand for science graduates will require that a greater share of females and racial/ethnic minorities are interested in, and qualified for, science careers (Muller, Stage, & Kinzie, 2001). Sociologists emphasize that disparities in scientific understanding and science achievement exacerbate social stratification in today's high-tech global economy (Drew, 2011; Muller et al., 2001) and that science achievement gaps foreshadow employment barriers for individuals seeking to enter fields offering high pay and prestige (Wang, 2013).

Despite the importance of monitoring and closing gender and racial/ethnic gaps in science achievement, researchers and policymakers tend to focus on math and literacy gaps. Studies that do examine science gaps often concern postsecondary education or high school, neglecting students' early foundational experiences. There is less research on how science gaps develop as students progress through elementary and middle school and the extent to which individual background characteristics and contextual factors explain these gaps. Understanding when gaps emerge and what factors predict them are critical first steps toward developing appropriate education interventions and policies. In this article, we contribute to the literature by analyzing science gaps in elementary and middle school using nationally

[1]Harvard University, Cambridge, MA
[2]The University of Texas at Austin, Austin, TX

Educational Researcher, Vol. 44 No. 6, pp. 336–346
DOI: 10.3102/0013189X15598539

representative data from the *Early Childhood Longitudinal Study, Kindergarten Class of 1998-99* (*ECLS-K:99*).[1]

Background

Why Study Science Test Score Gaps?

In the United States, White males have traditionally dominated STEM occupations (Riegle-Crumb & King, 2010) (although our interest is in the sciences specifically, we discuss STEM broadly because STEM fields are often analyzed collectively in this literature). Researchers studying the "STEM pipeline" into these occupations have found that females are less likely than males to declare a STEM major in college, and little progress in this area has been made over the past 50 years (Riegle-Crumb & King, 2010). Although racial minorities are no less likely than White students to declare a STEM major, racial/ethnic gaps appear in STEM persistence during college (Chen & Weko, 2009; Griffith, 2010). For example, in 2001, 15.5% of Black students and 16.3% of Hispanic students who had entered a bachelor's program in STEM obtained a bachelor's degree in STEM, compared to 29.5% for White students and 31.2% for Asian students (Chen & Weko, 2009). Gender gaps in STEM persistence appear in some datasets but not others (Chen & Weko, 2009; Griffith, 2010).

Although a student's decision to pursue a career in the sciences is influenced by many factors—including cultural norms and stereotypes related to the field and the student's perceived value of scientific occupations (Eccles, 2007)—preparation and achievement in science play key roles (Wang, 2013). Studies show that high school students' science background, achievement, and attitudes predict whether they will choose a STEM major in college (Riegle-Crumb, King, Grodsky, & Muller, 2012; Tai, Liu, Maltese, & Fan, 2006; Wang, 2013), and high school achievement and STEM preparation explain large portions of the racial/ethnic gaps in STEM major persistence in college (Griffith, 2010; Price, 2010). However, much less is known about the trends and predictors of science achievement gaps before high school, when the "leaky pipeline" may begin.

Documenting and Explaining Gaps

Most of what we know about science gaps in elementary and middle school comes from the National Assessment of Educational Progress (NAEP). The NAEP–Long Term Trend (NAEP-LTT), which is designed for examining achievement trends over time, included a science assessment from 1969 and 1999. Over those 30 years, there was little consistent progress in closing racial/ethnic or gender science gaps (Campbell, Hombo, & Mazzeo, 2000). Today, national estimates of science test score gaps come from the Main NAEP, which differs from the NAEP-LTT in that its content is subject to change over time. Results from the 2011 Main NAEP (the most recent for which data are available) showed large eighth grade Black–White (–1.03 *SD*) and Hispanic–White (–.79 *SD*) science gaps and somewhat smaller Asian/Pacific Islander–White (–.10 *SD*) and female–male (–.15 *SD*) gaps (authors' calculations based on data from U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, n.d.).

When it comes to documenting and explaining test score gaps, however, NAEP data are limited in two important ways. First, the NAEP does not follow the same students over time, preventing an examination of how gaps change as students progress through elementary and middle school. Second, the NAEP does not link students to their past (or concurrent) achievement in other subjects, preventing an examination of how students' foundational math and reading skills may affect their later science learning. The *ECLS-K:99* is the only nationally representative study without these limitations.

Explaining racial/ethnic gaps. Two of the most studied explanations for racial/ethnic gaps in student achievement are racial/ethnic differences in socioeconomic status (SES) and school quality. Students from higher SES backgrounds tend to outperform their lower SES peers for several reasons. For instance, financial resources enable parents to access—among other things—stimulating learning materials and environments (Entwisle, Alexander, & Olson, 2000) and high-quality health care, which promotes cognitive development in utero and in early childhood (Currie, 2005). Educated parents also tend to use more complex language with their children, a behavior that predicts children's later academic success (Hart & Risley, 2003).

As for school quality, Black and Hispanic students have less access (compared with White students) to school resources that promote science achievement (Jacob, 2007). They are less likely to be taught by qualified science teachers, are less likely to have important science lab facilities and equipment, and tend to be exposed to less rigorous curricula (Banilower et al., 2013; Oakes, Ormseth, Bell, & Camp, 1990; Ruby, 2006). Additionally, their teachers tend to place less emphasis on scientific inquiry and problem-solving and are less likely to use techniques that promote active student involvement (Oakes et al., 1990). This can be detrimental because hands-on learning and group activities inspire students' interest in science (Oakes et al., 1990) and may promote greater learning gains (NCES, 2012a; Stohr-Hunt, 1996). Unequal access to these resources and experiences often occurs at the school level through de facto segregation (Hanushek & Rivkin, 2006) and at the classroom level due to inequitable student assignment practices and curricular tracking (Lankford, Loeb, & Wyckoff, 2002; Oakes et al., 1990). This suggests that racial/ethnic science gaps among students from the same schools or the same classrooms may be narrower than gaps in the general population.

