

School Context, Peers and the Educational  
Achievement of Girls and Boys

Joscha Legewie

Submitted in partial fulfillment of the  
requirements for the degree of  
Doctor of Philosophy  
in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2013

© 2013  
Joscha Legewie  
All rights reserved

ABSTRACT

**School Context, Peers and the Educational  
Achievement of Girls and Boys**

Joscha Legewie

Today, boys dominate among high school dropouts, special education students, and literally any failed or special needs category throughout adolescence pinpointing boys as the troublemaker in modern educational systems. The notorious under-performance of boys in school and their tendency to disrupt the learning process in the class room has sparked intense academic as well as public debates about the causes of what many now call the “problem with boys.” Yet, historically, the lower performance of boys in school is not a new phenomenon. In fact, researchers overwhelmingly agree that girls and boys have similar levels of mental ability and generally observe relatively small changes in academic performance over the last decades. What is new is the striking reversal of the gender gap in educational attainment, which has changed from a male to a female advantage. At the same time, girls continue to lag behind in terms of science, engineering, and technology degrees. These persisting gender differences are not only relevant for gender equality but also for the supply of qualified labor—a linchpin for the future of the U.S. economy in an increasingly competitive global environment.

A widespread argument among parents, teachers, and policy makers alike has been that boys resistance to school is part of their masculinity: Boys are simply more active and disobedient to authority. Others blame schools for what they see as a de-masculinized learning environment and a tendency to negatively evaluate boys for fitting into this environment less well than girls. Yet, the role of the school context and the connection between school resources and

the gender gap remains controversial. Research on the effect of schools dates back to the 1966 Coleman report and developed out of the concern for equality of educational opportunity by social class and race. This original focus and much subsequent work condemned the unequal access to high quality schools for black and white kids and called for the desegregation of schools. Now that a growing gender gap in educational attainment has emerged, it is natural 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. The goal of this dissertation is to address this question and examine the role of the school context for gender differences in education and thereby challenge the view of boys as universally disengaged from school and opposed to authority. For this purpose, the three papers in this dissertation each examine different aspects of this broader question. Together, these three articles make important contributions to our understanding of gender differences in educational outcomes, and suggest concrete policy implications about the educational shortcomings of boys, and the persisting gender gap in STEM degrees. They show that peer effects are larger for boys than girls and that this gender difference can be explained by differences in the social support for academic work in the male and female peer culture. These findings shift the focus from masculinity as inherently based on resistance to school towards the importance of the local school environment for the construction of gender identities as well as school-related attitudes, behavior, and the performance of boys and girls. My findings also point to the high school years as the life course period that should be targeted to increase the number of women with STEM BAs, and provide evidence that high school interventions might be effective to achieve that goal.

# Contents

<b>Introduction</b>	<b>1</b>
<b>1 School Context and the Gender Gap in Educational Achievement</b>	<b>9</b>
1.1 Introduction . . . . .	10
1.2 Educational Outcomes and Schools . . . . .	12
1.2.1 The School Context and the Gender Gap in Education . . .	14
1.2.2 The Under-Achievement of Boys, Gender Identity and School Climate . . . . .	15
1.3 Data and Methods . . . . .	20
1.3.1 School-Level Fixed Effects as a Quasi-Experimental Ident- tification Strategy . . . . .	22
1.3.2 The Assignment of Students to Classrooms <i>within</i> Schools	23
1.3.3 Variables and Treatment of Missing Data . . . . .	30
1.4 Results . . . . .	33
1.4.1 Variation of the Gender Gap across Schools . . . . .	33
1.4.2 SES Composition and the Gender Gap in Education . . . .	33
1.4.3 Targeted Sensitivity Analysis . . . . .	36
1.4.4 Explaining the Observed Difference in the Causal Effect between Boys and Girls . . . . .	39
1.5 Discussion . . . . .	43

