American Sociological Review
77(3) 463–485

DOI: 10.1177/0003122412440802
http://asr.sagepub.com

School Context and the Gender Gap in Educational Achievement

Joscha Legewie[a] and Thomas A. DiPrete[a]

Abstract

Today, boys generally underperform relative to girls in schools throughout the industrialized world. Building on theories about gender identity and reports from prior ethnographic classroom observations, we argue that school environment channels conceptions of masculinity in peer culture, fostering or inhibiting boys' development of anti-school attitudes and behavior. Girls' peer groups, by contrast, vary less strongly with the social environment in the extent to which school engagement is stigmatized as un-feminine. As a consequence, boys are more sensitive than girls to school resources that create a learning-oriented environment. To evaluate this argument, we use a quasi-experimental research design and estimate the gender difference in the causal effect of peer socioeconomic status (SES) as an important school resource on test scores. Our design is based on the assumption that assignment to 5th-grade classrooms within Berlin's schools is as good as random, and we evaluate this selection process with an examination of Berlin's school regulations, a simulation analysis, and qualitative interviews with school principals. Estimates of the effect of SES composition on male and female performance strongly support our central hypothesis, and other analyses support our proposed mechanism as the likely explanation for gender differences in the causal effect.

Keywords

causal inference, education, gender, gender gap, peer effects

Throughout adolescence, boys are overrepresented among high school dropouts, special education students, and every failed or special needs category. Boys' notorious underperformance in school and their tendency to disrupt the learning process in classrooms has sparked intense academic and public debates about the causes of what many now call the "problem with boys." Some see the gender gap as largely biological in origin. Others blame schools for a de-masculinized learning environment and a tendency to evaluate boys negatively for fitting into this environment less well than girls. Yet, the true impact of school context on the size of the gender gap in academic performance remains controversial. The 1966 Coleman Report raised the profile of research on school effects, and much attention since then has been motivated by a concern for equality of educational opportunity by social class and race. Now that

[a]Columbia University

Corresponding Author:
Joscha Legewie, Columbia University, Department of Sociology–MC9649, 606 W. 122nd Street, New York, NY 10027
E-mail: jpl2136@columbia.edu

a growing gender gap in educational attainment has emerged, it is important to extend this line of research and ask whether schools affect gender inequality, and if so, what are the mechanisms by which this occurs.

Integrating theories about gender identity, adolescent culture, and findings from prior ethnographic classroom observations, we argue that the school environment channels conceptions of masculinity in the peer culture, fostering or inhibiting boys' development of anti-school attitudes and behavior. An academically oriented environment suppresses the construction of masculinity as oppositional and instead facilitates boys' commitment by promoting academic competition as an aspect of masculine identity. Lower quality schools, by contrast, implicitly encourage—or at least do not inhibit—development of a peer culture that constructs resistance to schools and teachers as valued masculine traits. Girls' peer groups, by contrast, vary less strongly with the social environment in the extent to which school engagement is stigmatized as un-feminine. As a result, boys, in particular, benefit from school resources that create a learning-oriented peer culture, and the size of the gender gap in educational performance depends on environmental factors connected to the quality of schools.

We evaluate our argument with a quasi-experimental research design using reading test scores as an outcome variable and the socioeconomic composition of the student body as the focal treatment variable. This design is based on within-school variation across classes using the ELEMENT data from one German city-state (i.e., Berlin). In contrast to the United States, the lack of performance-based tracking in Berlin's elementary schools, and parents' smaller influence on classroom assignment, makes it plausible that student assignment to elementary school classrooms in Berlin is almost random. To develop a detailed understanding of the actual selection process, we examine official school regulations, provide statistical evidence from simulation analyses, and conduct qualitative interviews with school principals. Results suggest that randomness indeed plays an important role in the assignment process, but they also point at potential sources of bias. We address these potential biases statistically with targeted sensitivity analyses using instrumental variable and sample restriction methods. We supplement the ELEMENT analysis with estimates from a large-scale nationally representative German dataset (PISA-I-Plus 2003) to address potential concerns about generalizability of the findings.

Results of our investigation support our core hypothesis. In addition, a systematic comparison of our preferred explanation with alternative accounts suggests that our hypothesized mechanism brings about the gender difference in the causal effect of SES composition on student achievement. Our findings speak to recent political debates about the educational shortcomings of boys by deepening our understanding of their notorious underperformance. Our analytic strategy also makes a methodological contribution by illustrating how a detailed study of selection processes using simulations and qualitative interviews can assist estimation of causal effects.

EDUCATIONAL OUTCOMES AND SCHOOLS

The 1966 Coleman Report (Coleman 1966) claimed that family is the most important determinant of achievement, but performance improves when classroom peers have greater socioeconomic resources and are racially integrated (see also Coleman 1961; Jencks and Mayer 1990; Kahlenberg 2001). As Coleman and others have subsequently argued, students are motivated to invest more heavily in their studies when adolescent culture rewards academic performance and thereby supports parents' and teachers' reward systems. But when adolescent culture values other behaviors more highly (e.g., sports, being popular with the opposite sex, or opposition to school authority), and especially when adolescent culture denigrates academic achievement, it inhibits academic investment and weakens academic achievement. Simply put, students who are highly motivated and capable (attributes more common at higher SES levels) create a learning-oriented peer culture (Jencks and Mayer 1990; Rumberger and Palardy 2005; Sewell, Haller, and Portes 1969).

For about 20 years following the release of the Coleman Report, the literature reported that school effects were relatively small in comparison to family effects, and therefore "schools are not an effective agent for the redistribution of societal resources" (Hallinan 1988:255; see also Hanushek 1989). This pessimistic view of schools began to change with the rise of the accountability and standards movement to improve schools (Schneider and Keesler 2007). Reanalysis of earlier studies suggested a more consistently positive relationship between school resources and student achievement (Greenwald, Hedges, and Laine 1996) and found that teacher quality, in particular, was a major input into student learning (see also Murnane 1983).

The renewed focus on schools' impact on learning has not obscured attention to the central conclusion of the Coleman Report that "the social composition of the student body is more highly related to achievement, independent of the student's own social background, than is any school factor"(Coleman 1966:325). Far more than was historically appreciated, estimation of peer effects is challenging (Angrist and Pischke 2008) because of non-random selection and unmeasured confounding variables (e.g., teacher quality) that affect student outcomes. The most persuasive recent studies use natural experiments to estimate the impact of changes in class composition on outcomes (e.g., Imberman, Kugler, and Sacerdote 2009). A second strategy is to exploit potentially random assignment of students to classes within schools. This strategy is only persuasive when applied in school districts that make it difficult for parents to teacher shop (Ammermueller and Pischke 2009). A third strategy examines arguably random fluctuations in adjacent cohorts (e.g., of gender or race composition) for the same school and grade (Gould, Lavy, and Pawerman 2009; Hoxby 2000), although these studies have not looked at peer effects related to socioeconomic characteristics. The magnitude of estimated effects is not large (about .15 standard deviations), but it is about the same as some of the most believable estimates of teacher effects, whether for academic, social, or behavioral outcomes (Jennings and DiPrete 2010; Rockoff 2004). Meanwhile, recent studies whose primary estimation strategies control for observable potential confounders have found a similar effect size on test scores (Crosnoe 2009; Rumberger and Palardy 2005).

SCHOOL CONTEXT AND THE GENDER GAP IN EDUCATION

The original focus on school effects developed out of a concern for equality of educational opportunity by social class and race. Now that a growing gender gap in educational attainment has emerged, it is natural to ask whether schools also affect gender inequality, and if so, what are the mechanisms by which this occurs. Starting in the 1970s and early 1980s (Spender 1982; Stanworth 1984), ethnographic studies documented girls' and boys' gendered behavior at school as well as the different ways that teachers treat girls and boys. Although overt discrimination against girls in the classroom has declined over the past three decades, recent studies suggest that boys still verbally dominate the classroom (Jovanovic and King 1998; Sadker and Zittleman 2009). Meanwhile, the once celebrated coeducation of boys and girls as a pivotal step toward gender equality is now challenged by the increasing popularity of single-sex private schools, the opening of girls-only public schools, and the claimed educational shortcomings of coeducation for girls (Morse 1998; Salomone 2003).

Despite these important strands of research and the general recognition that schools are an important context for socialization of young adolescents, literature on the educational gender gap has widely ignored the school as a potential source of variation in this gap. To our knowledge, Dresel, Stöger, and Ziegler (2006), Machin and McNally (2005), and Schöps and colleagues (2004) are the only studies that examine variation in the size of the gender gap across a number of schools. Using data from a specific region in Germany (i.e., Baden-Württemberg), Dresel and colleagues (2006) found substantial variation in the educational gender gap across schools and classes, while Schöps and colleagues (2004) obtained a similar finding using the German PISA data. Machin and McNally (2005), by contrast, argue that specific school-based characteristics, such as school inputs, teaching practices, and the examination system, have no effect on the gender gap. We extend this line of research by building on reports from prior ethnographic classroom observations and theories about gender identity to understand the role of school context for boys' underachievement.

