

School Context and the Gender Gap in Educational Achievement*

Joscha Legewie

(Columbia University)

Thomas A. DiPrete

(University of Wisconsin - Madison)

May 6, 2011

forthcoming in

American Sociological Review

Word count: 11,950 (without appendices and figure/table captions)

*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 the data and Prof. Rainer Lehmann for his help with the ELEMENT data. Please direct correspondence to Joscha Legewie, Columbia University, Department of Sociology – MC9649, 606 W 122nd Street, New York, NY 10027 (jpl2136@columbia.edu).

Abstract

Today, boys generally under-perform 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 the school environment channels the conception of masculinity in the peer culture, and thereby either fosters or inhibits the development of anti-school attitudes and behavior among boys. Girls' peer groups, in contrast, do not vary as strongly with the social environment in the extent to which school engagement is stigmatized as "un-feminine." As a consequence, boys are more sensitive to school resources that create a learning oriented environment than are girls. Our analyses use a quasi-experimental research design to estimate the gender difference in the causal effect on test scores, and focus on peer SES as an important school resource. We argue that assignment to 5th grade classrooms within Berlin schools is practically random, and we evaluate this selection process by an examination of Berlin's school regulations, by simulation analysis, and by 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 of the gender differences in the causal effect.

Introduction

Today, boys dominate among high school dropouts, special education students, and literally any failed or special needs category throughout adolescence. The notorious underperformance of boys in school and their tendency to disrupt the learning process in the classroom has sparked intense academic as well as 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 an allegedly de-masculinized learning environment and an alleged 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. Research on school effects was given a high profile by the 1966 Coleman report, and much of the attention since then has been motivated by 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 important to extend this line of research and ask whether schools affect gender inequality as well, and if so, what are the mechanisms by which this occurs.

Integrating theories about gender identity, adolescent culture, and the findings from prior ethnographic classroom observations, we argue that the school environment channels the conception of masculinity in the peer culture, and thereby either fosters or inhibits the development of anti-school attitudes and behavior among boys. An academically oriented environment suppresses a construction of masculinity as oppositional and instead facilitates boys’ commitment by promoting academic competition as an aspect of masculine identity. Lower quality schools, in contrast, implicitly encourage – or at least do not inhibit – the development of a peer culture that constructs resistance to both school and teacher as valued masculine traits. Girls’ peer groups, in contrast, do not vary as strongly with the social environment in the extent to which school engagement is stigmatized as “un-feminine.” As a result, boys benefit particularly 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 so-called ELEMENT data from one German state (the city-state of Berlin). In contrast to the US, the lack of performance-based tracking in Berlin elementary schools and the smaller extent of parents' influence on classroom assignment makes it plausible that student assignment to elementary school classrooms in Berlin is almost random. In order to develop a detailed understanding of the actual selection process, we examine the official school regulations, provide statistical evidence from simulation analyses, and conduct qualitative interviews with school principals. The results suggest that randomness indeed plays an important role in the assignment process, but 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 obtained from a large-scale nationally representative dataset from Germany (PISA-I-Plus 2003) to address potential concerns about the generalizability of the results.

The 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 is the source of the gender difference in the causal effect of SES composition on student achievement. Our findings speak to the recent political debate about the educational shortcomings of boys by deepening our understanding of their notorious under-performance. Our analytical strategy also makes a methodological contribution by illustrating how a detailed study of the selection process using simulations and qualitative interviews can assist the estimation of causal effects.

Educational Outcomes and Schools

The 1966 Coleman report (Coleman, 1966) claimed that, while family was the most important determinant of achievement, performance was improved 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 the adolescent culture rewards academic performance and thereby supports the reward system of parents and teachers. But when the adolescent culture values other behaviors more highly (e.g., sports, being popular with the opposite sex, or opposition to school authority), and especially when the adolescent culture denigrates academic achievement, it inhibits academic investment and weakens academic achievement. Simply put, students who are highly motivated and capable (attributes that are more common at higher SES levels) create a learning oriented peer culture (Sewell et al., 1969; Jencks and Mayer, 1990; Rumberger and Palardy, 2005, 125).

For about twenty years following the release of the Coleman report, the literature reported that school effects were relatively small in comparison with family effects, and therefore that “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 movements to improve schools in order to improve learning (Schneider and Keesler, 2007). Reanalysis of earlier studies suggested a more consistently positive relationship between school resources and student achievement (Greenwald et al., 1996), and found that teacher quality in particular was a major input into student learning (see also Murnane, 1983).

The renewed focus on the impact of schools 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 his-

torically appreciated, the estimation of peer effects is challenging (Angrist and Pischke, 2008, 192ff) because of non-random selection and unmeasured confounding variables (like teacher quality) that affect student outcomes. The most persuasive recent studies have used natural experiments to estimate the impact of changes in class composition on outcomes (e.g. Imberman et al., 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 has examined arguably random fluctuations in adjacent cohorts (e.g. of gender or race composition) for the same school and grade (Hoxby, 2000; Gould et al., 2009), though these studies have not looked at peer effects related to socioeconomic characteristics. Although the magnitude of estimated effects is not large (about 0.15 standard deviations), it is about the same as some of the most believable estimates of teacher effects, whether for academic, or social and behavioral outcomes (Rockoff, 2004; Jennings and DiPrete, 2010). Meanwhile, recent studies whose primary estimation strategy controls for observable potential confounders have found a similar effect size on test scores (Crosnoe, 2009; Rumberger and Palardy, 2005).

The 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 affect gender inequality as well, 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 the gendered behavior of girls and boys at school as well as the different ways that teachers treat girls and boys. Although the overt discrimination of girls in the classroom has declined over the past three decades, recent studies suggest that boys still ‘monopolize the linguistic space’ of the classroom (Jovanovic and King, 1998; Sadker and Zittleman,

2009). Meanwhile, the once celebrated coeducation of boys and girls as a pivotal step towards 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 (Salomone, 2003; Morse, 1998).