Students' science achievement is also partly a function of their math and reading skills (Snow, 2010). Science classes often require that students perform mathematical calculations and read complex texts with academic language, high information density, and abstract ideas (Fang, 2011; Snow & Uccelli, 2008). Thus, students with lower math and reading skills will be at a disadvantage in science class. Because racial/ethnic disparities in math and reading develop early on, also due in part to SES and school quality disparities (e.g., Quinn, 2015), students' prior achievement in these subjects may help explain racial/ethnic science test score gaps.

Explaining gender gaps. Unlike racial/ethnic test score gaps, gender gaps are unlikely to be explained by SES or school quality because SES and school quality do not differ systematically by gender. Gender gaps in science are more likely to be explained by cultural forces such as societal norms and stereotypes, which affect the encouragement that girls receive to pursue science and the messages they internalize about their potential in the field (Hill, Corbett, & Rose, 2010). Because such forces also negatively affect girls' math achievement, prior math test scores are more likely to explain science gender gaps than are SES and school quality. Most of the research on explaining gender gaps in science has focused on explaining gaps in science course-taking or degree pursuit, however; less has been done on explaining gaps in science test scores (Eccles, 2007; Griffith, 2010; Ost, 2010; Price, 2010; Riegle-Crumb & Moore, 2014; Riegle-Crumb et al., 2012; Riegle-Crumb & King, 2010).

Summary and Research Questions

Developing scientific literacy is an important goal for all students, and disparities in science achievement by gender and race/ethnicity are indicators of educational inequity that can foreshadow other inequities in adulthood. Research has shown that high school students' achievement in STEM predicts whether they will enter and persist in a STEM field in college, and high school STEM achievement helps to explain racial gaps in STEM persistence in college. This raises questions about how young students' early experiences lay the foundation for these inequalities, yet little is known about how science gaps may develop as a cohort of students progresses through elementary and middle school, or the extent to which early science gaps may be explained by individual factors (such as student SES and prior math and reading achievement) and contextual factors that vary between school or within school. In this study, we address these limitations of the literature by using nationally representative longitudinal data to answer the following research questions:

Research Question 1 (RQ1): (a) What are the science test score gaps by gender and race/ethnicity in Grades 3, 5, and 8? (b) Do these gaps change from Grade 3 to Grade 8?

Research Question 2 (RQ2): To what extent do (a) individual factors and (b) schools, teachers, and classrooms explain eighth grade science test score gaps?

Methods

Data Source and Measures

Data. We use data from the NCES's *ECLS-K:99*, a study of U.S. school children's academic and social development. The *ECLS-K:99* drew a nationally representative sample of 21,409 kindergarteners from nearly 1,000 schools during the 1998–1999 school year and followed these students through eighth grade. Data collection included interviews with parents; surveys of principals, teachers, and students; and direct cognitive assessments of students in math and reading. In Grades 3, 5, and 8, students also took science assessments (for more information on the *ECLS-K:99*, see Tourangeau, Nord, Lê, Sorongon, & Najarian, 2009).

Dependent variable: Science achievement. Our outcome measure is students' scores on the *ECLS-K:99* standardized, multiple-choice science test. Assessment items, which were based on other large-scale studies such as the NAEP, covered three main areas: Earth and space science (Earth's structure and systems, Earth's place in the universe), physical science (matter, energy, and their transformations; the motion of light, sound, and physical objects), and life science (cells and their functions, organisms, diversity, and ecology) (Najarian, Pollack, & Sorongon, 2009). In order to avoid floor and ceiling effects, students took a set of routing items that determined the difficulty of the questions they would be asked. We use a standardized version of the theta test score metric, which was derived from a three-parameter item response theory model. Theta reliabilities were .88, .87, and .84 for Grades 3, 5, and 8, respectively. For more information on these assessments, see Najarian et al. (2009).[2]

Question predictors. Our key predictors include a vector of gender and race/ethnicity indicator variables. Using non-Hispanic Whites as the omitted group, the race/ethnicity vector includes indicators for non-Hispanic Black, Hispanic (race specified or unspecified), Asian, and "other race" (due to the small sample sizes for Native Hawaiian/Pacific Islander, American Indian/Alaska Native, and multiracial, we combined these groups into a single category).

Controls. We use two sets of control variables to explain gaps in eighth grade science achievement. The first is a set of student background characteristics, including SES and prior achievement. SES is operationalized as the average of three standardized components (M = 0, SD = 1): log household income, parental education, and parental occupational prestige (see Tourangeau et al., 2009, for more information). As prior achievement measures, we include students' scores on the fifth grade *ECLS-K:99* math and reading assessments (standardized to M = 0 and SD = 1; see Najarian et al., 2009, for more information on these assessments).[3]

The second set of controls includes vectors of indicator variables for schools, teachers, or classrooms (each of which is included in a separate model). These fixed effects allow us to estimate the average gap sizes among students who attended the same school, were taught by the same science teacher, or were assigned to the same science classroom, respectively. In other words, the fixed effects models allow us to control for all observable and unobservable school-, teacher-, or classroom-level variables that might affect students' science scores. If racial/ethnic science gaps are reduced when school fixed effects are added to the model, for example, this would suggest that science gaps may be related to differences in the schools attended by students from different racial/ethnic groups. Although this is a logical first step for investigating the roles that broader school, teacher, and classroom characteristics may play in explaining science gaps, we note that these models cannot identify causal effects of schools, teachers, or classrooms.

Analytic samples. We employ two different analytic samples to answer our research questions. To answer RQ1, we use data from all students with nonmissing science scores in Grades 3, 5, and 8 (n = 8,721). Other sample restrictions applied to RQ2. In Grade 8, approximately half of the sample had information collected about their science teacher, whereas the other half had information collected about their math teacher. Consequently, half of the students in the sample cannot be linked to an eighth grade science teacher and therefore cannot be included in the teacher or classroom fixed effects models. To stabilize the sample across models for RQ2, in all models we include only students who can be linked to a science teacher and who have nonmissing outcome and control data (n = 3,853). Results from models for RQ1 that apply the same sample restrictions as RQ2 and results from models for RQ2 that use all available data are similar to the results reported here (see online Appendix A, available on the journal website, for these results as well as results from models that use multiple imputation).