Appendices . . . . .	48
<b>2 Disruptive Change: Peer Effects and the Social Adjustment Process of Mobile Students</b>	<b>69</b>
2.1 Introduction . . . . .	70
2.2 Peer Effects and Educational Outcomes: Theory and Research . . . . .	72
2.2.1 Mechanisms of Peer Effects . . . . .	74
2.3 Student Mobility, Social Integration and the Influence of Peers . . . . .	76
2.3.1 The Effect of Student Mobility . . . . .	76
2.3.2 Peer Influence and Student Mobility . . . . .	79
2.4 Data and Methods . . . . .	81
2.4.1 Difference in Difference Estimation Strategy . . . . .	82
2.4.2 Variables and Missing Data . . . . .	89
2.5 Results . . . . .	90
2.5.1 Patterns of Student Mobility . . . . .	90
2.5.2 The Effect of Changing School . . . . .	92
2.5.3 Student Mobility, Exposure Dynamics and the Effect of Peers . . . . .	94
2.6 Conclusion . . . . .	98
Appendices . . . . .	103
<b>3 High School Environments, STEM Orientations, and the Gender Gap in Science and Engineering Degrees</b>	<b>115</b>
3.1 Introduction . . . . .	116
3.2 Explanations for the Persisting Gender Gap in STEM Degrees . . .	119
3.3 Timing and Local Variations of the Gender Gap in Science and Engineering Orientation . . . . .	122

3.3.1	Life Course Timing in the Emergence of the Gender Gap .	123
3.3.2	The Local and Global Environment, and the Formation of Educational and Occupational Plans . . . . .	126
3.3.3	Summary . . . . .	130
3.4	Pathways to a STEM Bachelor Degree . . . . .	131
3.5	The Role of High School for the Gender Gap in STEM Orientation . . . . .	138
3.5.1	The Effect of High School Curriculum on the Gender Gap in STEM Orientation . . . . .	144
3.5.2	Is the High School Effect Lasting and How Big is the Effect?	151
3.6	Conclusion . . . . .	152
	Appendices . . . . .	158
	<b>Conclusion</b>	<b>175</b>
	<b>Bibliography</b>	<b>181</b>





# List of Tables

1.1	Variables in Main Analysis at Individual and Class Level . . . . .	32
1.2	Effect of SES Composition for Boys and Girls in Standard Deviations . . . . .	35
1.3	Sensitivity Analysis . . . . .	38
1.4	Effects of Gender and SES Composition on School-Related Attitudes and Behavior . . . . .	41
1.5	Fixed Effects Models with School-Related Attitudes and Behavior	42
A1.6	Variables in Main Analysis at Individual and Class Level . . . . .	64
A1.7	Gender Differences in the Effect of Peer SES . . . . .	66
2.1	Sample Characteristics . . . . .	90
2.2	Effect of Peer Ability and Student Mobility . . . . .	96
A2.3	Evaluation of Common Trend Assumption . . . . .	107
3.1	Decomposition of Gender Gap in STEM BAs, NELS 1988–2000 . .	137
3.2	Gender Effect in STEM Orientation across Schools . . . . .	140
3.3	Logistic Regression Estimates for the HS Entry Pathway of STEM Orientation in 12th Grade on High School Math and Science Curriculum . . . . .	147
A3.4	Description of Variables . . . . .	159

A3.5 Decomposition of Gender Gap in STEM BAs, Academically Tal- ented Students and STEM Sub-fields . . . . .	162
A3.6 Logistic Regression Estimates for the Persistence Pathway of STEM Orientation in 12th Grade on High School Math and Science Cur- riculum . . . . .	172
A3.7 Post-High School Transition Rates for Full Sample, and School with Small/Big Gender Gap . . . . .	173
A3.8 Gender Gap in STEM BAs for Full Sample, Schools with Small Gender Gap, and Schools with Big Gender Gap . . . . .	173

# List of Figures

1	Gender Gap in Bachelor Degrees Awarded by Field of Study, 1969-2007 . . . . .	2
2	Gender Gap in 10th grade Reading Test Scores across Schools, 2002	3
1.1	Comparison of Observed Classroom Composition with Simulations Involving Random Assignment . . . . .	26
1.2	Gender Gap and Average Performance across Schools in Standard Deviation . . . . .	34
A1.3	Variation in SES composition of Adjacent Cohorts within a School within a Grade with School-Grade Average . . . . .	53
A1.4	Variance in the Share of Female Students across Cohorts in CPS Schools . . . . .	60
A1.5	Variance in the Share of Students with Free Lunch across Cohorts	61
2.1	Illustration of Alternative Estimation Strategies . . . . .	83
2.2	Change in School Quality in terms of Peer Ability for mobile Students across Chicago Public schools . . . . .	92
2.3	Effect of Student Mobility on Test-Scores by Type of Transfer and Gender . . . . .	93
2.4	Effect of Peer Ability by Years since Student Changed to School .	98
A2.5	Balance between Treatment and Control Group . . . . .	106