Boys' Underachievement, Gender Identity, and School Climate

In a classic study, Willis (1981) argued that working for academic success is in conflict with adolescent conceptions of masculinity. He portrayed working-class white boys' anti-school attitudes and behavior as arising from peer dynamics and a belief that their opportunity to use education to achieve success in the labor market was blocked (see also Kao, Tienda, and Schneider 1996; MacLeod 2008). In line with Willis's early findings, much of the literature on boys' underachievement focuses on disincentives to engage with school that stem from adolescent conceptions of masculinity, which are developed and reinforced in peer groups. Gender differentiation and creation of stereotypical gender identities begin in early childhood before children have had any experience with school (Davies 2003; Maccoby 1998; Thorne 1993). Gender-differentiated childhood cultures become the basis for gender-differentiated adolescent cultures, which are important influences on how children view school, whether they take school seriously, and how hard they work as students (Steinberg et al. 1997).

Classroom observations and other ethnographic studies document how gender identities are constructed in the classroom and how these gender cultures affect boys' and girls' interactions and approach to education (Francis 2000; Pickering 1997; Salisbury and Jackson 1996; Skelton 1997). Boys tend to be noisier, more physically active, and more easily distracted than are girls (Francis 2000; Howe 1997; Spender 1982; Younger, Warrington, and Williams 1999). Studies also find that masculine stereotypes portray boys as competitive, active, aggressive, and dominating, while girls are viewed as conciliatory and cooperative (Francis 2000). Other scholars argue that stereotypical gender identities perpetuate the belief that girls have to work hard to learn in school, whereas boys are naturally gifted (Cohen 1998; Epstein 1998; Mac an Ghaill 1994; Power et al. 1998; Quenzel and Hurrelmann 2010). Cohen (1998) shows that these gendered beliefs are reflected in boys' casual and detached attitude toward school, which accords with the ethnographic studies referenced earlier. Despite the transformation of gender relations in modern societies, stereotypical gender identities continue to shape orientations toward school and produce behaviors that reinforce these identities while potentially affecting children's academic success. Morris (2008:736) observed this process at a rural high school: "girls tended to direct considerable effort and attention to school" whereas "boys . . . took pride in their lack of academic effort" as an aspect of their masculine identity.[1]

Peers and the adolescent reward system reinforce gender identities and gendered behavior patterns. In some contexts, disruptive behavior produces status gains in lower SES students' peer groups. Working for academic achievement, by contrast, is labeled as feminine and thereby stigmatized. Girls, however, typically view school work as acceptable and sometimes even encouraged. In a lack of parallelism with male peer groups, working-class and lower-class female peer groups do not consider resistance to authority and disengagement from school to be core aspects of feminine identity (Maccoby 1998). As a result, girls' peer culture more readily encourages attachment to teachers and school.[2]

A diverse group of studies supports the role of peers in shaping attitudes toward school. Coleman (1961), Eitzen (1975), Steinberg and colleagues (1997), and, more recently, Bishop and colleagues (2003) argue that adolescents value attributes that make one cool or popular because these attributes are linked with high status. Based on her own and others' ethnographic work, Epstein (1998:106) argues that "the main demand on boys from within their peer culture . . . is to appear to do little or no work" whereas for girls "it seems as if working hard at school is not only accepted, but is, in fact, wholly desirable." Morris (2008:738; for other examples, see Epstein 1998) documents this attitude in a conversation between three boys in an English class:

> Kevin: "I don't want to put in a lot of extra effort like that. I'll just do the basic stuff and get a B." "I got an 87 in here," he says proudly. Warren chimes in, "Yeah, I hate these pussies who make like an A minus and then they whine about it." Kevin says, "Yeah it's like why do you care? Why does it have to be better? Nothin' wrong with a normal grade!"

Although ethnographic studies document substantial within-gender diversity in the construction of gender identities, evidence on typical gender differences is rather persuasive. Young boys tend to construct masculinity at least partly in terms of resistance to school. This conception of masculinity may be partially responsible for male underachievement (Francis 2000; Pickering 1997; Salisbury and Jackson 1996; Skelton 1997). The conception of female identity and female peer culture, by contrast, is not as closely tied to resistance to school, and indeed may even support schoolwork as a positive attribute of femininity. As a result, girls consistently have better work habits and a stronger pro-school orientation.

While Willis and others mainly focus on consequences of lower- and working-class backgrounds for anti-school attitudes among boys, we are interested in the school and class environment as a context that either encourages or limits development of anti-school attitudes and behavior. High-status parents generally foster an orientation for their boys that is at least instrumentally focused on high performance in school. These parents also have resources to intervene in their children's lives to counter signs of educational detachment or poor performance. As Coleman and others argue, schools can play a similar role in enhancing students' incentives to be engaged with academics by creating a learning-oriented peer culture. In this line, many scholars argue that the success of some charter schools, such as KIPP and the Harlem Children's Zone, comes from their ability to foster a learning-oriented environment (Ravitch 2010).

We argue that boys gain more than girls from a learning-oriented environment because it channels how masculinity is constructed in the school culture. Such an environment promotes academic competition as an aspect of masculinity and encourages development of adaptive strategies that enable boys to maintain a show of emotional coolness toward school while being instrumentally engaged in the schooling process. In other words, academic competition as one of the "different ways of 'doing' masculinity" (Francis 2000:60; see also Mac an Ghaill 1994) becomes a more important part of the construction of masculine identity in certain environments.

As is true in the family, production of an academically oriented environment in school is not effortless. It requires resources. Better facilities, better curriculum, better teachers, and better support staff all can produce more value-added in school. Both boys and girls will generally benefit from better schooling, of course, but we expect school inputs that strengthen a learning orientation in the student culture have the potential to enhance educational outcomes especially strongly for boys. Teachers, for example, can promote a learning-oriented student culture. Accordingly, we expect teachers with the right collection of skills might have especially positive effects on boys' achievement.

Here, the school resource of central interest is the socioeconomic composition of the student body. Peers' impact on school climate and student achievement has played a crucial role in the literature on schools ever since Coleman (1966:325) claimed that "the social composition of the student body is more highly related to achievement, independent of the student's own social background, than is any school factor." The mechanism behind this association is cultural; students with high motivation and achievement from high-SES backgrounds create a learning-oriented peer culture and assist teachers in the process of education (Jencks and Mayer 1990; Rumberger and Palardy 2005; Sewell et al. 1969). We expect the disadvantages of low-SES composition to be larger for boys than for girls because of the evidence that lower SES student bodies create a stronger oppositional culture in male than in female peer groups. Conversely, an academically oriented environment in schools channels conceptions of adolescent and pre-adolescent masculinity, suppresses boys' negative attitudes toward school, and facilitates academic competition as an aspect of masculine identity. Girls' peer groups, on the other hand, more readily and independently of school context encourage attachment to teachers and school, and do not identify femininity with disengagement from school. Based on this argument, we hypothesize that the female advantage in academic achievement is bigger in schools with a lower SES composition in their student body.[3]

DATA AND METHODS

We address our core hypothesis with the German ELEMENT dataset, using reading test scores as an outcome variable and SES composition of classroom peers as our focal treatment variable. The ELEMENT dataset is a longitudinal study that assessed development of reading and math ability in the 4th, 5th, and 6th grades in Berlin schools (Lehmann and Lenkeit 2008). It includes about 3,300 students who attended 4th grade during the 2002 to 2003 school year in 71 randomly selected elementary schools in Berlin, and all 1,700 students who attended 5th grade in 2003 to 2004 in one of Berlin's 31 upper-secondary schools that begin with 5th grade.[4] In our final models, we combine these two ELEMENT samples and control for school type through school-level fixed effects. We also examine whether relevant effects vary by school type using interaction terms (they do not). Part A in the Appendix provides a short introduction to the German educational system.

The ELEMENT dataset includes at least two classrooms for every school.[5] This feature provides the basis for a quasi-experimental design. It allows us to estimate contextual effects of 5th-grade class composition by gender using school-level fixed-effects models, because the original assignment to elementary school classes in 1st grade within schools is not subject to self-selection or parental control.[6] This estimation strategy provides a clear advantage over similar estimates based on data from U.S. schools, where performance-based tracking in elementary schools and parents' influence on assignment to classes are more pronounced.

Our quasi-experimental research design provides high internal validity and allows us to make a strong case for causal inference, but the analysis is geographically limited to a single German state. To address this limitation, we supplement the ELEMENT data with the German PISA-I-Plus 2003 data, a German extension of the international PISA study.[7] The PISA-I-Plus includes a nationally representative sample of 9,000 students in at least two 9th-grade classrooms in 220 schools (PISA-Konsortium Deutschland 2006).[8] The two datasets complement each other and provide strong internal and external validity for the estimation of causal effects.

School-Level Fixed Effects as a Quasi-Experimental Identification Strategy

Regression or matching estimates of school effects based on conditioning on observable variables as an identification strategy potentially suffer from endogeneity problems. These strategies rely on the assumption that students are randomly assigned to schools conditional on observable covariates in the model (Sørensen and Morgan 2006). This common identification strategy is especially problematic for estimation of school effects with cross-sectional data. Students clearly are not randomly assigned to schools, and it is unlikely that this non-random assignment can be perfectly modeled with the observed covariates.