Despite these important strands of research and the general recognition that schools are an important context for the socialization of young adolescents, the literature on the educational gender gap has widely ignored the school as a potential source of variation in the educational gender gap. To our knowledge, Dresel et al. (2006), Schöps et al. (2004), and Machin and McNally (2005) 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 (Baden-Württemberg), Dresel et al. (2006) found substantial variation in the educational gender gap across schools and classes, while Schöps et al. (2004) obtained a similar finding using the German PISA data. Machin and McNally (2005), in 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 the reports from prior ethnographic classroom observations and theories about gender identity in order to understand the role of the school context for the under-achievement of boys.

The Under-Achievement of Boys, 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 the anti-school attitudes and behavior of working-class white boys 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 MacLeod, 2008; Kao et al., 1996). In line with Willis' early findings, much of the literature on the under-achievement of boys 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 the creation of stereotypical gender identities begin in early childhood before children have had any experience with school (Maccoby, 1998; Thorne, 1993; Davies, 2003). Gender-differentiated childhood cultures become the basis for gender-differentiated adolescent cultures, which are important influences on how children view school, on whether they take school seriously, and on how hard they work as students (Steinberg et al., 1996). Classroom observations and other ethnographic studies have documented the ways in which gender identities are constructed in the classroom and how these gender cultures affect interactions and the approach to education of boys and girls (Francis, 2000; Pickering, 1997; Salisbury and Jackson, 1996; Skelton, 1997). They show that boys tend to be noisier, more physically active, and more easily distracted than are girls (Spender, 1982; Younger et al., 1999; Howe, 1997; Francis, 2000). The studies also find that masculine stereotypes portray boys as competitive, active, aggressive, and dominating, while girls are viewed as conciliatory and cooperative (Francis, 2000, 48). Others have argued that stereotypical gender identities perpetuate the belief that girls have to work hard in order to learn in school, whereas boys are naturally gifted (Cohen 1998; Epstein, 1998; Power et al., 1998; Mac an Ghaill, 1994; Quenzel and Hurrelmann, 2010, 75ff). Cohen (1998) shows that these gendered beliefs are reflected in a casual and detached attitude towards school among boys, which accords with the other ethnographic studies referenced above. Despite the transformation of gender relations in modern societies, stereotypical gender identities continue to shape orientations towards school and produce behaviors that reinforce these identities while potentially affecting a child's academic success. This is illustrated in Morris' observations (2008, 736) at a rural high-school. He found that "girls tended to direct considerable effort and attention to school" whereas "boys [...] took pride in their lack of academic effort" (Morris, 2008, 736) as an aspect of their masculine identity.¹

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

Gender identities and gendered behavior patterns are reinforced by peers and the adolescent reward system. In some contexts, disruptive behavior produces status gains in the peer groups of lower SES students. Working for academic achievement, in contrast, is labeled as feminine and thereby stigmatized. Among girls, however, school work is typically viewed 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

The role of peers in shaping attitudes towards school and working habits is supported by a diverse group of studies. Coleman (1961), Eitzen (1975), Steinberg et al. (1996), and more recently Bishop et al. (2003) have argued that adolescents value the 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". This is also exemplified in a conversation between three boys in an English class that was documented by Morris (2008, 738; for other examples see Epstein, 1998):

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!"

²These assertions do not imply that girls are always engaged in the learning process. In contrast, many studies have documented the ways in which girls resist the teacher and school (e.g. Francis, 2000, 62f). Nevertheless, one of the most common findings in ethnographic studies is that boys more actively resist the learning process.

Although ethnographic studies have documented substantial within-gender diversity in the construction of gender identities, the evidence on typical gender differences is rather persuasive. Masculinity tends to be constructed among young boys at least partly in terms of resistance to school. This conception of masculinity may be partially responsible for male underachievement in school (Salisbury and Jackson, 1996; Pickering, 1997; Skelton, 1997; Francis, 2000). The conception of female identity and their peer culture, in 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 working habits and a stronger pro-school orientation.

While Willis and others have mainly focused on the consequence of lower and working-class background for anti-school attitudes among boys, we are interested in the school and class environment as a context that either encourages or limits the development of anti-school attitudes and behavior. High status parents generally manage to foster an orientation for their boys that is at least instrumentally focused on high performance in school. They also have the resources to intervene in their children's lives to counter signs of educational detachment or poor performance. As Coleman and others have argued, schools can play a similar role in enhancing the incentives of students to be engaged with academics by creating a learning oriented peer culture. In this line, many argue that the success of some charter schools such as KIPP and the Harlem Children Zone comes from their ability to foster a learning oriented environment (Ravitch, 2010, 144f).

We argue that boys gain more from a learning oriented environment, because it channels how masculinity in the school culture is constructed. Such an environment promotes academic competition as an aspect of masculinity and encourages the development of adaptive strategies that enable boys to maintain a showing of emotional coolness towards 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 Ghail, 1994) becomes a more important part of the construction of

masculine identity in certain environments.

As is true in the family, the 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 that 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 potentially promote a learning-oriented student culture. Accordingly, we would expect that teachers with the right collection of skills might have especially positive effects on the achievement of boys.

The school resource of central interest in this paper is the socioeconomic composition of the student body. The impact of peers on school climate and student achievement has played a crucial role in the literature on schools ever since Coleman 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” (Coleman, 1966, 325). The mechanism behind this association is cultural; students with high motivation and achievement from a high class background create a learning oriented peer culture and assist the teacher in the process of education (Sewell et al., 1969; Jencks and Mayer, 1990; Rumberger and Palardy, 2005, 125). 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 the conception of adolescent and pre-adolescent masculinity, suppresses boys’ negative attitudes towards school, and facilitates academic competition as an aspect of masculine identity. Girls’ peer groups, on the other hand, more readily and independently of the 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 socioeconomic 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 the SES composition of classroom peers as our focal treatment variable. The ELEMENT dataset is a longitudinal study that assessed the development of reading and math ability in the 4th, 5th, and 6th grade in Berlin schools (Lehmann and Lenkeit, 2008). It includes about 3,300 students who attended the 4th grade during the school year 2002-2003 in 71 randomly selected elementary schools in Berlin and all 1,700 students who attended the 5th grade in 2003-2004 in one of the 31 Berlin upper secondary schools that begin with 5th grade.⁴ In our final models, we combine these two ELEMENT samples, and control for the school type through school-level fixed effects. We also examined whether the relevant effects vary by school type using interaction terms (they do not). Appendix A provides a short introduction into the German educational system.