Analytic Strategy

To address our first research question, we fit the following general model (with separate models for each wave of data):

$$Y_i = \frac{Sci_i - \overline{Sci}}{SD} = \beta_0 + \beta_1 Female_i + \beta_2 Black_i + \beta_3 Hispanic_i + \beta_4 Asian_i + \beta_5 OtherRace_i + \epsilon_i \quad (1)$$

The outcome is student i's science theta score (where scores are standardized to a mean of 0 and a standard deviation of 1 at each test wave). In Model 1, β_1 represents the standardized mean gender gap in science (controlling for race/ethnicity). Coefficients β_2 through β_5 represent the standardized mean gaps between the named racial/ethnic group and non-Hispanic Whites (controlling for gender). To estimate gap changes, we subtract the Grade 3 gap for a particular group from the Grade 8 gap for that group (for additional methodological details, including the formula for the standard error of the gap change, see online Appendix B, available on the journal website).

To address our second research question, we fit taxonomies of regression models by adding to Model 1 the set of individual-level covariates and school-, teacher-, or classroom fixed effects. In these models, we use Grade 8 science scores as the outcome because this is the end of the elementary/middle school science pipeline. In all models, we incorporate the appropriate longitudinal sample weights and adjust the standard errors to account for the complex sampling design used in the *ECLS-K: 99*.[4]

Results

Descriptive Statistics

In Table 1, we present descriptive statistics for the Grade 8 analytic sample used in RQ2 by gender and race/ethnicity. For all groups, the variance of science theta scores is larger in the eighth grade than in the third or fifth grade (see Appendix B, available on the journal website, for a discussion of how this influences gap change estimates).

General Trends

In Table 2, we present results from regression models examining how gender and racial/ethnic gaps in science achievement change as students progress through school. In Figure 1, we present these results graphically.

In Grade 3 (column 1 of Table 2), we found that females scored approximately .23 *SD* lower than males. Compared to White students, Black students scored 1.07 *SD* lower, Hispanic students scored .85 *SD* lower, and Asian students scored .31 *SD* lower.

As seen in column 4, the gender gap in eighth grade was slightly narrower than it was in Grade 3 (by .04 *SD*, $p < .10$). The Hispanic–White gap was .20 *SD* narrower in Grade 8, and most of this narrowing occurred between Grades 3 and 5. The Asian–White gap closed by Grade 8, and most of this closure occurred between Grades 5 and 8. The Black–White gap did not change significantly from Grades 3 to 8.[5]

In models not displayed in Table 2, we tested for interactions between gender and the race/ethnicity variables at each time point; these interactions were not significant individually or as a group.

Explaining Gaps

SES and prior achievement. In Table 3, we examined the extent to which science gaps in Grade 8 could be explained by student SES and prior achievement. Column 1 shows the unadjusted science gaps as a reference (note that estimates differ slightly from those in Table 2 due to the sample restrictions described earlier). As expected, adding the SES composite in column 2 had little effect on the gender gap (–.16 to –.19), but the Black–White gap was reduced by almost 30% (from –1.16 to –.80) and the Hispanic–White gap was reduced by over half (–.66 to –.31). Conditional on SES, Asian students scored .17 *SD* higher than White students, on average.

Column 3 of Table 3 displays Grade 8 science gaps conditional on Grade 5 math achievement. With an R^2 of .60, this model demonstrates that prior math achievement explained more variation in science scores than did SES. Prior math achievement explained all of the gender gap in science scores (adjusted gap = –.03 *SD, ns*). The Black–White gap was substantially reduced after controlling for prior math scores, to –.51 *SD*. The Asian–White gap, which significantly favored Asian students when controlling for SES, reversed direction to significantly favor White students (–.14 *SD*) when controlling for prior math achievement.

As seen in column 4 of Table 3, reading skills apparently influence science gaps differently from how math skills influence science gaps. Most strikingly, the gender gap conditional on prior reading achievement (–.24 *SD*) was larger than the unconditional gender gap in science. Prior reading achievement explained slightly less of the Black–White science gap than did prior math achievement (–.56 *SD*) but explained more of the Hispanic–White gap (–.27 *SD*) than did prior math achievement. Although the Asian–White science gap conditional on math scores was significant favoring White students, the Asian–White gap conditional on reading scores was significant favoring Asian students (.15 *SD*).

In column 5, we condition on both prior math and reading scores. Here, the adjusted gender gap (–.14 *SD*) is similar to the unadjusted gender gap. Controlling for both math and reading

Table 1
Weighted Descriptive Statistics by Race/Ethnicity and Gender for Grade 8 Analytic Sample

	White			Hispanic			Black			Asian			Female			Male		
	M	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*
Female	0.48	0.5	2,487	0.53	0.5	618	0.45	0.5	355	0.56	0.5	200						
Sci 3 Th.	−0.35	0.6	2,462	−0.97	0.63	603	−1.09	0.62	344	−0.5	0.69	193	−0.69	0.68	1,905	−0.52	0.69	1,886
Sci 5 Th.	0.21	0.59	2,487	−0.27	0.65	618	−0.52	0.63	355	0.1	0.63	200	−0.1	0.65	1,929	0.06	0.69	1,924
Sci 8 Th.	1.17	0.77	2,487	0.61	0.77	618	0.2	0.73	355	1.24	0.77	200	0.82	0.82	1,929	0.95	0.88	1,924
Sci 3 Th. std.	0.4	0.9	2,462	−0.52	0.94	603	−0.71	0.93	344	0.18	1.03	193	−0.11	1.02	1,905	0.15	1.03	1,886
Sci 5 Th. std.	0.33	0.89	2,487	−0.4	0.98	618	−0.77	0.96	355	0.16	0.95	200	−0.14	0.98	1,929	0.1	1.04	1,924
Sci 8 Th. std.	0.22	0.92	2,487	−0.44	0.91	618	−0.93	0.87	355	0.3	0.91	200	−0.19	0.98	1,929	−0.04	1.05	1,924
Math 5 std.	0.23	0.95	2,487	−0.24	0.92	618	−0.71	0.91	355	0.54	0.97	200	−0.11	0.97	1,929	0.07	1.04	1,924
Reading 5 std.	0.26	0.98	2,487	−0.32	0.93	618	−0.63	0.96	355	0.18	0.86	200	0.04	0.95	1,929	−0.08	1.1	1,924
SES	0.12	0.78	2,487	−0.54	0.7	618	−0.56	0.68	355	−0.03	0.89	200	−0.11	0.83	1,929	−0.16	0.8	1,924