To avoid these potential endogeneity problems, we estimate school-level fixed-effects models using ELEMENT and PISA-I-Plus data. Both datasets contain an additional level of analysis, the classroom. We argue that most students are randomly assigned to classrooms conditional on their school in Berlin's elementary and 5th-grade upper-secondary schools (for a similar strategy, see Ammermueller and Pischke 2009). Assuming students' random assignment to classes within schools, we can estimate the causal effect using school fixed-effect models and a measure of SES composition on the classroom level (we discuss the variables in detail below). We specify these models as

$$\begin{aligned} y_{ijk} &= \alpha_j + \gamma(female)_i + \theta(SES\,Comp)_k \\ &\quad + \delta((SES\,Comp)_k \times female_i) \\ &\quad + \beta_1 y_i^{4th\ grade} + \boldsymbol{X}_i \beta_2 + \boldsymbol{U}_k \beta_3 + \varepsilon_{ijk} \end{aligned} \tag{1}$$

where *i, j,* and *k* are indices for individuals, schools, and classes, respectively; α_j are school fixed-effects; $y_i^{4th\ grade}$ is a student's prior achievement measured in 4th grade; and X_i and U_k are sets of control variables on the individual and class level, respectively.[9] Due to data limitations, analysis with the PISA-I-Plus dataset omits the variable of prior achievement on the right-hand side.[10]

These models examine whether class-to-class variation in performance is systematically related to class-to-class variation in socioeconomic composition, controlling for all unobserved school characteristics (and therefore the non-random selection of students into schools). Coefficients of interest are θ, which captures the causal effect of the socioeconomic class composition, and δ, which captures the difference in this effect between boys and girls. We expect to find a positive effect of SES composition, as previously documented, and, more important for our theory, a negative estimate of the interaction term, indicating that boys are more sensitive than girls to peer SES. Pre-treatment control variables on student and class levels are of secondary interest; we include them to increase balance between the treatment and control groups (for a description of the control variables, see Table 1).

Assignment of Students to Classrooms within Schools

Our estimation strategy relies on the assumption that selection of students into different classes within schools is practically random. While students self-select into schools, their allocation to different classes within schools is arguably less selective, but it might not be completely random. In particular, the allocation process, and therefore selection into treatment, might involve three potential biases: (1) parents might influence which class their children attend; (2) schools might allocate students based on certain characteristics (e.g., performance-based tracking or subject choice); and (3) children might self-select over time when certain children have to repeat a class or change schools. Even if students are randomly assigned to classes, certain teachers might be assigned to specific classes based on a classroom's composition, which could create a bias in relevant estimates of classroom composition.

Table 1. Variables in Main Analysis at Individual and Class Level

Variables	Description	Mean	SD	Mean (Male)	Mean (Female)	Diff. in Means
Dependent Variable						
Achievement	5th-grade reading test scores	109.92	13.65	108.79	111.05	−.17**
Independent Variables (Individual Level)						
Prior achievement	4th-grade reading test scores	103.22	15.96	101.39	105.07	−.23**
Female	0 = Male; 1 = Female	.50	.50			
Family background	ISEI scale	51.45	16.63	51.09	51.80	−.04
Migration background	Categorical Variable:					
	1 = both parents born in Germany	.64	.48	.64	.65	−.01
	2 = one parent not born in Germany	.12	.33	.12	.13	−.01
	3 = both parents not born in Germany	.17	.38	.18	.16	.02*
	4 = child not born in Germany	.06	.24	.06	.06	−.01
Class repeater	0 = not repeated; 1 = repeated at least once	.06	.23	.07	.04	.03**
Interaction terms	Female x Family background					
	Female x Migration background					
	Female x SES composition					
Independent Variables (Class Level)						
Size of class	Number of students in the class	23.33	3.41	23.33	23.34	−.00
SES composition	Average ISEI at class level (aggregated)	51.45	9.40	51.08	51.10	−.01

Source: ELEMENT data.

Note: N = 4,372. The difference in means refers to the mean for boys minus the mean for girls divided by the pooled standard deviation. The difference in means is not standardized for the binary variables (i.e., migration background dummies and indicator for class repeater). All continuous variables are standardized for the final analysis.

* $p < .05$; ** $p < .01$; *** $p < .001$ (two-tailed *t*- and *z*-tests).

To develop a deeper understanding of the actual selection process, we conducted a three-part analysis. First, we studied official school regulations in Berlin. Second, we used a simulation-based approach to compare observed composition of classes with simulations involving random assignment of students to classrooms within schools. Third, we conducted qualitative interviews with school principals in Berlin. The resulting detailed picture of the actual selection process allows us to evaluate our argument that self-selection is practically random and to design targeted statistical sensitivity analyses that address potential sources of biases.

School regulations and general considerations. Primary school regulations in Berlin (Grundschulverordnung Berlin, §8) prohibit allocation of students based on gender,

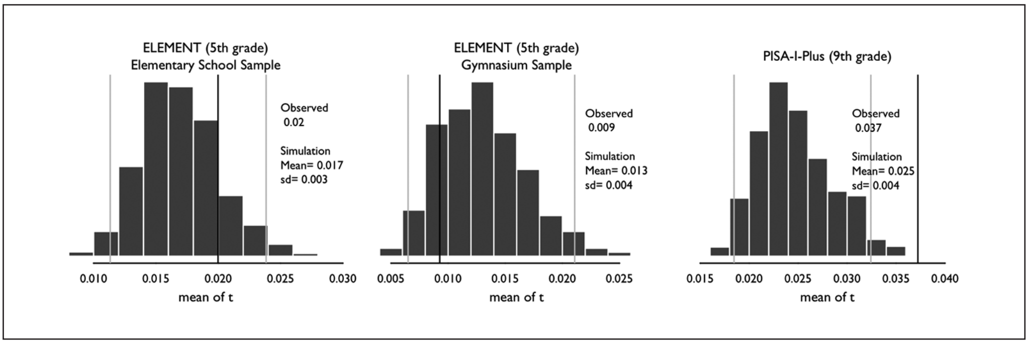

Figure 1. Comparison of Observed Classroom Composition with Simulations Involving Random Assignment
Note: Graphs show average variation of class means within schools for the observed samples (vertical line) together with the sampling distribution of this statistic obtained from 1,000 random simulations (histogram). Vertical grey lines indicate 95% confidence intervals from the simulations.

first language, or performance and emphasize heterogeneity of classes in regard to these characteristics. These legal constraints rule out performance-based tracking, set limits on parental influence over classroom assignment, and provide guidelines for classroom assignment of grade repeaters and newcomers. As a consequence, allocation of students to classrooms based on family background is unlikely. Regulations also mention, however, that schools can consider existing friendships between new students and assign them to the same classroom. This practice, if common, might create a bias in the assignment process that could pose a problem for estimation of the causal effect.

In secondary schools such as those in the PISA-I-Plus data, class-specific tracking based on subject choice (e.g., foreign language) is more common, and a higher number of students have to repeat a class compared with elementary school. This creates potentially non-random allocation of students to classrooms, so the selection problem might be more pronounced in secondary schools. In Berlin, however, 5th-grade upper-secondary schools (grundständige Gymnasien) are different from other secondary schools. Their student population is more homogeneous compared to other secondary schools, which makes a purposeful allocation to different classes relatively inconsequential. In addition, assignment to 5th grade is not subject to selection over time through grade retention because students enter these schools for the first time at 5th grade.

Based on these considerations, we expect that assignment to 5th-grade classrooms is practically random in Berlin's elementary and upper-secondary schools, whereas assignment to 9th-grade classrooms in secondary schools is subject to more pronounced selection processes.

Simulation of random assignment. We use a simulation-based approach to evaluate whether within-school variation in socioeconomic composition across classrooms created by the actual (unknown) allocation process is consistent with random assignment. Figure 1 compares socioeconomic composition across classrooms obtained from simulations that randomly assign students to classrooms (histogram) with the observed composition (vertical line), that is, the average variation of class means within schools (see Part B in the Appendix for details on the simulation).

For the two ELEMENT samples, the observed mean is consistent with a random assignment process. This is in line with our expectation about assignment to classrooms in 5th grade. As expected, however, the observed value for secondary schools in the PISA-I-Plus is relatively unlikely to occur under random assignment. Similar simulations for the proportion of students with migration backgrounds suggest that assignment in regard to this characteristic is consistent with randomness for all three datasets. Finally, variation in gender composition across classrooms within schools is smaller in the actual data than in the simulated distribution (see the online supplement [http://asr.sagepub.com/supplemental]). This result suggests that schools distribute boys and girls equally across classrooms.

These results provide statistical evidence to support the previously described institutional evidence that assignment to classrooms within schools with respect to family background is as good as random in the ELEMENT dataset. By contrast, some non-random selection process seems to play a role for 9th-grade classrooms in secondary schools.

Interviews with school principals. Although simulations are informative, they do not provide information about the actual assignment process. It is still conceivable that non-random selection processes are at work that produce a distribution of students in terms of socioeconomic status that is consistent with a random assignment process. To develop a deeper understanding about the actual assignment process, we conducted 12 interviews with school principals in Berlin as the central actors in the allocation process (nine for elementary schools and three for upper-secondary schools). We selected schools using a random sample that we then supplemented with specific schools to ensure diversity in regard to neighborhood and ethnic composition. Interviews lasted about 15 to 20 minutes and focused on the actual procedure schools use to assign students to classes, criteria that play a role in assignment, the extent to which parents try to influence this process, and how schools deal with parental requests. Interviews also solicited information about how schools assign students who repeat a class or who transfer from other schools and about how teachers are assigned to classrooms. The online supplement contains a detailed description of the sampling procedure and a translation of the interview questions.