3.1	Gender Gap in Bachelor Degrees Awarded by Field of Study, 1969-2007 . . . . .	118
3.2	Persistence in the Pathway to a STEM BA Degree . . . . .	133
3.3	Leakage from the Pathway to a STEM BA Degree . . . . .	134
3.4	Late Entry as a Pathway to a STEM BA Degree . . . . .	135
3.5	Variation of Gender Gap in 12th Grade STEM Orientation across Schools, NELS 1988-92 . . . . .	142
3.6	Effect of High School Curriculum on STEM Orientation . . . . .	150
A3.7	Balance between Top and Bottom Quartile of Continuous Treat- ment Variable . . . . .	165
A3.8	Sensitivity of Estimate to Additional Confounding Variable . . . . .	169

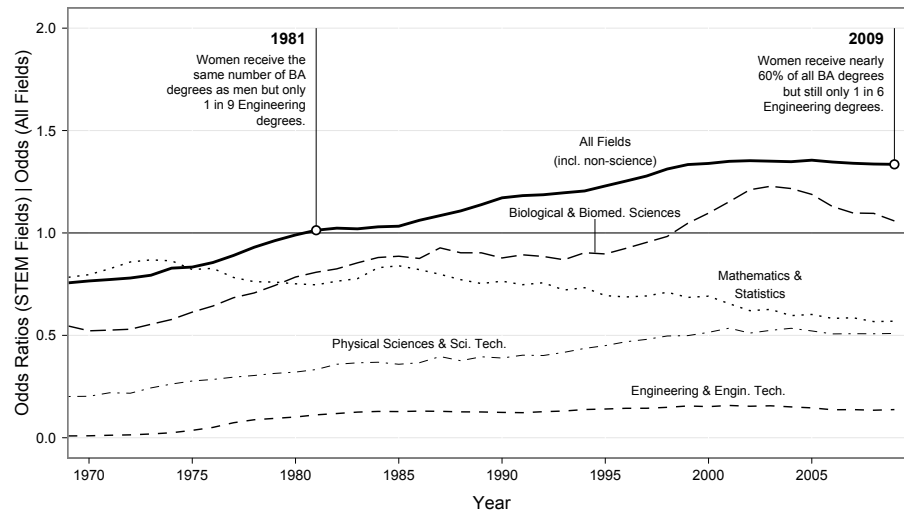
# Introduction

Today, boys dominate among high school dropouts, special education students, and literally any failed or special needs category throughout adolescence pinpointing boys as the troublemaker in modern educational systems.<sup>1</sup> The notorious under-performance of boys in school and their tendency to disrupt the learning process in the class room has sparked intense academic as well as public debates about the causes of what many now call the “problem with boys.” Yet, historically, the lower performance of boys in school is not a new phenomenon. In fact, researchers overwhelmingly agree that girls and boys have similar levels of mental ability and generally observe relatively small changes in academic performance over the last decades. What is new is the striking reversal of the gender gap in educational attainment, which has changed from a male to a female advantage (Buchmann and DiPrete 2006). At the same time, girls continue to lag behind in terms of science, engineering, and technology degrees. Figure 1 illustrates these trends. It shows, on the one hand, how women have made impressive gains in college attainment compared to men and now clearly outnumber men among college graduates in recent decades. On the other hand, women continue to lag behind in terms of bachelor degrees awarded in the physical sciences, mathematics, and engineering (illustrated in

---

<sup>1</sup>Out of 30 member nations of the Organization for Economic Cooperation and Development (OECD), men retain significant advantages only in Switzerland, Turkey, Japan and Korea (OECD 2007).