Note. Sci 3 Th. = Grade 3 science score in theta metric; Sci 3 std. = Grade 3 science theta, standardized; Math 5 std. = Grade 5 math score, standardized; Reading 5 std. = Grade 5 reading score, standardized; SES = composite of parental education, income, and occupational prestige.

Table 2
Gender and Racial/Ethnic Science Test Score Gaps in Grades 3, 5, and 8 and Changes in Gaps (Grade 8–Grade 3)

	(1) Grade 3 b (*SE*)	(2) Grade 5 b (*SE*)	(3) Grade 8 b (*SE*)	(4) Grade 8–Grade 3 Δ b (*SE*)
Female	–0.226***	–0.246***	–0.185***	0.041†
	(0.032)	(0.035)	(0.034)	(0.022)
Black	–1.073***	–1.121***	–1.103***	–0.030
	(0.087)	(0.073)	(0.071)	(0.055)
Hispanic	–0.852***	–0.693***	–0.653***	0.199***
	(0.050)	(0.054)	(0.050)	(0.034)
Asian	–0.306***	–0.238**	0.041	0.347***
	(0.085)	(0.089)	(0.075)	(0.055)
Other race	–0.436*	–0.462**	–0.442*	–0.006
	(0.185)	(0.176)	(0.181)	(0.123)
Constant	0.475***	0.428***	0.281***	
	(0.033)	(0.033)	(0.036)	
N	8,721	8,721	8,721	8,721
R^2	0.213	0.200	0.184	

Note. Test scores are standardized to $M = 0$ and $SD = 1$ at each test wave. Standard errors that account for the *ECLS-K:99*'s complex sampling design are in parentheses. Other race = student is Native American/Alaska Native, Native Hawaiian/Pacific Islander, or multiracial. Estimates in the third column differ slightly from those in the first column of Table 3 due to differences in sample restrictions (see text). Unweighted sample sizes: female = 4,370 (male = 4,351), Black = 880; Hispanic = 1,464; Asian = 478; other race = 460 (White = 5,439).
†$p < .10$. *$p < .05$. **$p < .01$. ***$p < .001$.

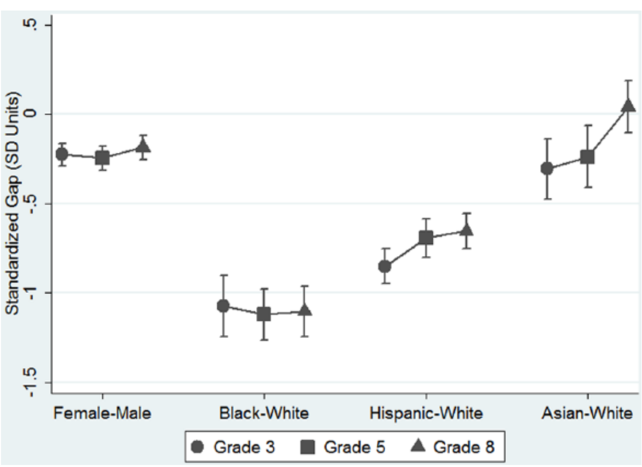

FIGURE 1. *Standardized science test score gaps in Grades 3, 5, and 8*
Test scores are standardized to $M = 0$ and $SD = 1$ at each test wave. Bands represent 95% confidence intervals. Grade 8 Hispanic–White and Asian–White gaps are significantly different from respective Grade 3 gaps. Unweighted sample sizes: female = 4,370 (male = 4,351);, Black = 880; Hispanic = 1,464; Asian = 478 (White = 5,439).

yields conditional Black–White and Hispanic–White science gaps that are slightly narrower than the gaps seen when conditioning on either subject alone. In column 6, we add SES to the model, which explained very little additional outcome variation above and beyond prior math and reading and only slightly reduced the Black–White and Hispanic–White gaps.

Fixed effects. In Table 4, we examined the extent to which gaps persisted after controlling for school, science teacher, or science classroom. Columns 1 through 3 control for fixed effects of school, teacher, and classroom, respectively; columns 4 through 6 repeat these models while also controlling for the full set of individual control variables. With an adjusted R^2 of .50, column 1 shows that schools explain a substantial amount of variation in science scores (approximately 30 percentage points more than race/ethnicity and gender alone). As expected, the gender science gap within schools was similar to the gender gap overall (–.18 *SD*). The magnitude of the mean within-school Black–White gap (–.71 *SD*) was approximately 60% of the overall Black–White gap and the within-school Hispanic–White gap (–.22 *SD*) was not statistically significant (though note the small sample size).[6] Controlling for science teacher and classroom fixed effects (columns 2 and 3, respectively), the Hispanic–White gap remained nonsignificant. Teacher fixed effects had only a small effect on the Black–White gap, but classroom fixed effects reduced this gap to –.51. Teacher and classroom fixed effects explained much of the variation in science scores, with an adjusted R^2 of .63 for the teacher fixed effects model and an adjusted R^2 of .70 for the classroom fixed effects model. As discussed below, some of this explanatory power is likely due to the intentional assignment of students to achievement tracks.