The ELEMENT dataset includes at least two classrooms for every school.⁵ This feature of the dataset 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

³Our expectations mainly relate to wealthy OECD countries because prior research has found that both the role of the school context (Chudgar and Luschei 2009) as well as gender relations differ substantially between high- and low-income countries.

⁴In contrast to most other states in Germany, students in Berlin usually attend elementary school until the 6th grade so that the 31 fifth grade upper secondary schools - the so called 'grundständige Gymnasien' - are different from the other 'normal' secondary schools.

⁵Elementary school students in Berlin who are assigned to the same classroom take virtually all their classes together, and so we use the terms "classroom" and "class" interchangeably in the text below.

⁶For the 5th grade upper secondary schools in ELEMENT the class assignment occurs in 5th grade because the students transfer after 4th grade from an elementary school.

estimation strategy provides a clear advantage over similar estimates based on data from U.S. schools, where both performance-based tracking in elementary schools and parents' influence on assignment to classes are more pronounced.

While our quasi-experimental research design provides high internal validity and allows us to make a strong case for causal inference, 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 together 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 the conditioning on observable variables as an identification strategy potentially suffer from endogeneity problems. They rely on the assumption that students are randomly assigned to schools conditional on the observable covariates in the model (Sørensen and Morgan, 2006, 155f). This common identification strategy is especially problematic for the 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.

⁷Both datasets were obtained from the Forschungsdatenzentrum at the Institute für Qualitätsentwicklung im Bildungswesen (IQB) HU-Berlin.

⁸As a substantive matter, the culture of fifth grade differs from the culture of ninth grade in the obvious sense that the students in fifth grade are pre-adolescent while the students in 9th grade 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 (Thorne, 1993; Maccoby, 1998). Thus, for both substantive and methodological reasons, we expect the comparison of results from fifth and ninth grades to be informative about our core hypothesis.

In order to avoid these potential endogeneity problems, we estimate school-level fixed effects models using both the ELEMENT and the PISA-I-Plus data. Both datasets contain an additional level of analysis, namely the classroom. We argue that students are almost randomly assigned to classrooms conditional on their school in both Berlin’s elementary schools and 5th grade upper secondary schools (for a similar strategy see Ammermueller and Pischke, 2009). Assuming the random assignment of students 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 (for detailed discussion of the variables see below). We specify these models as

$$y_{ijk} = \alpha_j + \gamma (female)_i + \theta (SES\ Comp)_k + \delta ((SES\ Comp)_k \times female_i) + \beta_1 y_i^{4th\ grade} + \mathbf{X}_i \beta_2 + \mathbf{U}_k \beta_3 + \epsilon_{ijk} \quad (1)$$

where i , j , and k are indices for individuals, schools, and classes respectively, α_j are the school fixed-effects, $y_i^{4th\ grade}$ is the prior achievement of the student measured in 4th grade, and \mathbf{X}_i and \mathbf{U}_k are sets of control variables on the individual and class level, respectively.⁹ The analysis with the PISA-I-Plus dataset omits the variable of prior achievement on the right hand side because of data limitations.¹⁰

These models examine whether the class-to-class variation in performance is systematically related to the class-to-class variation in socioeconomic composition controlling for all unobserved school characteristics (and therefore the non-random selection of students into schools). The coefficients of interest are θ , which captures the causal effect of the socioeconomic class composition, and δ , which captures the difference in this effect

⁹The three-level data structure might imply that the 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: 308ff).

¹⁰Although the PISA-I-Plus is a panel study and collected achievement data in both 9th and 10th grade, the panel component of these data is not yet available.

between boys and girls. We expect a positive effect of SES composition as previously documented and, more importantly for our theory, a negative estimate of the interaction term indicating that boys are more sensitive to peer SES. The pre-treatment control variables on the student and class level are of secondary interest, and are included to increase balance between the treatment and control group (for a description of the control variables, see Table 1).

The Assignment of Students to Classrooms *within* Schools

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

In order to develop a deeper understanding of the actual selection process, we conducted a three-part analysis of this process. First, we studied the official school regulations in Berlin. Second, we used a simulation-based approach to compare the 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 detailed picture of the actual selection process that results from this examination allowed us to evaluate our argument that the self-selection is practically random and to design targeted statistical sensitivity analyses that address potential sources of

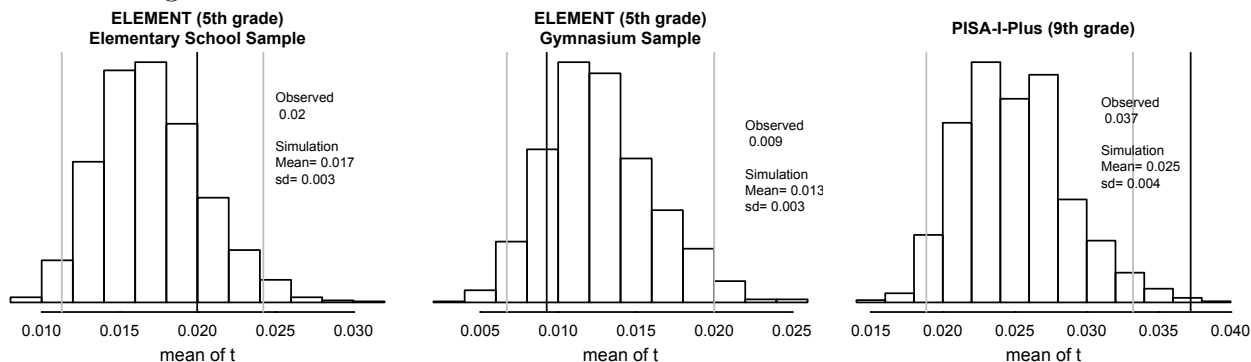
biases.

School Regulations and General Considerations The primary school regulations in Berlin (Grundschulverordnung Berlin, §8) prohibit the allocation of students based on gender, first language, or performance, and emphasize the 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 the classroom assignment of grade repeaters or newcomers. As a consequence, an allocation of students to classrooms based on family background is unlikely. The 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 can pose a problem for the estimation of the causal effect.