While the schools under study use different procedures to assign students to classes, a number of findings emerged from the interviews. First, none of the principals reported that they directly take family background or performance into account in the assignment process, and most schools do not respond to parents who try to influence the process (for an exception, see below). Second, schools try to have classes of similar size. This plays an important role in assignment of students who either repeat a grade or transfer from another school. Third, assignment of teachers to classrooms is generally not connected to socioeconomic composition or other characteristics of the class. Teacher assignment is based on scheduling issues and past experience with the teacher.[11]

A number of potential biases, however, do exist. First, while all school principals emphasized that a desire to equalize classroom size is the main criterion, principals also reported that students who repeat a grade are sometimes assigned to specific classes based on expectations about social dynamics. Second, some principals reported that they take into account whether groups of children attended the same kindergarten and try to assign these students to the same 1st-grade classroom. Other principals said they follow parent requests when they are related to friendships between two new students, which often developed because the children attended the same kindergarten. Third, while most principals reported distributing children with immigration backgrounds equally across classes, two principals said they create a separate class for children who are learning German. While the simulations suggest the contrary, this finding makes it unclear how common the practice of sorting students by migration background or language skills is. We take special care to address this potential issue statistically. Fourth, all principals reported that they try to ensure gender balance between the classrooms. This practice is consistent with results from the simulation insofar as variation in the proportion of female students across classes within schools is smaller than what we would expect from random assignment.

Except for the last criterion, which is irrelevant because boys and girls are equally distributed across families, these selection criteria might induce some systematic bias in the composition of classrooms. The importance of these selection criteria, however, seems to be limited. Most school principals independently and without knowledge of our study concluded that randomness plays an important role in the assignment process because they simply have little prior knowledge about entering students and because the whole process is not very systematic. One elementary school assistant principal and teacher, for example, emphasized that even decades of experience at elementary schools could not remove the inherent unpredictability about classroom dynamics, given the limited prior knowledge about entering students that schools have to work with:

> We have realized again and again that even if we try to make sense of the classroom composition based on names or other attributes we know about, there is no way to know how the class actually turns out in regard to its social composition. Even though I have been working at schools for 40 years now, there are always unexpectedly difficult or balanced classes, which really depends on the personalities of the students inside the classroom so that in the end randomness plays a big role. (translation by authors)

We elicited these and similar concluding remarks from interviewees at the end of interviews by asking how they would weigh the importance of different criteria and whether they thought randomness also plays a role. These observations are particularly interesting considering that we expected a social desirability bias in favor of principals reporting a sophisticated assignment procedure.

Conclusions about selection process. Based on evidence from school regulations, the simulations, and interviews with school principals, we conclude that the role of potential selection biases is limited. Results justify our quasi-experimental design and support our argument that using within-school variation across classrooms in Berlin elementary schools greatly improves our estimates compared to estimates based only on between-school variation. We also recognize the potential selection biases documented by the interviews, and we address these problems statistically by conducting a set of targeted sensitivity analyses. These robustness checks are based on instrumental variable analyses and sample restrictions specifically designed to address each potential source of bias.

Finally, we note that in contrast to most research on compositional school effects, we are not fundamentally interested in school performance as an outcome. Rather, we address contextual determinants of the gender gap in school performance. While evidence from the interviews indicates that students might select into certain classrooms, it seems unlikely there is differential selection of boys and girls into different classrooms. Non-random assignment to classrooms only matters for our key estimation results to the extent that schools treat boys and girls differently during the assignment process. Interviews did not provide any indication of differential treatment of boys and girls, even though school principals were asked directly about such a possibility. This fact enhances our confidence in the validity of our estimates.

Variables and Treatment of Missing Data

Our analysis uses reading test scores in 5th grade (ELEMENT) and 9th grade (PISA-I-Plus) as the main outcome variable (see Table 1 for descriptive statistics). Campbell and colleagues (2001:1) describe reading scores as "one of the most important abilities students acquire through their early school years. It is the foundation for learning across all subjects." Reading literacy also figures importantly in research on the gender gap in education, because boys' reading test scores lag notably behind that of females (Buchmann, DiPrete, and McDaniel 2008). Some researchers even argue that boys' failure in general is due to their deficits in reading (Whitmire 2010). We measure test scores on a common scale using item response theory and standardize them with a mean of zero and a standard deviation of one.

Our focal treatment variable is the student body's socioeconomic (SES) composition, which we measure at the classroom level as the average social status on the ISEI scale (Ganzeboom, Treiman, and Ultee 1991).[12] One could argue that peers' prior achievement is a more natural contextual measure for testing our core hypothesis. However, peer achievement is endogenous in our data because it is measured after random assignment. Moreover, the correlation between peer achievement and SES is too high to reliably distinguish effects of the two variables. Accordingly, SES composition provides a stronger test (i.e., one resting on weaker assumptions) of our theory than could be obtained using peer achievement. In addition, a long tradition in sociology, going back to the Coleman Report, sees SES composition as connected to a peer group's learning orientation because attributes such as high motivation and capability are more common among students from high-SES families. Consequently, a student body's SES composition is a school resource that fosters a learning orientation, and it is highly relevant for our study.

Along with SES composition, we use a comprehensive set of control variables, including 4th-grade test scores, as a measure of prior performance. Table 1 describes these variables together with descriptive statistics. All independent continuous variables are standardized to have a mean of zero and a standard deviation of one across the combined sample of males and females in both datasets.

The Forschungsdatenzentrum at the IQB provides five imputed versions of the ELEMENT dataset (see Lehmann and Lenkeit 2008). We performed each analysis separately for the five imputed datasets and then combined the different estimates to obtain the final results presented here. We employed a similar imputation strategy based on the chained equations approach for the PISA-I-Plus dataset.

RESULTS

Variation of the Gender Gap across Schools

In an average school, the female advantage in reading scores is about .12 standard deviations in 5th grade and .21 standard deviations in 9th grade. It ranges from –.04 to .28 standard deviations in 5th grade and from .07 to .35 standard deviations in 9th grade for 95 percent of schools. Expressed in terms of years of education, girls are .36 school years ahead in 5th-grade reading test scores in an average school, but the gap ranges across schools from a male advantage of .12 years to a female advantage of .83 years.[13] Figure 2 plots this variation in the gender gap on the school level against average performance at a school. The striking pattern in the figure indicates that schools with higher than average performance also have the smallest gender gap. This finding is consistent with our theoretical prediction; it suggests that boys do not fall as far behind in performance-oriented schools.

SES Composition and the Gender Gap in Education

Table 2 presents estimates from school-level fixed-effect regression of reading test scores in 5th grade on classroom-level SES composition, 4th-grade scores, and other control variables. The table shows the main effects of gender and SES composition on the classroom level together with the interaction between SES composition and gender (all coefficients are in standard deviation units).

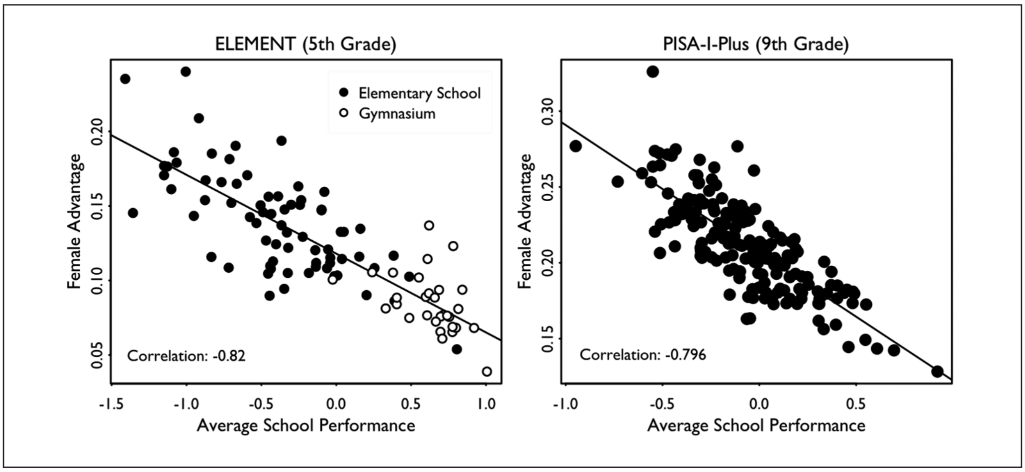

Figure 2. Gender Gap and Average Performance across Schools in Standard Deviation
Note: Estimates shown in the figure are based on a multilevel model with two levels (student and schools) and with a random intercept and a random slope for female on the school level so that average performance and effect of gender are allowed to vary across schools. Dots represent empirical Bayes predictions for the random intercept (i.e., average school performance) against the prediction for the random slope (i.e., the female advantage).