In columns 4 through 6 of Table 4, we present school, teacher, and classroom fixed effects models that also adjust for the individual control variables. With classroom fixed effects and all individual controls (column 6), both Black and Hispanic students' scores were statistically equivalent to those of White students. The gender gap was also not statistically significant, though the significance is sensitive to modeling strategy.[7]

Table 3
Gender and Racial/Ethnic Science Test Score Gaps in the Eighth Grade, With and Without Individual Controls

	(1) b (*SE*)	(2) b (*SE*)	(3) b (*SE*)	(4) b (*SE*)	(5) b (*SE*)	(6) b (*SE*)
Female	–0.162***	–0.186***	–0.032	–0.240***	–0.138***	–0.143***
	(0.035)	(0.032)	(0.032)	(0.029)	(0.031)	(0.030)
Black	–1.157***	–0.795***	–0.509***	–0.564***	–0.437***	–0.393***
	(0.071)	(0.060)	(0.064)	(0.049)	(0.053)	(0.051)
Hispanic	–0.661***	–0.308***	–0.344***	–0.266***	–0.245***	–0.185***
	(0.055)	(0.048)	(0.051)	(0.040)	(0.040)	(0.039)
Asian	0.085	0.168*	–0.135*	0.146*	0.000	0.023
	(0.092)	(0.084)	(0.061)	(0.060)	(0.053)	(0.053)
Other race	–0.459*	–0.299**	–0.196*	–0.146**	–0.123**	–0.106*
	(0.185)	(0.106)	(0.085)	(0.047)	(0.047)	(0.044)
SES		0.533***				0.134***
		(0.034)				(0.023)
Grade 5 math			0.684***		0.378***	0.362***
			(0.022)		(0.024)	(0.024)
Grade 5 reading				0.675***	0.411***	0.374***
				(0.018)	(0.022)	(0.021)
Constant	0.303***	0.251***	0.080**	0.168***	0.097***	0.097***
	(0.035)	(0.028)	(0.027)	(0.023)	(0.024)	(0.023)
N	3,853	3,853	3,853	3,853	3,853	3,853
R^2	0.207	0.364	0.604	0.619	0.678	0.685

Note. Test scores are standardized to $M = 0$ and $SD = 1$. Standard errors that account for the *ECLS-K:99*'s complex sampling design are in parentheses. Other race = student is American Indian/Alaska Native, Native Hawaiian/Pacific Islander, or multiracial.
*$p < .05$. **$p < .01$. ***$p < .001$.

Column 6 shows that a model with classroom fixed effects and all individual controls explains most of the variation in science achievement, with an adjusted R^2 of .83.[8]

Discussion

In this study, we sought to (a) describe national trends in science test score gaps by gender and race/ethnicity from Grade 3 to Grade 8 and (b) determine the extent to which Grade 8 science gaps could be explained by SES, prior achievement, and differences in school-, teacher-, and classroom-level factors. In Grade 3, we found large Black–White and Hispanic–White science test score gaps and smaller Asian–White and gender gaps. Although the Black–White gap remained fairly constant over time (Grade 8: –1.10 *SD*, or equivalent to the distance between the 50th percentile of a normal distribution and the 14th percentile), the Hispanic–White gap narrowed (Grade 8: –.65 *SD*, or the distance between the 50th percentile of a normal distribution and the 26th) and the Asian–White gap had disappeared by eighth grade. The narrowing of the gender gap from Grade 3 to Grade 8 was small and marginally significant (Grade 8: –.19 *SD*, or 50th percentile of a normal distribution to 42nd).

These unadjusted racial/ethnic gaps in science scores were somewhat larger than their corresponding math and reading gaps. The Grade 8 Black–White science gap was approximately .25 *SD* wider than the Black–White math (–.84 *SD*) and reading (–.85 *SD*) gaps, and the Hispanic–White Grade 8 science gap was larger than the Hispanic–White math (–.46 *SD*) and reading (–.59 *SD*) gaps (authors' calculations from *ECLS-K:99*). For eighth graders, the gender gap in science achievement was slightly wider than the gender gap in math (math gap: –.12 *SD*); the Grade 8 reading gap favors females by .21 *SD* (Robinson & Lubienski, 2011).

With our parsimonious set of controls, all eighth grade science gaps were greatly reduced and no longer statistically significant. The gender gap in science appears to be closely related to the gender gap in math, as fifth grade math achievement explains the entire eighth grade female–male science gap. Prior math and reading achievement, SES, and classroom fixed effects together reduce the gender gap, the Black–White gap, and the Hispanic–White gap to nonsignificance in Grade 8 (though again, the significance of the gender gap is sensitive to modelling approach).

Limitations

Before turning to a discussion of these findings, we note some limitations of the study. First, as with most longitudinal studies, the *ECLS-K:99* experienced high levels of attrition. Although the longitudinal sampling weights are designed to adjust for attrition, such adjustment is not perfect and bias can still appear. If science gaps among attriters differed compared to science gaps among those who remained in the study, the gaps estimated here may be biased. Similarly, if the relationships among the covariates and science scores differed among attriters compared to nonattriters, our estimates of adjusted gaps may be biased. Second, although the *ECLS-K:99* offers the best data available for answering this

Table 4
Gender and Racial/Ethnic Science Test Score Gaps in the Eighth Grade With and Without Various Fixed Effects and Individual Controls