Figure 1: Gender Gap in Bachelor Degrees Awarded by Field of Study, 1969-2007



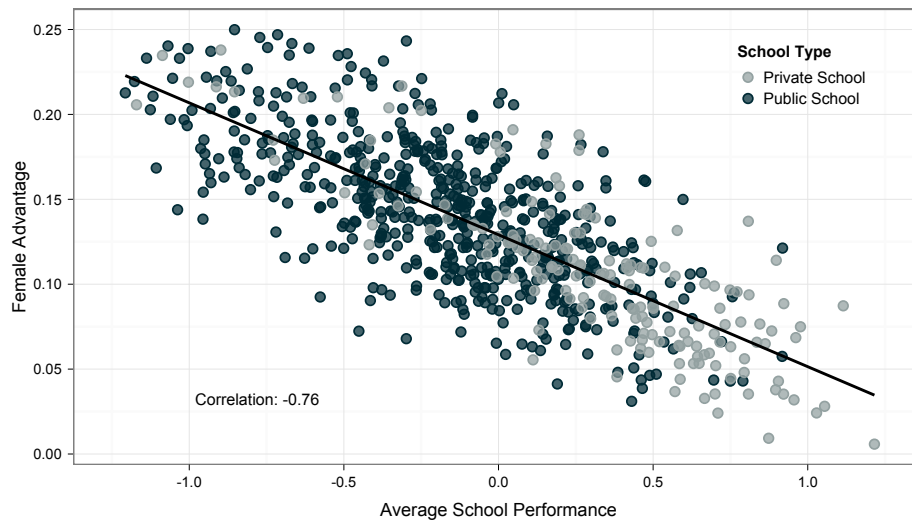
Source: *Digest of Educational Statistics 2011*, Table 268, 299, 303, 305, 312 and 313

the graph for different STEM sub-fields).<sup>2</sup> These persisting gender differences are not only relevant for gender equality but also for the supply of qualified labor—a linchpin for the future of the U.S. economy in an increasingly competitive global environment.

A widespread argument among parents, teachers, and policy makers alike has been that boys' resistance to school is part of their masculinity: Boys are simply more active and disobedient to authority. This view remains appealing in public debates and reflects a belief in deeply entrenched, possibly innate gender differences. Others blame schools for what they see as a de-masculinized learning environment and a tendency to negatively evaluate boys for fitting into this environment less well than girls. Yet, the role of the school context and the connection between school resources and the gender gap remains controversial. This is especially surprising considering the intriguing pattern revealed in

<sup>2</sup>Exceptions to this trend are the biological, biomedical and life sciences, in which women today outnumber men.

Figure 2: Gender Gap in 10th grade Reading Test Scores across Schools, 2002



Source: *Educational Longitudinal Study, 2002*

Note: The graph shows the 10th grade gender gap in standardized reading test scores plotted against the average performance across 751 high schools in the United States. The dots are based on empirical Bayes predictions for the the random intercept and the random slope from a multilevel model.

Figure 2. The graph shows substantial variations in the size of the gender gap in reading test scores across 751 high schools in the U.S. and indicates that boys do worse in schools with lower average performance. Indeed, in some schools boys excel in reading nearly at the same rate as their female peers do while in others they fall behind as much as a quarter of a standard deviation. Building on this striking pattern, the goal of this dissertation is to examine the role of the school context for gender differences in education and thereby challenge the view of boys as universally disengaged from school and opposed to authority.

Research on the effect of schools dates back to the 1966 Coleman report and developed out of the concern for equality of educational opportunity by social class and race. This original focus and much subsequent work condemned the unequal access to high quality schools for black and white kids and called for the desegregation of schools. Now that a growing gender gap in educa-

tional attainment has emerged, it is natural 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. For this purpose, the three papers in this dissertation each examine different aspects of the role of the school context for gender differences in education.

The first article "*School Context and the Gender Gap in Educational Achievement*" begins by developing a theoretical argument about the role of peers in school for the educational performance of boys and girls. Building on theories about gender identity and reports from prior ethnographic classroom observations, the article argues 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. To evaluate this argument, the paper uses a quasi-experimental research design that estimates the gender difference in the causal effect on test scores focusing on peer SES as an important school resource. The quasi-experimental research design is based on the argument that the assignment to 5th grade classrooms within Berlin schools is as good as random, and I 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 my central hypothesis, and other analyses confirm the proposed mechanism as the likely explanation of the gender differences in the causal effect. A second quasi-experimental case study based on data from Chicago public schools reaffirm these findings.