Table 2. Effect of SES Composition for Boys and Girls in Standard Deviations

Model	Prior Performance	Female Coef.	Female (se)	SES Comp. Coef.	SES Comp. (se)	SES Comp. x Female Coef.	SES Comp. x Female (se)
1. FE Estimate (ELEMENT)	yes	.007	(.02)	.091*	(.04)	−.060**	(.02)
2. FE Estimate (ELEMENT)	no	.120***	(.03)	.178***	(.06)	−.057*	(.03)
3. FE Estimate (PISA-I-Plus 2003)	no	.196***	(.03)	.237***	(.03)	−.052*	(.02)
4. MLM Estimate (PISA-I-Plus 2003)	no	.143	(.11)	.303***	(.05)	−.099*	(.04)

Note: FE = fixed effect. Table 1 describes control variables. The full set of coefficient estimates for Models 1 and 2 are in Table S1 in the online supplement. The number of students for models based on ELEMENT is 4,372, the number of schools is 101, and the average number of students per school is 43.3. *N* for PISA-I-Plus is 8,559.
* $p < .05$; ** $p < .01$; *** $p < .001$ (two-tailed tests); standard errors adjusted for clustering on class level.

Other coefficients are omitted from the table (for the full regression results, see Table S1 in the online supplement). The table also shows fixed-effect (FE) estimates from the PISA-I-Plus data for 9th-grade reading test scores without a measure of prior performance and estimates from a multilevel (MLM) model on the school level with a broad set of control variables. We include MLM estimates as a comparison because they reflect one of the most common estimation strategies (i.e., conditioning on observable covariates) used in sociology to identify compositional peer effects (e.g., Rumberger and Palardy 2005).

Results in Table 2 show that SES composition has a positive and highly significant effect on reading test scores in all models for gain scores (top row) and raw scores. This result conforms with previous findings in the literature on effects of SES composition (Jencks and Mayer 1990; Rumberger and Palardy 2005). In all models, the point estimate for the interaction between SES composition and female is negative and significant. Most important, estimates from the fixed-effect model using the ELEMENT data along with a control variable for prior performance show that boys learn more in classes with higher average SES. Adding additional peer characteristics, such as the proportion of foreign-born students, to this specification does not affect this finding (results not shown here). Results from the two FE models based on ELEMENT and PISA-I-Plus data without 4th-grade performance show the same results (we include ELEMENT results for direct comparison). In particular, the main effect of SES composition in the model based on PISA-I-Plus data seems to be upwardly biased (.237 compared to .178), and both estimates are somewhat larger than the .15 effect size estimated by Crosnoe (2009). However, the estimated size of the interaction between female and SES composition is very similar across the three fixed-effect models. This finding supports our argument that even if students self-select into classes (and self-selection appears to be more important in 9th grade), boys and girls are unlikely to differ in this selection process, which increases our confidence in the ELEMENT estimates. Results from the MLM model point in the same direction but appear to be upwardly biased. In particular, the estimate for the interaction is about 90 percent higher in the MLM model compared to the corresponding FE model. This could reflect the fact that the MLM estimate is based on non-random school-level variation, while the fixed-effect estimate is based on almost-random classroom-level variation within schools. The school-based estimate's larger size might also reflect spillover effects between SES composition of two classrooms in the same school. Given the possibility of selection bias in the MLM estimates, we consider the fixed-effects classroom-based estimates to be a more definitive test of our theoretical prediction.

Overall, our estimates provide strong evidence that boys are more sensitive than girls to the important school resource of classroom SES composition. Our statistical evidence is strengthened by the fact that institutional, simulation-based, and qualitative evidence indicates that randomness plays a central role in the allocation of students to 5th-grade classrooms in Berlin.

Targeted Sensitivity Analysis

In this section, we investigate whether our results are sensitive to three potential selection biases documented in interviews with school principals. Our detailed knowledge about the assignment process allows us to design a set of sensitivity analyses based on instrumental variables (IV) and certain sample restrictions targeted to address these potential biases. The FE model specified in Equation 1 and shown in the top row of Table 2 serves as the starting point. Table 3 presents results from the different sensitivity analysis and repeats estimates from the school FE model based on ELEMENT data for direct comparison.

The first selection process documented in the interviews refers to the non-random assignment of students who have repeated a grade to specific classrooms. While all principals reported that classroom size plays an important role, some principals also mentioned that they take potential implications for classroom culture into account. To address this potential selection problem, we treat SES composition on the class level as endogenous and instrument it with the average SES of the subset of students who never repeated a grade. This instrument is highly correlated with the total composition (the treatment indicator), and is arguably not affected by potentially non-random selection of grade repeaters because it is based only on students who never repeated a grade. The instrument should only be connected with the outcome through the actual class composition (i.e., it satisfies the exclusion restriction). Model 1 in Table 3 presents results and shows that the interaction between SES composition and female remains negative and significant. This indicates that

Table 3. Sensitivity Analysis

	Female		SES Comp.		SES Comp. x Female	
Model	Coef.	(se)	Coef.	(se)	Coef.	(se)
FE Estimate (full sample)	.007	(.02)	.091*	(.04)	–.060**	(.02)
(1) FE/IV Estimate Instrument: SES comp. of students who never repeated a class	.008	(.02)	.108*	(.05)	–.066**	(.02)
(2) FE/IV Estimate Instrument: SES comp. of students who did not go to kindergarten, skipped a class, or transferred to school	.009	(.02)	.113	(.06)	–.068*	(.03)
(3) FE Estimate (restricted sample) Sample restriction: only schools that do not allocate based on ethnicity (24 schools excluded)	.008	(.03)	.117*	(.05)	–.052*	(.03)

Note: $N = 4{,}372$. First stage results show that the two instruments are highly correlated with SES composition (i.e., the treatment). F-statistics are over 700 (highly significant), which is far above the commonly used threshold of 10. Table 1 describes control variables.
* $p < .05$; ** $p < .01$; *** $p < .001$ (two-tailed tests); standard errors adjusted for clustering on class level.

selection of students who repeat a grade into specific classes does not significantly bias the estimated effects.

The second potential selection process is assignment of students to the same class who attended the same kindergarten or who were friends before entering school. Using a similar strategy as in the last sensitivity analysis, we instrument peer SES by SES composition calculated for the subset of students who either did not attend kindergarten or who skipped a grade or transferred from another school. This set of students was certainly not assigned to classrooms based on the kindergarten criterion, and students who skip a class or transfer from a different school are most likely assigned to classrooms based on the number of students in different classrooms. For these reasons, the instrument is unaffected by the kindergarten criteria and (for the most part) by friendship self-selection. Results, presented in Model 2 of Table 3, again support our previous finding and indicate that the estimated causal effect is not sensitive to selection of connected students (either through the same kindergarten or through friendship) into the same class.

Finally, some principals reported that, in violation of school regulations, they assign students with migration backgrounds to the same class. To address this potential selection bias, we estimated the fixed-effect model reported earlier on a restricted sample. For this purpose, we assessed which schools allocate students with migration backgrounds non-randomly to classes, and we exclude these schools from the analysis.[14] Results, presented in Model 3 of Table 3, show that self-selection of students with migration backgrounds into specific classrooms in some schools does not affect our results.

Overall, results from the targeted sensitivity analyses specifically designed to address the potential selection processes identified in the interviews provide strong evidence that our estimates of gender-specific effects of classroom composition are not biased by these selection processes.

Explaining the Observed Difference in the Causal Effect between Boys and Girls

The theoretical argument presented earlier suggests that school context plays an important

Table 4. Effects of Gender and SES Composition on School-Related Attitudes and Behavior

		Female		SES Comp.		SES Comp. x Female	
		Coef.	(se)	Coef.	(se)	Coef.	(se)
Panel A	Attitudes toward school	.300***	(.04)	.057	(.07)	–.075*	(.04)
	Learning orientation	.131***	(.04)	.044	(.07)	–.033	(.03)
	Work habits	.166***	(.04)	.147*	(.07)	–.086*	(.04)
Panel B	Self-evaluation reading	.140***	(.04)	–.098	(.06)	–.028	(.03)
	Self-evaluation German	.207***	(.04)	.012	(.08)	–.056	(.03)
	Self-evaluation general	–.294***	(.04)	–.020	(.07)	–.025	(.03)

Note: $N = 4,372$. Table 1 describes control variables.
*$p < .05$; **$p < .01$; ***$p < .001$ (two-tailed tests); standard errors adjusted for clustering on class level.

role for the size of the gender gap. An academically oriented environment in schools with high-SES peers shapes how masculinity is constructed; this suppresses boys' negative attitudes toward school, facilitates their commitment, and enhances students' incentives to be engaged with academics. Other mechanisms, however, may account at least in part for the observed difference in the causal effect of SES composition for male and female students.

The literature on compositional school and classroom effects offers an alternative explanation for the relationship between SES composition and student performance, which focuses on social comparison processes (Jencks and Mayer 1990; Rumberger and Palardy 2005; Thrupp, Lauder, and Robinson 2002). This alternative account argues that students use their classmates as a reference group to evaluate their own performance and thereby develop academic self-perceptions that in turn may affect their performance (Crosnoe 2009; Dai and Rinn 2008). To adjudicate between our proposed explanation and this alternative account, we estimate models based on ELEMENT data that are identical to the school-level fixed-effects regression described in Equation 1, but that replace the reading score dependent variable with measures of student attitudes, student behavior, and self-perception about academic ability.[15] Our core hypothesis implies that class environment has a more pronounced effect on attitudes toward school, learning orientation, and academic effort for boys than for girls. Accordingly, a higher positive effect of SES composition on these outcomes for boys than for girls would provide further evidence for this mechanism. An explanation for gender differences based on reference group processes, however, would imply that a class's SES composition affects boys' and girls' academic self-perceptions differently. In other words, this alternative account suggests that boys and girls react differently to their reference group.