	(1) b (*SE*)	(2) b (*SE*)	(3) b (*SE*)	(4) b (*SE*)	(5) b (*SE*)	(6) b (*SE*)
Female	–0.178***	–0.175**	–0.213**	–0.117**	–0.116**	–0.114
	(0.050)	(0.054)	(0.082)	(0.038)	(0.042)	(0.058)
Black	–0.709***	–0.670***	–0.507*	–0.295**	–0.210	–0.178
	(0.127)	(0.132)	(0.206)	(0.110)	(0.123)	(0.186)
Hispanic	–0.215	–0.164	–0.233	0.040	0.126	0.175
	(0.110)	(0.109)	(0.175)	(0.082)	(0.081)	(0.133)
Asian	–0.042	–0.176	–0.405	0.045	0.002	–0.114
	(0.125)	(0.167)	(0.257)	(0.084)	(0.109)	(0.179)
Other race	0.056	–0.035	0.095	0.155	0.188	0.390*
	(0.133)	(0.147)	(0.223)	(0.086)	(0.127)	(0.170)
Grade 5 math				0.416***	0.409***	0.390***
				(0.026)	(0.033)	(0.051)
Grade 5 reading				0.348***	0.321***	0.329***
				(0.028)	(0.035)	(0.048)
SES				0.043	0.042	0.044
				(0.029)	(0.034)	(0.050)
Constant	0.138***	0.128**	0.134*	0.004	–0.027	–0.046
	(0.042)	(0.043)	(0.066)	(0.033)	(0.034)	(0.053)
Fixed Effects	**School**	**Teacher**	**Classroom**	**School**	**Teacher**	**Classroom**
N	3,853	3,853	3,853	3,853	3,853	3,853
R^2	0.695	0.832	0.915	0.860	0.911	0.954
Adjusted R^2	0.497	0.630	0.695	0.769	0.804	0.834

Note. Test scores are standardized to $M = 0$ and $SD = 1$. Standard errors that account for the *ECLS-K:99*'s complex sampling design are in parentheses. Other race = student is American Indian/Alaska Native, Native Hawaiian/Pacific Islander, or multiracial.
$^*p < .05$. $^{**}p < .01$. $^{***}p < .001$.

study's research questions, the students in this sample were eighth graders in 2007 and may not be representative of eighth graders today. Finally, as is always the case with observational studies such as this, the possibility of omitted variables bias prevents us from interpreting parameter estimates causally.

Explaining Gaps

SES and prior achievement. The relative importance of SES and prior math and reading achievement in explaining science gaps varied depending on the gap in question. Prior math achievement alone explained all of the gender gap in science, approximately 56% of the Black–White science gap, and approximately 52% of the Hispanic–White science gap. Math achievement did more to explain the Black–White science gap than did reading, but reading did more to explain the Hispanic–White gap than did math. Although SES on its own explained meaningful portions of the Black–White and Hispanic–White science gaps, SES explained very little of these gaps above and beyond what was explained by prior math and reading scores.

Scholars have previously established sound theory describing how students' math and literacy skills may affect their science learning (Fang, 2011; Snow, 2010; Snow & Uccelli, 2008). Our analyses offer the first empirical estimates of the extent to which math and reading skills explain later science gaps in a nationally representative sample. Although we cannot infer that closing fifth grade math and reading gaps would result in subsequent narrowing of the eighth grade science gaps of the magnitudes seen in these models, the fact that math and reading scores behave differently in explaining various gaps suggests that these variables' explanatory power may derive from more than just their covariance with general cognitive skills. Future research should explore this area more fully. Additionally, our results show that although math and/or reading skills help explain large portions of the science gaps, meaningful Black–White and Hispanic–White gaps remain after controlling for math and reading. Gaps in science achievement should therefore be investigated separately from math and reading gaps, as they appear to have distinct influences. Some of these separate causes likely relate to subject-matter-specific inequities at the school and classroom levels.

School and classroom effects. The results of our school, teacher, and classroom fixed effects models are consistent with past research demonstrating that, compared to White students, Black and

Hispanic students tend to experience less rigorous and engaging science curricula and have access to fewer or lower quality science facilities and less qualified teachers (Banilower et al., 2013; Jacob, 2007; Oakes et al., 1990). For Hispanic students, school-level variables appear to play an important role in science gaps; although White students scored .19 *SD* higher than Hispanic students with similar SES and prior achievement (column 6 of Table 3), this gap switched direction and lost statistical significance when school fixed effects were added to the model (column 4 of Table 4).

Classroom-level variables may be relatively more important in shaping the Black–White science gap. School fixed effects reduced the Black–White gap by only .10 *SD* (when added to a model conditioning on SES and prior achievement), and the adjusted within-school Black–White gap remained statistically significant. In contrast, classroom fixed effects reduced the adjusted Black–White gap such that the gap was statistically indistinguishable from zero. This classroom effect is consistent with a situation in which White students tend to receive higher quality instruction than Black students who attend the same school. However, if students were assigned to eighth grade science classrooms based on their prior science achievement, part of the observed classroom effect will be due to this sorting process (it is also worth noting that a sizeable Black–White science gap still exists within classrooms in models that do not control for SES and prior achievement). At the very least, these results demonstrate that more could be done to equalize the science achievement of Black and White students within schools.

Unmeasured psychological factors. Past research points to additional factors influencing test score gaps that we were unable to examine in the *ECLS-K:99* data. For example, stereotypes of Black and Hispanic students as low achievers, and of females as being less suited for science than males, are well documented (Hill et al., 2010; Nguyen & Ryan, 2008). These stereotypes can affect observed science test score gaps through various channels, including stereotype threat, whereby scores of students from stereotyped groups are temporarily depressed through the activation of stereotypes prior to test-taking (Nguyen & Ryan, 2008). Stereotypes can also lower students' motivation to learn science, either because they wish not to enter a field in which, according to stereotypes, members of their racial/ethnic group or gender do not belong, or because they have internalized these biases themselves (Hill et al., 2010). Such damage to students' self-efficacy can cause them to give up on scientific endeavors more easily and can prevent them from pursuing experiences from which they would develop skills in science (Hill et al., 2010).

Conclusion

The urgency of developing a scientifically literate citizenry stems from the demands of living in a high-tech and global economy (DeBoer, 2000; Muller et al., 2001), and science achievement gaps raise concerns about equity, efficiency, and the nation's future. In this study, we found that science gaps by gender and race/ethnicity tended to remain stable or narrow as students progressed through elementary and middle school. However, the overall sizes of these unadjusted gaps were often larger than gaps in mathematics and reading. Our findings indicate that the "leaky" science pipeline may begin as early as third grade, suggesting that interventions aimed at closing gaps should begin when students are young. The results of our explanatory analyses point to areas of potentially fruitful future research on the causal roles that prior math and reading skills, school quality, teacher quality, and curriculum may play in expanding or closing science achievement gaps. Such research could eventually result in an understanding of the malleable factors that can be manipulated so as to equalize opportunity within the scientific fields.