In secondary schools such as those in the PISA-I-Plus data, class-specific tracking based on subject choice such as 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 that in secondary schools the selection problem might be more pronounced. The situation at the 5th grade upper secondary schools (grundständige Gymnasien) in Berlin, however, is different from other secondary schools. The population of students who attend these schools is more homogeneous compared to other secondary schools, which makes a purposeful allocation to different classes relatively inconsequential. In addition, the assignment to 5th grade is not subject to selection over time through grade retention because students enter these schools for the first time at grade 5.

Based on these considerations, we expect that assignment to 5th grade classrooms is practically random both in elementary schools and Berlin's upper secondary schools (grundständige Gymnasien), whereas assignment to 9th grade classrooms in secondary

Figure 1: Comparison of Observed Classroom Composition with Simulations Involving Random Assignment



Note: The graphs show the average variation of class means within schools for the observed samples (vertical line) together with the sampling distribution of this statistic obtained from 1000 random simulations (histogram). The vertical grey lines indicate the 95% confidence intervals from the simulations.

schools is subject to more pronounced selection processes.

Simulation of Random Assignment We use a simulation-based approach in order to evaluate whether the within-school variation in the socioeconomic composition across classrooms created by the actual (unknown) allocation process is consistent with random assignment. Figure 1 compares the socioeconomic composition across classrooms obtained from simulations that randomly assign students to classrooms (histogram) with the observed composition (vertical line) in terms of the average variation of class means within schools (see Appendix B 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 the secondary schools in the PISA-I-Plus is relatively unlikely to occur under random assignment. Similar simulations for the proportion of students with migration background suggest that the assignment in regard to this characteristic is consistent with randomness for all three datasets. Finally, the observed statistic (i.e. variation across classrooms within schools) is smaller than the simulated distribution for the case of gender composition (see Online Appendix). 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 the assignment to classrooms within schools with respect to family background is practically random in the ELEMENT dataset. In contrast, some non-random selection process seems to play a role for 9th grade in secondary schools.

Interviews with School Principals Although the 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, who are the central actor in the allocation process in Berlin elementary schools (9 interviews) and *grundständige Gymnasien* (3 interviews). The schools were selected using a random sample that we then supplemented with specific schools to ensure diversity in regard to neighborhood and ethnic composition. The interviews lasted about 15-20 minutes and focused on the actual procedure the schools use to assign students to classes, the criteria that play a role in the assignment, the extent to which parents try to influence this process, and the ways in which the school deals with parental requests. The 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 Appendix 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 assignment process (for an exception see below). Second, schools try to have classes with similar

size. This plays an important role in the assignment of students who either repeat a grade or transfer from another school. Third, the assignment of teachers to classrooms is generally not connected to the socioeconomic composition or other characteristics of the class. Teacher assignment is based on scheduling issues and past experience with the teacher.¹¹

There are, however, also a number of potential biases: First, while all school principals emphasized that the desire to equalize classroom size is the main criteria, 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 first grade classroom. Other principals mentioned that 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 background equally across classes, two principals mentioned that they create a separate class for children who are German learners. While the simulations suggested 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 the results from the simulation insofar as the 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

¹¹In addition, all schools reported that class changes within a grade level are extremely rare, and resources are generally allocated equally across classes.

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 assignment process is not very systematic. One assistant school principal and teacher, for example, emphasized that even decades of experience in working at elementary schools could not remove the inherent unpredictability about the dynamics of classrooms, given the limited prior knowledge about entering students that the 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” (assistant school-principal at an elementary school in Berlin, translation by authors).

These and similar concluding remarks were elicited from the interviewees at the end of the interview by asking how they would weigh the importance of the different criteria and whether they thought that 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.

Conclusion about Selection Process Based on the evidence from the school regulations, the simulations, and the interviews with school principals, we conclude that the role of potential selection biases is limited. As such, the 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 analysis. These robustness checks are based on instrumental variable analyses and sample restrictions specifically designed to address each of the potential sources 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 the evidence from the interviews indicates that students might select into certain classrooms, it seems unlikely that 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. The interviews did not provide any indication of differential treatment of boys and girls even though the 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). Reading scores have been described as “one of the most important abilities students acquire through their early school years. It is the foundation for learning across all subjects” (Campbell et al., 2001, 1). Reading literacy has also figured importantly in research on the gender gap in education, because reading is the cognitive area where male achievement on test scores lags notably behind that of females (Buchmann et al., 2008). Some researchers have even argued that boys’ failure in general is due to their deficits in reading (Whitmire, 2010). The test scores are measured on a common scale using item response theory, and are standardized with a mean of zero and a standard deviation of one.

Our focal treatment variable is the socioeconomic (SES) composition of the student body, which is measured at the classroom level as the average social status on the ISEI

scale (Ganzeboom et al., 1991).¹² An argument can be made that prior achievement of peers 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 the 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 the learning orientation of the peer group because attributes such as high motivation and capability are more common among students from high SES families. Consequently, the SES composition of the student body is a school resource that fosters a learning orientation, and is highly relevant for our study.

Aside from SES composition, we use a comprehensive set of control variables including 4th grade test scores as a measure of prior performance. These variables are described in Table 1 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, 13ff). We performed each analysis separately for the five imputed datasets and then combined the different estimates to obtain the final results presented in this paper. We employed a similar imputation strategy based on the chained equations approach for the PISA-I-Plus dataset.

¹²We also explored alternative specifications of SES composition effects, such as allowing separate effects of the SES composition of male and female peers. These alternative specifications yield essentially the same results as those reported in the tables.

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

Variables	Description	Mean	SD	Mean (Male)	Mean (Female)	Std. Diff. in Means
<i>Dependent Variable</i>						
Achievement	5th grade reading test scores	109.92	13.65	108.79	111.05	-0.17**
<i>Independent Variables (Individual Level)</i>						
Prior Achievement	4th grade reading test scores	103.22	15.96	101.39	105.07	-0.23**
Female	0 - Male; 1 - Female	0.50	0.50			
Family Background	ISEI scale	51.45	16.63	51.09	51.80	-0.04
Migration Background	categorical variable:					
	1 - both parents born in Germany	0.64	0.48	0.64	0.65	-0.01
	2 - one parent not born in Germany	0.12	0.33	0.12	0.13	-0.01
	3 - both parents born in Germany but child	0.17	0.38	0.18	0.16	0.02*
	4 - child not born in Germany	0.06	0.24	0.06	0.06	-0.01
Class Repeater	0 - not repeated; 1 - repeated at least once	0.06	0.23	0.07	0.04	0.03**
<i>Interaction Terms</i>						
	Female x Family Background					
	Female x Migration Background					
	Female x SES Composition					
<i>Independent Variables (Class Level)</i>						
Size of Class	Number of students in the class	23.33	3.41	23.33	23.34	-0.00
SES Composition	Average ISEI at class level (aggregated)	51.45	9.40	51.08	51.10	-0.01

Source: ELEMENT; n=4372; * $p < 0.05$, ** $p < 0.01$

Note: 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 (the migration background dummies and the indicator for class repeater). The asterisks indicate statistical significance at the 0.05 and 0.01 level based on t- and z-tests. Note that all the continuous variables are standardized for the final analysis.