Table 4 shows results from school-level fixed-effect models of the indicated variables on classroom socioeconomic composition, controlling for variables described in Table 1. Panel A, which reports regression results using attitudes toward school, learning orientation, and work habits as dependent variables, provides further evidence for our core hypothesis. Point estimates for SES composition and the interaction with female are not all significant but consistently point in the expected direction. This pattern of results suggests that boys' attitudes toward school, their learning orientation, and their work

Table 5. Fixed-Effects Models with School-Related Attitudes and Behavior

	Model 1		Model 2		Model 3	
Variable	Coef.	(se)	Coef.	(se)	Coef.	(se)
Female	.007	(.02)	–.009	(.03)	–.009	(.03)
SES Composition	.091*	(.04)	.037	(.04)	.033	(.04)
SES Composition x Female	–.060**	(.02)	–.061**	(.02)	–.040*	(.02)
Attitude Toward School			.041*	(.01)	.040*	(.02)
Learning Orientation			.019	(.01)	.017	(.02)
Work Habits			.072***	(.01)	.092***	(.01)
Attitude Toward School x Female					–.047*	(.02)
Learning Orientation x Female					.002	(.02)
Work Habits x Female					–.059***	(.02)
Control Variables	yes		yes		yes	
Constant	–.673***	(.35)	–.538	(.27)	–.533*	(.25)

Note: N = 4,372. Table 1 describes control variables.
*$p < .05$; **$p < .01$; ***$p < .001$ (two-tailed tests); standard errors adjusted for clustering on class level.

habits are more sensitive to the school environment than are girls' attitudes and work habits. Panel B, by contrast, reports small and insignificant interaction effects between gender and social classroom composition on self-evaluations of performance in reading, performance in German, and performance "in general." The lack of gender differences in the effect of SES composition on self-perceptions of ability favors our preferred explanation over the alternative account based on reference group processes.

We further extend this examination of mechanisms by building on the initial FE model for 5th-grade performance (defined in Equation 1), and add school-related attitudes and behavior as independent variables in a stepwise fashion. Compared to the models presented so far, the elaborated model is less rigorous from a causal point of view because the causal ordering of performance and school-related attitudes and behavior is not clear-cut. Nonetheless, it can be informative about potential mechanisms. Results in Table 5 suggest that the effect of SES composition is clearly reduced by the addition of variables for school-related attitudes and behavior (Model 2). They also suggest that part of the gender difference in the effect of SES composition (33 percent) may be explained by its gender-specific effect on school-related attitudes and behavior; this provides further support for our proposed mechanism.

Finally, we investigate the possibility that boys benefit from a stronger academic peer culture not because they are boys, but rather because underperforming students benefit in general, and boys are a disproportionate fraction of underperforming students. Accordingly, we again extend the model described in Equation 1 by adding an interaction term between performance in 4th grade (the year prior to our measured outcomes in the regressions) and SES composition in 5th grade. Results (available from the authors) show that the impact of SES composition is significantly stronger for low-performing students, which is in line with findings from other studies (Bryk, Lee, and Holland 1993; Coleman 1966, 1970). Inclusion of this interaction also weakens the direct benefit of being male in a high-SES class by about 27 percent (from –.060 to –.044). However, the interaction between SES composition and gender remains statistically significant (p = .021) and substantively important. Results suggest that boys do indeed benefit indirectly from a stronger academic climate because they are disproportionately low-performing students. Nonetheless, the bulk of the effect stems from boys' greater sensitivity than girls to classrooms' academic orientation.

DISCUSSION

Throughout the industrialized world, girls have made dramatic gains in educational attainment, while boys' underperformance and their tendency to disrupt the learning process have sparked intense academic and public debates about the causes of what many now call the "problem with boys." Some scholars and pundits blame schools for fostering a de-masculinized learning environment. Yet, the role of school context and the connection between school resources and the gender gap is underdeveloped in the literature to date. In this article, we have extended research on schools' effect on class and race inequality by asking whether schools affect gender inequality as well, and if so, what are the mechanisms by which this occurs.

Building on theories about gender identity, adolescent culture, and prior ethnographic classroom observations, we developed a theoretical argument about the role of environmental factors for the educational gender gap and boys' underachievement. In particular, we argue that school and class environments shape how masculinity is constructed in peer culture and thereby influence boys' orientation toward school. Resources that create a learning-oriented environment raise the valuation of academics in adolescent male culture and facilitate commitment. Girls' peer groups, by contrast, do not vary as strongly with the social environment in the extent to which they encourage academic engagement, and they are less likely to stigmatize school engagement as un-feminine. As a consequence, boys differentially benefit from these school resources, and the female advantage in test scores shrinks in higher quality schools.

Results from our analysis of German ELEMENT and PISA-I-Plus 2003 data provide clear support for this hypothesis. We find substantial variation in the gender gap in academic performance across schools and that this variation is related to average school performance. We then used a quasi-experimental research design to establish that boys are more sensitive to peer SES composition as an important dimension of school quality related to the learning environment. This quasi-experimental research design is based on the argument that randomness plays an important role for student assignment to classes within Berlin's elementary and 5th-grade upper-secondary schools. To evaluate this argument, we examined Berlin's school regulations, compared observed classroom composition with simulations of random assignment, and conducted qualitative interviews with school principals in Berlin. Findings from this evaluation of the selection process generally support our argument but also point at potential biases, which we addressed with targeted sensitivity analyses. Results from these analyses show little effect of potential selection biases on our core results. In addition, we considered alternative mechanisms that might explain the observed difference in the causal effect between boys and girls. Results from this analysis provide further support for our own explanation. They suggest that boys benefit indirectly (because low-performing students benefit in general) and directly (because the effect is bigger for boys than for girls) from being in a classroom with high-SES composition.

Our findings contribute to several areas of research. First, our study makes an important contribution to the debate about boys' well-publicized underperformance. The outlined cultural mechanism explains why boys are more sensitive to human and cultural capital resources in schools, which plays an important role for their underperformance and the gender gap in educational achievement. This argument suggests that boys' resistance to school is not purely a function of either their class background—as many studies suggest—or the fact of their masculinity—as other research suggests—but instead depends on schools' and classrooms' local cultural environment. As such, the findings broaden our understanding of boys' notorious underperformance. Results point at an important mechanism connected to how school and class environments shape boys' and girls' learning orientations, and in the process reveal a pattern similar to what has been found in families (Buchmann and DiPrete 2006). In both cases, boys seem to be more sensitive to the level of resources in the local environment, so that the size of the gender gap is a function of environmental resources.

Second, our results point to useful directions for new research on policies to raise boys' achievement levels. It is obviously important to know that boys respond especially positively to an academic orientation among their peers. However, while local governments could decide to invest more resources in their schools, they cannot, as a practical matter, produce more high-SES children for their school systems. An important unanswered question raised by our research concerns whether schools can accomplish the same cultural enrichment through alternative means. The most obvious alternative resource would be better teachers. Teachers directly influence schools' academic environment and can raise academic performance. They have the potential to modify student behavior and produce a stronger academic student culture, even without socioeconomic enrichment of a school's student body. At present, however, too little is known about what makes a quality teacher, or the extent of the effect of better teachers on higher academic performance and the academic climate. Our research suggests, for example, that teaching methods that emphasize academic competition are particularly beneficial for boys. These are important questions for future research.

Finally, we make a methodological contribution to the literature on estimation of causal effects. Our work illustrates how a detailed study of the relevant selection process—in our case, examination of official regulations, statistical simulations, and qualitative interviews—can facilitate the estimation of causal effects. This detailed understanding of the actual selection process not only allows researchers to evaluate the extent of bias but also enables the design of targeted sensitivity analysis (in our case, based on instrumental variables and sample restrictions). Overall, we believe that knowledge about the selection process can help researchers improve the accuracy of causal effect estimates. Considering these benefits, we invite sociologists to take selection processes seriously as an independent object of study—an argument previously made by Sampson (2008:189), who conceptualizes "selection bias as a fundamental social process worthy of study in its own right rather than a statistical nuisance" (for an earlier statement of this argument, see DiPrete 1993).

Our findings are limited in some regards. Most important, our theoretical argument applies to all kinds of school resources that create a learning-oriented environment. Our empirical analysis, however, only focuses on one (although important) dimension, peer socioeconomic composition. Given this limitation, future studies should establish the extent to which conclusions from this study apply to other kinds of school-based resources. Additionally, due to lack of adequate data, our study neglects teachers' role in shaping boys' and girls' learning orientations. While our interviews indicate that teachers are not assigned to classrooms based on classroom composition, teachers might react to classroom dynamics in certain ways that play an important role for the processes studied here. Finally, our study focuses on only one major dimension of cognitive achievement, reading. On average, boys do as well or better than girls in mathematics, and the male advantage is larger on the right tail of the distribution. Whether boys gain a stronger advantage than girls from being in a classroom with higher mean SES, or whether their special advantage occurs only for academic subjects where they otherwise lag behind girls, is an important question for future research.