Although this study was motivated in part by the fact that students' early preparation in science predicts their attainment of a science bachelor's degree, we close with a reminder that inequities persist beyond degree attainment and entry into a scientific occupation. For example, White chemists earn higher pay than Black or Hispanic chemists; for Hispanics, the pay gap seems to be explained by training, and for Blacks, the gap appears to be due to discrimination (Broyles & Fenner, 2010). Gender pay gaps also exist in STEM fields, and people tend to hold negative opinions of women who enter what are perceived to be masculine occupations, such as those in STEM fields (Hill et al., 2010). Women in these jobs must therefore outperform their male colleagues in order to be judged equally competent (Hill et al., 2010). Eliminating science test score gaps and improving rates of STEM entry and persistence therefore cannot be the final goals; efforts directed towards these outcomes must also be accompanied by efforts to correct inequities experienced by women and minorities in the STEM workforce.

NOTES

We are grateful to Soojin Oh, Celia Gomez, and our anonymous referees for their helpful comments on this article. All errors and omissions are our own.

[1]We found little research that analyzes science gaps in the *ECLS-K:99*. Kohlhaas, Lin, and Chu (2010) use ANOVA to test for differences in fifth grade science scores by gender, race/ethnicity, and poverty level; Lin and Wilson (2014) conducted similar analyses using eighth grade scores. Neither examined gap changes over time nor attempted to explain gaps.

[2]We note that multiple-choice tests have limitations in measuring some important aspects of science proficiency, which may have implications when interpreting gap estimates. For example, in 2009, the NAEP administered science assessments with computer-based tasks and hands-on tasks. In contrast to multiple-choice tests, the computer-based tasks showed no gender gap in Grades 4, 8, or 12, and girls scored higher than boys on the hands-on tasks in all grades. Asian/Pacific Islanders scored similarly, or higher, to White students on these tasks; similar to multiple-choice tests, however, Black–White and Hispanic–White gaps existed on these tests (NCES, 2012b).

[3]We chose Grade 5 math and reading scores (as opposed to scores from earlier grades) because the higher level skills assessed in Grade 5 are more consistent with our theory about how math and reading skills affect students' ability to develop and express understanding of Grade 8 science objectives. In supplementary analyses, we replaced Grade 5 scores with Grade 1 scores; as expected, Grade 1 math and reading scores help to explain science gaps in Grade 8, but not to the same extent as do Grade 5 scores.

[4]In models estimating gaps changes, we use weight C567CW0. In models explaining Grade 8 gaps, we use weight C67PW0. We incorporate strata and PSU variables through the *svy* command suite in Stata;

in models with school, teacher, or classroom fixed effects, we use the sampling weight with robust standard errors.

[5]Results that adjust gap changes for the differing theta reliabilities at each time point yield similar results.

[6]Note that the school fixed effects model does not imply that closing the mean within-school gap would result in a proportional closure of the overall gap in the population (see Hanushek & Rivkin, 2006).

[7]The adjusted within-classroom gender gap is significant (with nearly identical magnitude) when indicators for all groups are used, as opposed to the "other race" indicator.

[8]Overall, models using Grade 5 science scores as the outcome show similar patterns of results as those seen in Table 4. In Grade 3, prior math scores and teacher fixed effects explain somewhat less of the gaps compared to later grades.

REFERENCES

Association of American Universities. (2006). *National defense education and innovation initiative: Meeting America's economic and security challenges in the 21st century*. Washington, DC: Association of American Universities.

Banilower, E. R., Smith, P. S., Weiss, I. R., Malzahn, K. A., Campbell, K. M., & Weis, A. M. (2013). *Report of the 2012 national survey of science and mathematics education*. Chapel Hill, NC: Horizon Research, Inc.

Broyles, P., & Fenner, W. (2010). Race, human capital, and wage discrimination in STEM professions in the United States. *International Journal of Sociology and Social Policy*, *30*, 251–266. doi: 10.1108/01443331011054226

Campbell, J. R., Hombo, C. M., & Mazzeo, J. (2000). *NAEP 1999 trends in academic progress: Three decades of student performance* (NCES 2000–469). Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics.

Chen, X., & Weko, T. (2009). *Students who study science, technology, engineering, and math (STEM) in postsecondary education* (NCES 2009–161). Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Currie, J. M. (2005). Health disparities and gaps in school readiness. *Future of Children*, *15*, 117–138.

DeBoer, G. E. (2000). Scientific literacy: Another look at its historical and contemporary meanings and its relationship to science education reform. *Journal of Research in Science Teaching*, *37*, 582–601.

Drew, D. E. (2011). *STEM the tide: Reforming science, engineering, and math education in America*. Baltimore: Johns Hopkins University Press.

Eccles, J. S. (2007). Where are all the women? Gender differences in participation in physical science and engineering. In S. J. Ceci & W. M. Williams (Eds.), *Why aren't more women in science? Top researchers debate the evidence* (pp. 199–210). Washington, DC: American Psychological Association.

Entwisle, D. R., Alexander, K. L., & Olson, L. S. (2000). Summer learning and home environment. In R. D. Kahlenberg (Ed.), *A notion at risk* (pp. 9–30). New York: Century Foundation Press.

Fang, Z. (2011). The language demands of science reading in middle school. *International Journal of Science Education*, *28*, 491–520.

Griffith, A. L. (2010). Persistence of women and minorities in STEM field majors: Is it the school that matters. *Economics of Education Review*, *29*, 911–922.

Hanushek, E. A., & Rivkin, S. G. (2006). *School quality and the black-white achievement gap* (No. w12651). Cambridge, MA: National Bureau of Economic Research.

Hart, B., & Risley, T. R. (2003). The early catastrophe: The 30 million word gap by age 3. *American Educator*, *27*, 4–9.