The findings from the first paper speak to a long tradition of research about



the important role of neighborhood, school, and peer effects and indicate that boys benefit particularly from school resources that create a learning orientated environment. Yet, the empirical evidence for the success of policy interventions that aim to place families in low-poverty neighborhoods or move students to better schools is mixed and partly shows larger benefits for girls. Most prominently, the Moving to Opportunity (MTO) experiment, which randomly gave low-income families in high-poverty housing projects the opportunity to move to low-poverty neighborhoods, provoked a debate about the lack of neighborhood effects for many of the outcome measures (Kling et al. 2007; Sampson 2008) with larger benefits for girls than boys (Kling et al. 2005; Clampet-Lundquist et al. 2011). In the second article "*Disruptive Change: Peer Effects and the Social Adjustment Process of Mobile Students*", I integrate the literature on student mobility and peer effects to understand these seemingly contradictory findings. Most theories of context effects describe mechanisms that are based on social integration and relations to peers in the local environment as well as knowledge about available resources. Student mobility, however, disrupts existing relations and as a consequence not only has a temporary negative effect on test-score growth itself but also reduces the influence of peers in the years after students change school. This social adjustment process is particularly pronounced for boys who have more problems integrating in the new environment. Accordingly, this temporal adjustment process initially dampens the benefits of transferring to a school with higher achieving peers particularly for boys but over the years students begin to experience the positive effect that is commonly associated with higher quality schools. To study this temporal adjustment process, I use a large-scale administrative dataset and a quasi-experimental research design based on a difference-in-difference, matching approach. This design compares the performance of mobile students who change school with the performance of similar (i.e. matched) students in their

current and previous schools. I find that the effect of peers is substantially smaller for mobile students with a clear temporal adjustment process that is more pronounced for boys. These results have important implications for our understanding of context effects and reconcile the opposing findings in the literature on gender differences in exposure and policy effects.

The final and third paper "*High School Environments, STEM Orientations, and the Gender Gap in Science and Engineering Degrees*" shifts the focus from educational performance to the persisting gender gap in field of study. Despite the striking reversal of the gender gap in education, women pursue science, technology, engineering, and mathematics (STEM) degrees at much lower rates than their male peers do. This paper extends existing explanations for these gender differences and examines two important and related dimensions: the life-course timing of a stable gender gap in STEM orientation, and variations across high schools. I argue that the high school years play an important role for gender differences in orientation towards STEM fields as students develop a more realistic and cognitively grounded understanding of their future work lives. During this period, the gender-specific formation of career aspirations is not only shaped by widely shared and hegemonic gender beliefs but also by the local environment in school. Together these two dimensions highlight the importance of the high school context for the gender gap in STEM degrees and open concrete avenues for policy intervention. Using the National Education Longitudinal Study (NELS), I then decompose the gender gap in STEM bachelor degrees and show that the solidification of the gender gap in STEM orientations is largely a process that occurs during the high school years. Far from being a fixed attribute of adolescent development, however, I find that the size of the gender gap in STEM orientation is quite sensitive to local high school influences; going to school at a high school that is supportive of a positive orientation by females towards math and science can reduce the gender

gap in STEM bachelor degrees by 25% or more.

The three articles highlight different aspects of the role of schools and peers in school for gender differences in education. They broaden our understanding of the processes that explain the pattern of inequality and promote a long tradition of research that has examined the role of schools for different dimensions of social inequality. Together, the three articles contribute to several areas of research with important policy implications. First, the three papers make critical contributions to the debate about the well-publicized under-performance of boys and the persisting gender difference in field of study. The outlined mechanisms attribute a critical role to schools and peers in school, which challenges the focus on deeply entrenched, possibly innate gender differences and instead emphasizes the importance of the local cultural environment. While the findings in terms of educational performance and field of study might initially contradict each other, a broader theoretical argument reconciles the different results: supportive peers or more generally a supportive school environment are particularly beneficial for the disadvantaged group – boys in the case of work habits and educational performance, and girls in the case of STEM interests. Second, our results point to useful directions for new research on policies to raise boys' achievement levels and reduce the persisting gender differences in field of study. It is obviously important to know that school resources affect the educational performance of boys and girls and how this influence depends on exposure dynamics connected to student mobility. At the same time, the pathway analysis shows that high school is the decisive life period during which the gender gap emerges, and the examination of variations across schools indicates that the local context in high school plays an important role for the gender gap in orientations towards STEM fields. These findings have a number of major policy implications that raise the achievement level of boys and girls, provide opportunities to move to different schools, and try to attract boys and girls to

a STEM oriented career path. Finally, the three papers make a methodological contribution to the literature on the estimation of causal effects. They illustrate how a detailed study of and a close attention to the relevant selection processes can facilitate the design of quasi-experiments, which provide a promising alternative to the focus on regression and matching methods.