APPENDIX

Part A. Education and the Educational Gender Gap in Germany

Although our main focus here is the theoretical argument, background information can help contextualize findings from the German case. In Germany, children usually attend elementary school from age 6 until age 10 or 12, depending on state (Bundesland) regulations. After finishing elementary school, students transfer to secondary schools, which are distinct from U.S. middle and high schools because of performance-based tracking on the school level. Although the system has become more differentiated in recent decades, three school types have traditionally been of great importance. The *Gymnasium* is the highest secondary school type, the *Realschule* is for intermediate students, and the *Hauptschule* is the low secondary school track. As a response to critiques of this tripartite secondary school system, some states have introduced comprehensive schools that either integrate all three school tracks or just the Haupt- and Realschule (*Gesamtschule* and *Schule mit mehreren Bildungsgängen*). After finishing secondary school, students have the option to obtain a higher education degree, to continue their education in one of the vocational programs (which figure importantly in the German educational system), or to enter the labor market immediately. Overall, the German educational system is distinct from the U.S. system and other countries primarily because of the early school-based tracking in secondary school, the strong vocational track as an alternative to higher education, and the limited role of the federal government, which is evident in the many differences in the specific structure of German schools across German states. Similar to other industrialized countries, the gender gap in Germany has closed over the past decades. Legewie and DiPrete (2009) emphasize, however, that the female advantage in higher education is less pronounced than in the United States due in large part to women's failure to converge with men in rates of obtaining degrees from *Fachhochschulen* (universities of applied sciences).

Part B. Simulation of Random Assignment

Our simulation-based approach allows us to evaluate whether within-school variation in the composition of classes is consistent with a random allocation process. To compare observed composition with composition obtained under complete randomization, we proceed in the following way. For each school, we randomly allocate students to classrooms in the school they attend, keeping the number and size of classrooms constant. We then compare socioeconomic composition across classes obtained from the simulation with the observed composition. Accordingly, the simulation evaluates whether the actual (unknown) allocation process is consistent with a completely randomized classroom assignment. The statistic to compare the actual and simulated distribution for some variable x (e.g., SES, migration background, or gender) for classroom k in school j is defined as the average square deviation of the classroom means from the school mean

$$t_j = \frac{1}{n_j}\sum_{k=1}^{n_j}\left(\bar{x}_{jk} - \bar{x}_j\right)^2$$

where j and k are indices for schools and classrooms, respectively; $\bar{x}_j$ is the average for school j; $\bar{x}_{jk}$ is the average for classroom k in school j; and n_j is the number of classrooms in school j. If the number of students is the same in each classroom within a school, this measure is simply the variance of the class specific means in a school.

Acknowledgments

We thank Claudia Buchmann, Jennifer L. Jennings, Merlin Schaeffer, and participants at seminars at the Mailman School, Columbia University, the University of Wisconsin, Madison, and the 2010 RC28 conference in Haifa, Israel for helpful comments and suggestions. We thank the Institut zur Qualitätsentwicklung im Bildungswesen (IQB) for providing access to data and Prof. Rainer Lehmann for his help with ELEMENT data.

Notes

1. Of course, stereotypical gender identities also affect girls. Correll (2001), for example, shows how cultural beliefs about gender can bias women's self-perceptions of math ability, controlling for actual performance, and thereby deter women from careers in science, math, or engineering.

2. These assertions do not imply that girls are always engaged in the learning process. Many studies document how girls resist teachers and school (e.g., Francis 2000). Nevertheless, one of the most common findings in ethnographic studies is that boys more actively resist the learning process.
3. Our expectations mainly relate to wealthy OECD countries because prior research finds that the role of school context (Chudgar and Luschei 2009) and gender relations differ substantially between high- and low-income countries.
4. In contrast to most other states in Germany, students in Berlin usually attend elementary school until 6th grade, so the 31 5th-grade upper-secondary schools—the *grundständige Gymnasien*—are different from normal secondary schools.
5. In Berlin, elementary school students assigned to the same classroom take virtually all their classes together, so we use the terms "classroom" and "class" interchangeably in the text.
6. For 5th-grade upper-secondary schools in ELEMENT, class assignment occurs in 5th grade because students transfer from elementary school after 4th grade.
7. We obtained both datasets from the Forschungsdatenzentrum at the Institute für Qualitätsentwicklung im Bildungswesen (IQB) HU-Berlin.
8. As a substantive matter, 5th-grade culture differs from 9th-grade culture in the obvious sense that 5th-grade students are pre-adolescent while 9th-grade students have generally passed through puberty. At the same time, studies of childhood and adolescent culture find continuity in the emerging masculine culture between middle childhood and high school (Maccoby 1998; Thorne 1993). Thus, for substantive and methodological reasons, we expect comparison of results from 5th and 9th grades to be informative about our core hypothesis.
9. The three-level data structure might imply that error terms of students in the same classroom are correlated even after controlling for school fixed-effects. We address this problem by correcting the standard error for clustering on the class level using the Moulton factor (Angrist and Pischke 2008).
10. Although the PISA-I-Plus is a panel study and collected achievement data in 9th and 10th grades, the panel component of these data is not yet available.
11. In addition, all schools reported that class changes within a grade level are extremely rare, and resources are generally allocated equally across classes.
12. We explored alternative specifications of SES composition effects, such as allowing separate effects of SES composition of male and female peers. These alternative specifications yield essentially the same results as those reported in the tables.
13. One additional school year corresponds to the estimated test score difference between 5th and 6th grade in the ELEMENT dataset.
14. We use a simple z-test to identify schools in which the difference in the proportion of students with migration backgrounds between classes is higher than what we would expect under randomness. Using conservative criteria, we exclude schools with a p-value smaller than .1 (24 schools).
15. We constructed the measures from a range of indicators using exploratory factor analysis (see the online supplement).

References

Ammermueller, Andreas and Jörn-Steffen Pischke. 2009. "Peer Effects in European Primary Schools: Evidence from the Progress in International Reading Literacy Study." *Journal of Labor Economics* 27:315–48.

Angrist, Joshua D. and Jörn-Steffen Pischke. 2008. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ: Princeton University Press.

Bishop, John H., Matthew Bishop, Lara Gelbwasser, Shanna Green, and Andrew Zuckerman. 2003. "Nerds and Freaks: A Theory of Student Culture and Norms." *Brookings Papers on Education Policy* 6:141–99.

Bryk, Anthony, Valerie Lee, and Peter Holland. 1993. *Catholic Schools and the Common Good*. Cambridge, MA: Harvard University Press.

Buchmann, Claudia and Thomas A. DiPrete. 2006. "The Growing Female Advantage in College Completion: The Role of Family Background and Academic Achievement." *American Sociological Review* 71: 515–41.

Buchmann, Claudia, Thomas A. DiPrete, and Anne McDaniel. 2008. "Gender Inequalities in Education." *Annual Review of Sociology* 34:319–37.

Campbell, Jay R., Dana L. Kelly, Ina V. Mullis, Michael O. Martin, and Marian Sainsbury. 2001. *Framework and Specification for PIRLS Assessment 2001*. Chestnut Hill, MA: International Study Center, Lynch School of Education, Boston College.

Chudgar, Amita and Thomas F. Luschei. 2009. "National Income, Income Inequality, and the Importance of Schools: A Hierarchical Cross-National Comparison." *American Educational Research Journal* 46:626–58.

Cohen, Michele. 1998. "A Habit of Healthy Idleness: Boys Underachievement in Historical Perspective." Pp. 19–33 in *Failing Boys? Issues in Gender and Achievement,* edited by D. Epstein, J. Elwood, V. Hey, and J. Maw. Buckingham, UK: Open University Press.

Coleman, James S. 1961. *The Adolescent Society: The Social Life of the Teenager and its Impact on Education*. New York: Free Press.

Coleman, James S. 1966. *Equality of Educational Opportunity*. Washington, DC: U.S. Dept. of Health, Education, and Welfare, Office of Education.

Coleman, James S. 1970. "Reply to Cain and Watts." *American Sociological Review* 35:242–49.

Correll, Shelley J. 2001. "Gender and the Career Choice Process: The Role of Biased Self-Assessments." *American Journal of Sociology* 106:1691–1730.

Crosnoe, Robert. 2009. "Low-Income Students and the Socioeconomic Composition of Public High Schools." *American Sociological Review* 74:709–730.

Dai, David Yun and Anne N. Rinn. 2008. "The Big-Fish-Little-Pond Effect: What Do We Know and Where Do We Go from Here?" *Educational Psychology Review* 20:283–317.

Davies, Bronwyn. 2003. *Frogs and Snails and Feminist Tales: Preschool Children and Gender*. New York: Hampton Press.

DiPrete, Thomas A. 1993. "Discrimination, Choice, and Group Inequality: A Discussion of How Allocative and Choice-Based Processes Complicate the Standard Decomposition." *Social Science Research* 22:415–40.

Dresel, Markus, Heidrun Stöger, and Albert Ziegler. 2006. "Klassen- und Schuleffekte bei Geschlechtsunterschieden im schulischen Leistungsbereich." *Psychologie in Erziehung und Unterricht* 1:44–61.

Eitzen, D. Stanley. 1975. "Athletics in the Status System of Male Adolescents: A Replication of Coleman's 'The Adolescent Society.'" *Adolescence* 10:267–76.

Epstein, Debbie. 1998. "Real Boys Don't Work: Underachievement, Masculinity and the Harassment of Sissies." Pp. 96–108 in *Failing boys? Issues in Gender and Achievement,* edited by D. Epstein, J. Elwood, V. Hey, and J. Maw. Buckingham, UK: Open University Press.