Hill, C., Corbett, C., & Rose, A. (2010). *Why so few? Women in science, technology, engineering, and mathematics*. Washington, DC: American Association of University Women.

Jacob, B. (2007). The challenge of staffing urban schools with urban teachers. *The Future of Children*, *17*, 129–153.

Kohlhaas, K., Lin, H., & Chu, K. (2010). Disaggregated outcomes of gender, ethnicity, and poverty on fifth grade science performance. *Research in Middle Level Education Online*, *33*, 1–12.

Lacey, T. A., & Wright, B. (2009). Occupational employment projections to 2018. *Monthly Labor Review*, *132*, 82–123.

Lankford, H., Loeb, S., & Wyckoff, J. (2002). Teacher sorting and the plight of urban schools: A descriptive analysis. *Educational Evaluation and Policy Analysis*, *24*(1), 37–62.

Lin, H., & Wilson, B. (2014). The comparisons between ethnicity and poverty level on eight [sic] graders' science performance. *Journal of Studies in Education*, *4*, 122–130. doi:10.5296/jse.v4i4.6600

Muller, P. A., Stage, F. K., & Kinzie, J. (2001). Science achievement growth trajectories: Understanding factors related to gender and racial-ethnic differences in precollege science achievement. *American Educational Research Journal*, *38*, 981–1012.

Najarian, M., Pollack, J. M., & Sorongon, A. G. (2009). *Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 (ECLS-K), psychometric report for the eighth grade* (NCES 2009–002). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

National Center for Education Statistics. (2012a). *The Nation's Report Card: Science 2011* (NCES 2012–465). Washington, DC: Institute of Education Sciences, U.S. Department of Education.

National Center for Education Statistics. (2012b). *The Nation's Report Card: Science in Action: Hands-on and interactive computer tasks from the 2009 Science Assessment* (NCES 2012-468). Washington, DC: Institute of Education Sciences, U.S. Department of Education.

Nguyen, H. D., & Ryan, A. M. (2008). Does stereotype threat affect test performance of minorities and women? A meta-analysis of experimental evidence. *Journal of Applied Psychology*, *93*, 1314–1334. doi: 10.1037/a0012702

Oakes, J., Ormseth, T., Bell, R., & Camp, P. (1990). *Multiplying inequalities: The effects of race, social class, and tracking on opportunities to learn mathematics and science*. Santa Monica, CA: RAND.

Ost, B. (2010). The role of peers and grades in determining major persistence in the sciences. *Economics of Education Review*, *29*, 923–934.

President's Council of Advisors on Science and Technology. (2012). *Engage to excel: Producing one million additional college graduates with degrees in science, technology, engineering, and mathematics*. Retrieved from http://www.whitehouse.gov/sites/default/files/microsites/ostp/pcast-engage-to-excel-final_2-25-12.pdf

Price, J. (2010). The effect of instructor race and gender on student persistence in STEM fields. *Economics of Education Review*, *29*, 901–910.

Quinn, D. M. (2015). Kindergarten black-white test score gaps: Re-examining the roles of socioeconomic status and school quality with new data. *Sociology of Education*, *88*, 120–139. doi: 10.1177/0038040715573027

Riegle-Crumb, C., & King, B. (2010). Questioning a white male advantage in STEM: Examining disparities in college major by gender and race/ethnicity. *Educational Researcher*, *39*(9), 656–664.

Riegle-Crumb, C., King, B., Grodsky, E., & Muller, C. (2012). The more things change, the more they stay the same? Prior achievement fails to explain gender inequality in entry into STEM college majors over time. *American Educational Research Journal*, *49*, 1048–1073. doi: 10.3102/0002831211435229

Riegle-Crumb, C., & Moore, C. (2014). The gender gap in high school physics: Considering the context of local communities. *Social Science Quarterly*, *95*, 253–268. doi: 10.1111/ssqu.12022

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. doi: 10.3102/0002831210372249

Ruby, A. (2006). Improving science achievement at high-poverty urban middle schools. *Science Education*, *90*(6), 1005–1027.

Snow, C. E. (2010). Academic language and the challenge of reading for learning about science. *Science*, *328*, 450–542.

Snow, C. E., & Uccelli, P. (2008). The challenge of academic language. In D. R. Olson & N. Torrance (Eds.), *The Cambridge handbook of literacy* (pp. 112–133). New York: Cambridge University Press.

Stohr-Hunt, P. M. (1996). An analysis of frequency of hands-on experience and science achievement. *Journal of Research in Science Teaching*, *33*(1), 101–109.

Tai, R. H., Liu, C. Q., Maltese, A. V., & Fan, X. (2006). Planning early for careers in science. *Science*, *312*, 1143–1144.

Tourangeau, K., Nord, C., Lê, T., Sorongon, A. G., & Najarian, M. (2009). *Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 (ECLS-K), combined user's manual for the ECLS-K eighth-grade and K–8 full sample data files and electronic codebooks* (NCES 2009–004). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics. (n.d.). *NAEP data explorer*. Available at http://nces.ed.gov/nationsreportcard/naepdata/dataset.aspx

Wang, X. (2013). Why students choose STEM majors: Motivation, high school learning, and postsecondary context of support. *American Educational Research Journal*, *50*, 1081–1121.

AUTHORS

DAVID M. QUINN is an advanced doctoral student at the Harvard Graduate School of Education, 248 Gutman Library, 6 Appian Way, Cambridge, MA 02138; *david_quinn@mail.harvard.edu*. His research interests relate to measuring, explaining, and ending educational inequity.

NORTH COOC, EdD, is an assistant professor of special education at the University of Texas at Austin, 1 University Station, Stop 5300, Austin, TX 78712; *ncooc@austin.utexas.edu*. His research focuses on racial disparities in special education enrollment and academic trajectories for children with disabilities.

Manuscript received January 18, 2015
Revisions received April 23, 2015, and July 7, 2015
Accepted July 10, 2015