## Article 1

# School Context and the Gender Gap in Educational Achievement<sup>1</sup>

*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*

---

<sup>1</sup>This article without Appendix C was previously published as Legewie, Joscha, and Thomas A. DiPrete. 2012. "School Context and the Gender Gap in Educational Achievement." *American Sociological Review* 77(3):463-85.

*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.*

## **1.1 Introduction**

Today, boys dominate among high school dropouts, special education students, and literally any failed or special needs category throughout adolescence. The notorious under-performance 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 po-

tential 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. In addition, Appendix C contains a second quasi-experimental case study based on data from Chicago public schools, which replicates the main findings discussed in the main text of this article.

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.

## 1.2 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 1990a; 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 au-



thority), 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. 1969b; Jencks and Mayer 1990a; Rumberger and Palardy 2005a, 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 historically 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. 2012a). 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 2005a).

### **1.2.1 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), SchÄ¶ps 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-WÄ¶rttemberg), Dresel et al. (2006) found substantial variation in the educational gender gap across schools and classes, while SchÄ¶ps 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.

### **1.2.2 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 be-

gin 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 1997a). 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 Ghail 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.<sup>2</sup>

---

<sup>2</sup>Stereotypical 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

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.<sup>3</sup>

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

---

actual performance and thereby deter women from a career in science, math, or engineering.

<sup>3</sup>These 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.

why do you care? Why does it have to be better? Nothin' wrong with a normal grade!"

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 1997b; 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 Ghaill 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. 1969a; Jencks and Mayer 1990a; Rumberger and Palardy 2005a, 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.<sup>4</sup>

### 1.3 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.<sup>5</sup> 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.

---

<sup>4</sup>Our 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.

<sup>5</sup>In 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.



The ELEMENT dataset includes at least two classrooms for every school.<sup>6</sup> 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.<sup>7</sup> This 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.<sup>8</sup> 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).<sup>9</sup> The two datasets complement each other and together provide strong internal and external validity for the estimation of causal effects.

---

<sup>6</sup>Elementary 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.

<sup>7</sup>For 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.

<sup>8</sup>Both datasets were obtained from the Forschungsdatenzentrum at the Institute fr Qualittsentwicklung im Bildungswesen (IQB) HU-Berlin.

<sup>9</sup>As 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.

### 1.3.1 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 (Sorensen 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.

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

$$\begin{aligned}
 y_{ijk} = & \alpha_j + \gamma(\text{female})_i + \theta(\text{SES Comp})_k + \delta((\text{SES Comp})_k \times \text{female}_i) \\
 & + \beta_1 y_i^{\text{4th grade}} + \mathbf{X}_i \beta_2 + \mathbf{U}_k \mathbf{f}_3 + \epsilon_{ijk}
 \end{aligned} \tag{1.1}$$

where  $i$ ,  $j$ , and  $k$  are indices for individuals, schools, and classes respectively,  $\alpha_j$  are the school fixed-effects,  $y_i^{\text{4th grade}}$  is the prior achievement of the student

measured in 4th grade, and  $X_i$  and  $U_k$  are sets of control variables on the individual and class level, respectively.<sup>10</sup> The analysis with the PISA-I-Plus dataset omits the variable of prior achievement on the right hand side because of data limitations.<sup>11</sup>

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  $\theta$ , which captures the causal effect of the socioeconomic class composition, and  $\delta$ , which captures the difference in this effect 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).

### 1.3.2 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 chil-

<sup>10</sup>The 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).

<sup>11</sup>Although 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.

dren 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, Â§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.

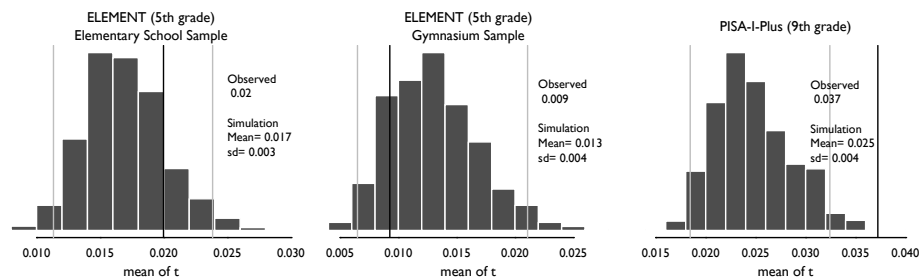
In secondary schools such as those in the PISA-I-Plus data, class-specific tracking based on subject choice such as foreign language is more common, and a higher number of students have to repeat a class compared with elementary school. This creates potentially non-random allocation of students to classrooms 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 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

Figure 1.1: 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.*

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.<sup>12</sup>

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 concluded that randomness plays an important role in the assignment process because they simply have little

---

<sup>12</sup>In addition, all schools reported that class changes within a grade level are extremely rare, and resources are generally allocated equally across classes.