Results

Variation of the Gender Gap across Schools

In an average school, the female advantage in reading scores is about 0.12 standard deviations in 5th grade and 0.21 standard deviations in 9th grade. It ranges from -0.04 to 0.28 standard deviations in 5th grade and from 0.07 to 0.35 standard deviations in 9th grade for 95% of the schools. Expressed in terms of years of education, girls are 0.36 school years ahead in 5th grade reading test scores in an average school, but the gap ranges across schools from a male advantage of 0.12 years to a female advantage of 0.83 years.¹³ Figure 2 plots this variation in the gender gap on the school level against the average performance at a school. The striking pattern in the figure indicates that schools with higher average performance are also schools where the gender gap is small. This pattern is consistent with our theoretical prediction; it suggests that boys do not fall as far behind in schools that are performance oriented. The following section scrutinizes this initial finding using the quasi-experimental research design described above.

SES Composition and the Gender Gap in Education

The estimates from the school-level fixed effect regression of reading test scores in 5th grade on classroom level SES composition, 4th grade scores, and other control variables on the right-hand side are presented in Table 2. The table shows the main effect of gender and of SES composition on the classroom level together with the interaction between SES composition and gender (all coefficients are in standard deviation units). The other coefficients are omitted from the table (for the full regression results, see Online Appendix). The table also shows the FE-estimates from the PISA-I-Plus data for 9th grade reading test scores without a measure of prior performance and the estimates

¹³One additional school year corresponds to the estimated test score difference between 5th and 6th grade in the ELEMENT dataset.

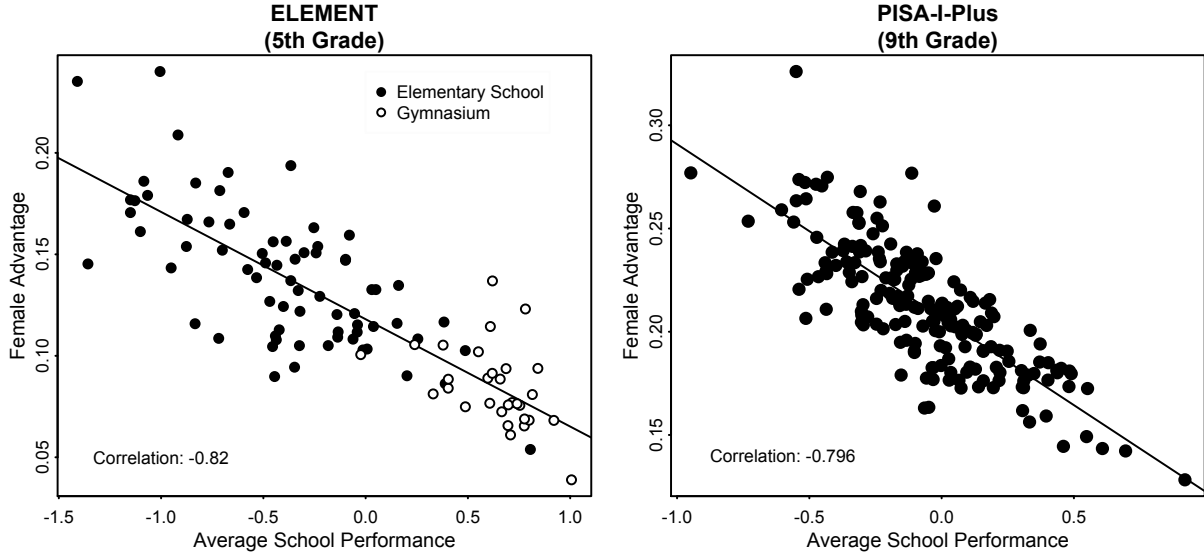


Figure 2: Gender Gap and Average Performance across Schools in Standard Deviation
Note: The estimates shown in the figure are based on a multilevel model with two levels (student and schools) and with a random intercept and a random slope for female on the school level so that both the average performance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

from a multilevel (MLM) model on the school level with a broad set of control variable. The MLM estimates are included as a comparison, because they reflect one of the most common estimation strategy (conditioning on observable covariates) used in sociology to identify compositional peer effects (e.g. Rumberger and Palardy, 2005).

The results in Table 2 show that SES composition has a positive and highly significant effect on reading test scores in all models and therefore both for gain scores (top row) and raw scores. This result conforms with previous findings reported in the literature on the effects of SES composition (Rumberger and Palardy, 2005; Jencks and Mayer, 1990). In all models, the point estimate for the interaction between SES composition and female is negative and significant. Most importantly, the estimates from the fixed effect model using the ELEMNT 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

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

Model	Prior Perf.	Female		SES Comp.		SES Comp. x Female	
		coef.	(se)	coef.	(se)	coef.	(se)
1. FE - Estimate (ELEMENT)	yes	0.007	(0.02)	0.091*	(0.04)	-0.060**	(0.02)
2. FE - Estimate (ELEMENT)	no	0.120***	(0.03)	0.178***	(0.06)	-0.057*	(0.02)
3. FE - Estimate (PISA-I-Plus 2003)	no	0.196***	(0.03)	0.237***	(0.03)	-0.052*	(0.02)
4. MLM - Estimate (PISA-I-Plus 2003)	no	0.143	(0.11)	0.303***	(0.05)	-0.099*	(0.04)

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; standard errors adjusted for clustering on level

Note: Control variables are listed and described in table 1. The full set of coefficient estimates for Models 1 and 2 are in appendix table A1. The number of students for the models based on ELEMENT is 4372, the number of schools is 101, and the average number of students per school is 43.3. n for PISA-I-Plus is 8559.

finding (results not shown here). The results from the two FE-models based on the ELEMENT and the PISA-I-Plus data without 4th grade performance show the same results (the ELEMENT results are included for direct comparison). In particular, the main effect of SES composition in the model based on the PISA-I-Plus data seems to be upwardly biased (0.237 compared to 0.178), and both estimates are somewhat larger than the 0.15 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. The 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% 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 larger size of the school-based estimate might also reflect spillover effects between the SES composition of one classroom and the SES composition of another classroom 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 are 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 classrooms within 5th grade in Berlin.