Francis, Becky. 2000. *Boys, Girls and Achievement: Addressing the Classroom Issues*. London: Routledge Falmer.

Ganzeboom, Harry B. G., Donald J. Treiman, and Wout C. Ultee. 1991. "Comparative Intergenerational Stratification Research: Three Generations and Beyond." *Annual Review of Sociology* 17:277–302.

Gould, Eric D., Victor Lavy, and M. Daniele Paserman. 2009. "Does Immigration Affect the Long-Term Educational Outcomes of Natives? Quasi-Experimental Evidence." *The Economic Journal* 119:1243–69.

Greenwald, Rob, Larry V. Hedges, and Richard D. Laine. 1996. "The Effect of School Resources on Student Achievement." *Review of Educational Research* 66:361–96.

Hallinan, M. T. 1988. "Equality of Educational Opportunity." *Annual Review of Sociology* 14:249–68.

Hanushek, Eric A. 1989. "The Impact of Differential Expenditures on School Performance." *Educational Researcher* 18:45–62.

Howe, Christine. 1997. *Gender and Classroom Interaction: A Research Review*. Glasgow: The SCRE Centre.

Hoxby, Caroline. 2000. *Peer Effects in the Classroom: Learning from Gender and Race Variation*. NBER Working Paper No. 7867.

Imberman, Scott, Adriana Kugler, and Bruce Sacerdote. 2009. "Katrina's Children: A Natural Experiment in Peer Effects from Hurricane Evacuees." NBER Working Paper No. 15291.

Jencks, Christopher and Susan E. Mayer. 1990. "The Social Consequences of Growing Up in a Poor Neighborhood." Pp. 111–86 in *Inner-City Poverty in the United States,* edited by L. E. Lynn and M. G. H. McGeary. Washington, DC: National Academy Press.

Jennings, Jennifer L. and Thomas A. DiPrete. 2010. "Teacher Effects on Social/Behavioral Skills in Early Elementary School." *Sociology of Education* 83:135–59.

Jovanovic, Jasna and Sally Steinbach King. 1998. "Boys and Girls in the Performance-Based Science Classroom: Who's Doing the Performing?" *American Educational Research Journal* 35:477–96.

Kahlenberg, Richard D. 2001. *All Together Now: Creating Middle-Class Schools through Public School Choice*. Washington, DC: Brookings Institution Press.

Kao, Grace, Marta Tienda, and Barbara Schneider. 1996. "Racial and Ethnic Variation in Academic Performance." *Research in Sociology of Education and Socialization* 11:263–97.

Legewie, Joscha and Thomas A. DiPrete. 2009. "Family Determinants of the Changing Gender Gap in Educational Attainment: A Comparison of the U.S. and Germany." *Schmollers Jahrbuch (Journal of Applied Social Science Studies)* 129:1–13.

Lehmann, Rainer and Jenny Lenkeit. 2008. *ELEMENT: Abschlussbericht über die Untersuchungen 2003, 2004 und 2005 an Berliner Grundschulen und grundständigen Gymnasien*. Berlin: Humboldt Universität zu Berlin.

Mac an Ghaill, Mairtin. 1994. *The Making of Men: Masculinities, Sexualities and Schooling*. Buckingham, UK: Open University Press.

Maccoby, Eleanor E. 1998. *The Two Sexes: Growing up Apart, Coming Together*. Cambridge, MA: Belknap Press of Harvard University Press.

Machin, Stephen and Sandra McNally. 2005. "Gender and Student Achievement in English Schools." *Oxford Review of Economic Policy* 21:357–72.

MacLeod, Jay. 2008. *Ain't No Makin' It: Aspirations and Attainment in a Low-Income Neighborhood*. Oxford, UK: Westview Press.

Morris, Edward W. 2008. "'Rednecks,' 'Rutters,' and 'Rithmetic." *Gender & Society* 22:728–51.

Morse, Susan. 1998. *Separated by Sex: A Critical Look at Single-Sex Education for Girls*. Washington, DC: American Association of University Women Educational Foundation.

Murnane, Richard J. 1983. "Quantitative Studies of Effective Schools: What Have We Learned?" Pp. 193–209 in *School Finance and School Improvement,* edited by A. Odden and L. D. Webb. Cambridge: Ballinger.

Pickering, Jon. 1997. *Raising Boys' Achievement*. London: Network Continuum Education.

PISA-Konsortium Deutschland. 2006. *PISA 2003: Untersuchungen zur Kompetenzentwicklung im Verlauf eines Schuljahres*. Münster: Waxmann Verlag.

Power, Sally, Geoff Whitty, Tony Edwards, and Valerie Wigfall. 1998. "Schoolboys and Schoolwork: Gender Identification and Academic Achievement." *International Journal of Inclusive Education* 2:135–53.

Quenzel, Gudrun and Klaus Hurrelmann. 2010. "Geschlecht und Schulerfolg: Warum wird die schulische Leistungsbilanz von jungen Männern immer schlechter?" *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 62:61–91.

Ravitch, Diane. 2010. *The Death and Life of the Great American School System: How Testing and Choice Are Undermining Education*. New York: Basic Books.

Rockoff, Jonah E. 2004. "The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data." *American Economic Review* 94:247–52.

Rumberger, Russell W. and Gregory J. Palardy. 2005. "Does Segregation Still Matter? The Impact of Student Composition on Academic Achievement in High School." *Teachers College Record* 107:1999–2045.

Sadker, David and Karen Zittleman. 2009. *Still Failing at Fairness: How Gender Bias Cheats Girls and Boys in School and What We Can Do About It*. New York: Scribner.

Salisbury, Jonathan and David Jackson. 1996. *Challenging Macho Values: Practical Ways of Working with Adolescent Boys*. London: Falmer Press.

Salomone, Rosemary C. 2003. *Same, Different, Equal: Rethinking Single-Sex Schooling*. New Haven, CT: Yale University Press.

Sampson, Robert J. 2008. "Moving to Inequality: Neighborhood Effects and Experiments Meet Social Structure." *American Journal of Sociology* 114:189–231.

Schneider, Barbara L. and Venessa A. Keesler. 2007. "School Reform 2007: Transforming Education into a Scientific Enterprise." *Annual Review of Sociology* 33:197–217.

Schöps, Katrin, Oliver Walter, Karin Zimmer, and Manfred Prenzel. 2004. "Disparitäten zwischen Jungen und Mädchen in der mathematischen Kompetenz." In *PISA 2003: Der Bildungsstand der Jugendlichen in Deutschland- Ergebnisse des zweiten internationalen Vergleichs,* edited by PISA-Konsortium Deutschland. Münster: Waxmann.

Sewell, William H., Archibald O. Haller, and Alejandro Portes. 1969. "The Educational and Early Occupational Attainment Process." *American Sociological Review* 34:82–92.

Skelton, Christine. 1997. "Primary Boys and Hegemonic Masculinities." *British Journal of Sociology of Education* 18:349–69.

Sørensen, Aage B. and Stephen L. Morgan. 2006. "School Effects: Theoretical and Methodological Issues." Pp. 137–60 in *Handbook of the Sociology of Education,* edited by M. T. Hallinan. New York: Springer.

Spender, Dale. 1982. *Invisible Women: The Schooling Scandal*. London: Writers and Readers.

Stanworth, Michelle. 1984. *Gender and Schooling: A Study of Sexual Divisions in the Classroom*. London: Routledge.

Steinberg, Laurence D., L. Steinberg, Benson Bradford Brown, and Sanford M. Dornbusch. 1997. *Beyond the Classroom: Why School Reform Has Failed and What Parents Need to Do*. New York: Simon and Schuster.

Thorne, Barrie. 1993. *Gender Play: Girls and Boys in School*. Piscataway, NJ: Rutgers University Press.

Thrupp, Martin, Hugh Lauder, and Tony Robinson. 2002. "School Composition and Peer Effects." *International Journal of Educational Research* 37:483–504.

Whitmire, Richard. 2010. *Why Boys Fail: Saving Our Sons from an Educational System That's Leaving Them Behind*. New York: AMACOM.

Willis, Paul. 1981. *Learning to Labor: How Working Class Kids Get Working Class Jobs*. New York: Columbia University Press.

Younger, Michael, Molly Warrington, and Jacquetta Williams. 1999. "The Gender Gap and Classroom Interactions: Reality and Rhetoric?" *British Journal of Sociology of Education* 20:325–41.

Joscha Legewie is a PhD student in the Department of Sociology at Columbia University. His dissertation research focuses on how school and peer environment shape gender inequalities in education using two quasi-experimental case studies from Germany and the United States. Other research looks at the persisting gender gap in science and engineering education, the role of peer relations and social support for academic work in school, and the effect of terrorist events on the perception of immigrations using a natural experiment.

Thomas A. DiPrete is Giddings Professor of Sociology at Columbia University. His research interests include social stratification, demography, education, economic sociology, and quantitative methodology. Recent research projects include causes of the widening gender gap in higher education in favor of women, the persisting gender gap in science and engineering education, social polarization in the United States and its link with segregation in social networks along several potential dimensions of social cleavage, and the role of social comparison and cumulative advantage processes in the trend toward rising inequality at the top of the earnings distribution.