Targeted Sensitivity Analysis

In this section, we investigate whether our results are sensitive to the three potential selection biases documented in the interviews with school principals. Our detailed knowledge about the assignment process allows us to design a set of sensitivity analysis based on instrumental variables (IV) and certain sample restrictions that are 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 the results from the different sensitivity analysis and also repeats the estimates from the school FE model based on the 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 school principals reported that the size of the different classrooms plays an important role, some principals also mentioned that potential implications for the classroom culture are also taken into account. In order to address this potential selection problem, we treat the SES composition on the class level as endogenous and instrument it with the average SES of

Table 3: Sensitivity Analysis

Model	Female		SES Comp.		SES Comp. x Female	
	coef.	(se)	coef.	(se)	coef	(se)
FE-Estimate (full sample)	0.007	(0.02)	0.091*	(0.04)	-0.060**	(0.02)
FE/IV - Estimate						
(1) Instrument: SES comp. of students who never repeated a class	0.008	(0.02)	0.089*	(0.04)	-0.065**	(0.02)
FE/IV - Estimate						
(2) Instrument: SES comp. of students who didn't go to kinderg., skipped a class or transferred to school	0.009	(0.02)	0.113*	(0.06)	-0.068*	(0.03)
FE - Estimate (restricted sample)						
(3) Sample Restriction: Only schools that do not allocate based on ethnicity (24 schools excluded)	0.008	(0.03)	0.117*	(0.05)	-0.052*	(0.02)

n=4372; * p < 0.05, ** p < 0.01, *** p < 0.001; standard errors adjusted for clustering on level

Note: The first stage results show that the two instruments are highly correlated with SES composition (i.e. the treatment). The F-statistics are over 700 (highly significant), which is far above the commonly used threshold of 10. The control variables are described in table 1.

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 only based on those students who never repeated a grade. The instrument should also only be connected with the outcome through the actual class composition (i.e., it satisfies the exclusion restriction). The results are presented in Table 3 Model 1 and show that the interaction between SES composition and female remains negative and significant. This indicates that the selection of students who repeat a class into specific classes does not significantly bias the estimated effects.

The second potential selection process is the assignment of those 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 the 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 the students who skipped a class or transferred from a different school were most likely assigned to classrooms based on the number of students in the different classrooms. For these reasons, the instrument is unaffected by the kindergarten criteria and (for the most part) by friendship self-selection. The results, which are presented in Model 2 of Table 3, again support our previous finding and indicate that the estimated causal effect is not sensitive to the selection of connected students (either through the same kindergarten or through friendship) into the same class.

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

Overall, the 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 above suggests that the school context plays an important role for the size of the gender gap. An academically oriented environment in schools with high SES peers shapes the ways in which masculinity is constructed and

¹⁴We use a simple z-test to identify the schools in which the difference in the proportion of students with migration background between classes is higher than what we would expect under randomness. Using a conservative criteria, we exclude those schools with a p-value smaller than 0.1 (24 schools).

thereby suppresses boys' negative attitude towards school, facilitates their commitment, and enhances the incentives of students to be engaged with academics. It might well be the case, however, that other mechanisms 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 (Thrupp et al., 2002; Rumberger and Palardy, 2005; Jencks and Mayer, 1990). This alternative account argues that students use their classmates as a reference group to evaluate their own performance and thereby develop academic self-perceptions, which in turn may affect their performance (Dai and Rinn, 2008; Crosnoe, 2009). To adjudicate between our proposed explanation and this alternative account, we estimate models based on the 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 the class environment has a more pronounced effect on attitudes towards 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 the academic self-perceptions of boys and girls are affected differently by the socioeconomic composition of the class. In other words, this alternative account suggests that boys and girls react differently to their reference group.

Table 4 shows the results from school-level fixed effect models of the indicated variables on classroom socioeconomic composition, controlling for the variables described in Table 1. Panel A, which reports regression results using attitudes towards school, learning orientation, and working habits as dependent variables, provides further evidence

¹⁵The measures are constructed from a range of indicators using exploratory factor analysis (see Online Appendix).

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

		Female		SES Composition		SES Composition x Female	
		coef.	(se)	coef.	(se)	coef.	(se)
Panel A	Attitude Towards School	0.301***	(0.04)	0.054	(0.06)	-0.079*	(0.03)
	Learning Orientation	0.131***	(0.04)	0.043	(0.06)	-0.035	(0.03)
	Working Habits	0.166***	(0.04)	0.147*	(0.07)	-0.086*	(0.04)
Panel B	Self-Evaluation Reading	0.140***	(0.04)	-0.098	(0.06)	-0.028	(0.03)
	Self-Evaluation German	0.207***	(0.04)	0.012	(0.08)	-0.056	(0.03)
	Self-Evaluation general	-0.294***	(0.04)	-0.020	(0.07)	-0.025	(0.03)

n=4372; * p < 0.05, ** p < 0.01, *** p < 0.001; standard errors adjusted for clustering on level

Note: Control variables are described in Table 1.

for our core hypothesis. The 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 implies that boys' attitudes towards school, their learning orientation, and their working habits are more sensitive to the school environment than are the attitudes and working habits of girls. Panel B, in 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-

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

	Model 1		Model 2		Model 3	
	Coef.	(se)	Coef.	(se)	Coef.	(se)
Female	0.007	(0.02)	-0.009	(0.02)	-0.009	(0.02)
SES Composition	0.091*	(0.04)	0.037	(0.04)	0.033	(0.04)
SES Composition x Female	-0.060**	(0.02)	-0.060**	(0.02)	-0.040*	(0.02)
Attitude Towards School			0.041*	(0.01)	0.058*	(0.02)
Learning Orientation			0.006	(0.01)	0.004	(0.02)
Working Habits			0.067***	(0.01)	0.093***	(0.01)
Attitude Towards School x Female					-0.047*	(0.02)
Learning Orientation x Female					-0.001	(0.02)
Working Habits x Female					-0.069***	(0.02)
Control Variables	yes		yes		yes	
Constant	-1.000***	(0.35)	-0.470	(0.27)	-0.593*	(0.253)

n=4372; * p < 0.05, ** p < 0.01, *** p < 0.001; standard errors adjusted for clustering on level

Note: Control variables are described in Table 1.

cut. It can nonetheless be informative about potential mechanisms. The 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%) may be explained by its gender-specific effect on school-related attitudes and behavior, and therefore provide 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 because 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 fourth grade (the year prior to our measured outcomes in the regressions) and SES composition in fifth grade. The 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 (Coleman, 1966; Coleman, 1970; Bryk et al., 1993). The inclusion of this interaction also weakens the direct benefit of being male in a high SES class by about 27% (from -.060 to -

.044). However, the interaction between SES composition and gender remains both statistically significant (p-value 0.021) and substantively important. These results suggest that boys indeed do benefit indirectly from a stronger academic climate because they are disproportionately low-performing students. Nonetheless, the bulk of the effect stems from a greater sensitivity of boys than girls to the academic orientation of the classroom culture.

Discussion

Throughout the industrialized world, girls have made dramatic gains in educational attainment, while the under-performance of boys and their tendency to disrupt the learning process has sparked intense academic as well as public debates about the causes of what many now call the “problem with boys.” Some have blamed schools for fostering a demasculinized learning environment. Yet, the role of the school context and the connection between school resources and the gender gap has been under-developed in the literature to date. In this paper, we have extended research on the effect of schools on class and race inequality dating back to the 1966 Coleman report 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 the underachievement of boys. In particular, we argue that the school and class environment shapes the ways in which masculinity in the peer culture is constructed and thereby influences boys’ orientation towards school. Resources that create a learning oriented environment raise the valuation of academics in the adolescent male culture and facilitate commitment. Girls’ peer groups, in contrast, do not vary as strongly with the social environment in the extent to which they encourage academic engagement, and 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. The results from our analysis of the German ELEMENT and PISA-I-Plus 2003 data provide clear support for this hypothesis. We first showed that there is 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 the 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 the assignment of students to classes within Berlin elementary and 5th grade higher secondary schools. To evaluate this argument, we examined Berlin’s school regulations, compared the observed classroom composition with simulations involving random assignment, and conducted qualitative interviews with school principals in Berlin. The 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. The results from these analyses showed little effect of these 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. The results from this analysis provide further support for our own explanation. They suggest that boys benefit both indirectly (because low-performing students benefit in general) and directly (because the effect is bigger for boys than 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 the well-publicized under-performance of boys. The outlined cultural mechanism explains why boys are more sensitive to the presence of human and cultural capital resources in schools, which turns out to play an important role for the under-performance of boys and the gender gap in educational achievement.

This argument suggests that boys' resistance to school is not purely a function either of their class background – as suggested by many studies – or the fact of their masculinity – as suggested by other studies – but instead depends on the local cultural environment of the school and classroom. As such, the findings broaden our understanding of the notorious under-performance of boys. They point at an important mechanism connected to how the school and class environment shape the learning orientation of boys and girls, and in the process reveal a pattern similar to what has previously 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 the achievement level of boys. 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 that is 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 the academic environment of the school, and raise academic performance. They have the potential to modify student behavior and produce a stronger academic student culture even in the absence of socioeconomic enrichment of the school's student body. At present, however, too little is known about what makes a quality teacher, or the extent to which higher academic performance induced by better teachers has a strong effect on the academic climate. These are important questions for further research.

Finally, the paper makes a methodological contribution to the literature on the estimation of causal effects. Our work illustrates how a detailed study of the relevant selection process – in our case, the examination of official regulations, statistical simu-

lations, and qualitative interviews – can facilitate the estimation of causal effects. This detailed understanding of the actual selection process not only allows the researcher 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 such as in our case for compositional peer effects in school. 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 also limited in some regards. Most importantly, our theoretical argument applies to all kinds of school resources that create a learning oriented environment. Our empirical analysis, however, only focuses on one (though important) dimension, namely peer socioeconomic composition. Given this limitation, future studies should establish the extent to which the conclusions from this study apply to other kinds of school-based resources. Additionally, due to the lack of adequate data, our study neglects the role of teachers in shaping the learning orientation of boys and girls. While our interviews indicate that teachers are not assigned to classrooms based on the classroom composition, it might still be the case that teachers react to the classroom dynamics in a certain way and thereby play an important role for the processes studied in this paper. Finally, our study focuses on only one major dimension of cognitive achievement, namely reading. Boys on average do as well or better than girls in mathematics, with the male advantage being larger on the right tail of the distribution. Whether boys nonetheless 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 further research.

Appendix A Education and the Educational Gender Gap in Germany

Although the main focus of the paper is the theoretical argument, the background information provided in this section helps to contextualize the findings from the German case. In Germany, children usually attend elementary school from 6 to 10 or 12 years of age depending on the state (Bundesland) regulations. After finishing elementary school, the students transfer to one of the secondary school types, which are distinct from the American middle and high school because of the 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 as the highest secondary school type, the Realschule for intermediate students, and the Hauptschule as 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 US 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 the German states. Similarly to other industrialized countries, the gender gap in Germany has closed over the last decades. Legewie and DiPrete (2009), however, also emphasize that the female advantage in higher education is less pronounced compared to the US due in large part to their failure to converge with men in rates of obtaining degrees from Fachhochschulen (universities of applied sciences).

Appendix B Simulation of Random Assignment

This appendix contains a detailed description of our simulation-based approach. The simulation allows us to evaluate whether the within-school variation in the composition of classes is consistent with a random allocation process. To compare the observed composition with the 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 the 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

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

where j and k are the indices for schools, and classrooms respectively, \bar{x}_j is the average for school j , \bar{x}_{jk} the average for classroom k in school j , and n_j 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.

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(3):315-348.
- Angrist, Joshua D., and Jörn-Steffen Pischke. 2008. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton: Princeton University Press.
- Bishop, JH, M 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):141-199.
- Bryk, Anthony, Valerie Lee, and Peter Holland. 1993. *Catholic Schools and the Common Good*. Cambridge: 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(4):515-541.
- Buchmann, Claudia, Thomas A. DiPrete, and Anne McDaniel. 2008. "Gender Inequalities in Education." *Annual Review of Sociology* 34:319-337.
- 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(3):626 -658.
- Cohen, Michele. 1998. "A habit of healthy idleness: boy's underachievement in historical perspective." in *Failing Boys? Issues in Gender and Achievement*, edited by Debbie Epstein, Jannette Elwood, Valerie Hey, and Janet Maw. Buckingham: Open University Press.
- Coleman, J. 1966. *Equality of Educational Opportunity*. U.S. Dept. of Health, Education, and Welfare, Office of Education.
- Coleman, JS. 1970. "Reply to Cain and Watts." *American Sociological Review* 35(2):242-249.
- Coleman, JS. 1961. *The Adolescent Society: The Social Life of the Teenager and its Impact on Education*. New York: Free Press.
- Correll, Shelley J. 2001. "Gender and the Career Choice Process: The Role of Biased Self-Assessments." *American Journal of Sociology* 106(6):1691-1730.
- Crosnoe, Robert. 2009. "Low-Income Students and the Socioeconomic Composition of Public High Schools." *American Sociological Review* 74(5):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(3):283-317.
- Davies, Bronwyn. 2003. *Frogs and Snails and Feminist Tales: Preschool Children and Gender*. Hampton Pr.
- 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(4): 415-440.
- 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. "Atheletics in the Status System of Male Adolescents: A Replication of Coleman's 'The Adolescent Society'." *Adolescence* 10(38):267-276.
- 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 Debbie Epstein, Jannette Elwood, Valerie Hey, and Janet Maw. Buckingham: 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(540):1243-1269.
- Greenwald, Rob, Larry V. Hedges, and Richard D. Laine. Fall. "The Effect of School Resources on Student Achievement." *Review of Educational Research* 66(3):361-396.
- Hallinan, M. T. 1988. "Equality of Educational Opportunity." *Annual Review of Sociology* 14(1):249-268.
- Hanushek, Eric A. 1989. "The Impact of Differential Expenditures on School Performance." *Educational Researcher* 18(4):45-62.
- Howe, Christine. 1997. *Gender and Classroom Interaction: A Research Review*. The SCRE Centre.
- Hoxby, Caroline. 2000. *Peer Effects in the Classroom: Learning from Gender and Race Variation*. National Bureau of Economic Research Retrieved (www.nber.org/papers/7867).
- Imberman, Scott, Adriana Kugler, and Bruce Sacerdote. 2009. "Katrina's Children: A Natural Experiment in Peer Effects from Hurricane Evacuees."
- Jencks, Christopher, and Susan E. Mayer. 1990. "The social consequences of growing

- up in a poor neighborhood.”P. 111–186 in *Inner-city poverty in the United States*, edited by Laurence E. Lynn and Michael 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(2):135-159.
- Jovanovic, Jasna, and Sally Steinbach King. Autumn. “Boys and Girls in the Performance-Based Science Classroom: Who’s Doing the Performing?” *American Educational Research Journal* 35(3):477-496.
- Kahlenberg, Richard D. 2001. *All Together Now: Creating Middle-Class Schools through Public School Choice*. Washington, D.C. Brookings Institution Press.
- Kao, G., M. Tienda, and B 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(2):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 Ghail, Mairtin. 1994. *The Making of Men: Masculinities, Sexualities and Schooling*. Buckingham: Open University Press.
- Maccoby, Eleanor E. 1998. *The Two Sexes: Growing up Apart, Coming Together*. Cambridge, Mass: 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(3):357-372.
- MacLeod, Jay. 2008. *Ain’t No Makin’ It: Aspirations and Attainment in a Low-Income Neighborhood*. Oxford: Westview Press.
- Morris, Edward W. 2008. “‘Rednecks,’ ‘Rutters,’ and ‘Rithmetic.” *Gender & Society* 22(6):728 -751.
- Morse, Susan. 1998. *Separated by Sex: A Critical Look at Single-Sex Education for Girls*. American Association of University Women Educational Foundation.
- Murnane, Richard J., Allan Odden, and L. Dean Webb. 1983. “Quantitative Studies of Effective Schools: What Have We Learned?”Pp. 193-209 in *School Finance and School Improvement*. Cambridge: Ballinger.
- Pickering, Jon. 1997. *Raising Boys’ Achievement*. Network Continuum Education.
- PISA-Konsortium Deutschland. 2006. *PISA 2003: Untersuchungen zur Kompetenzentwicklung im Verlauf eines Schuljahres*. Münster: Waxmann Verlag.
- Power, S., G. Whitty, T. Edwards, and V. Wigfall. 1998. “Schoolboys and schoolwork:

- gender identification and academic achievement.” *International Journal of Inclusive Education* 2(2):135–153.
- 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(1):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.” *The American Economic Review* 94(2):247-252.
- 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(9):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, Salisbury, 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: Yale University Press.
- Sampson, Robert J. 2008. “Moving to Inequality: Neighborhood Effects and Experiments Meet Social Structure.” *American Journal of Sociology* 114(1):189-231.
- Schneider, BL, and Venessa A. Keesler. 2007. “School Reform 2007: Transforming Education into a Scientific Enterprise.” *Annual Review of Sociology* 33(1):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(1):82-92.
- Skelton, Christine. 1997. “Primary Boys and Hegemonic Masculinities.” *British Journal of Sociology of Education* 18(3):349-369.
- Sørensen, Aage B., and Stephen L. Morgan. 2006. “School Effects: Theoretical and Methodological Issues.” Pp. 137-160 in *Handbook of the Sociology of Education*, edited by Maureen 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, LD, 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: Rutgers University Press.
- Thrupp, Martin, Hugh Lauder, and Tony Robinson. 2002. "School composition and peer effects." *International Journal of Educational Research* 37(5):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(3):325-341.