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.

### **1.3.3 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).<sup>13</sup> 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 em-

---

<sup>13</sup>We 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 1.1: Variables in Main Analysis at Individual and Class Level

Variables	Description	Mean	SD	Mean (Male)	Mean (Female)	Std. Diff. in Means
<i>Dependent Variables</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 2 - one parent not born in Germany 3 - both parents born in Germany but child 4 - child not born in Germany	0.64 0.12 0.17 0.06	0.48 0.33 0.38 0.24	0.64 0.12 0.18 0.06	0.65 0.13 0.16 0.06	-0.01 -0.01 0.02* -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: ELEMENTF; 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.

ployed a similar imputation strategy based on the chained equations approach for the PISA-I-Plus dataset.

## 1.4 Results

### 1.4.1 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.<sup>14</sup> 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.

### 1.4.2 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

---

<sup>14</sup>One additional school year corresponds to the estimated test score difference between 5th and 6th grade in the ELEMENT dataset.

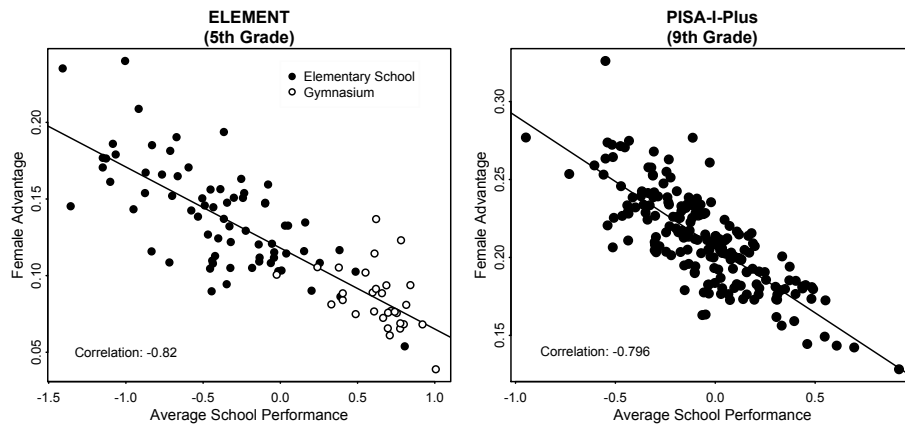


Figure 1.2: Gender Gap and Average Performance across Schools in Standard Deviation

*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). Accordingly, the model used to estimate the gender gap and its variation is specified as  $y_{ij} = \alpha_j + \theta_j \text{Female}_i + \epsilon_i$  whereby  $i$  and  $j$  are the indices for students, and schools respectively, and  $\theta_j \sim N(\gamma_\theta, \sigma_\theta^2)$ .

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 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

Table 1.2: 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.*

reported in the literature on the effects of SES composition (Rumberger and Palardy 2005a; Jencks and Mayer 1990a). 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 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). 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 be-

tween 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.

### **1.4.3 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 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

Table 1.3: 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)
(1) FE/IV - Estimate <i>Instrument: SES comp. of students who never repeated a class</i>	0.008	(0.02)	0.089*	(0.04)	-0.065**	(0.02)
(2) FE/IV - Estimate <i>Instrument: SES comp. of students who didn't go to kindergarten, skipped a class or transferred to school</i>	0.009	(0.02)	0.113*	(0.06)	-0.068*	(0.03)
(3) FE - Estimate (restricted sample) <i>Sample Restriction: Only schools that do not allocate based on ethnicity (24 schools excluded)</i>	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.

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.<sup>15</sup> 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.

#### **1.4.4 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

---

<sup>15</sup>We 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).

is constructed and 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 2005a; Jencks and Mayer 1990a). 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.<sup>16</sup> 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.

---

<sup>16</sup>The measures are constructed from a range of indicators using exploratory factor analysis (see Online Appendix).

Table 1.4: Effects of Gender and SES Composition on School-Related Attitudes and Behavior

		Female		SES Composition		SES Comp. 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.

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 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.

Table 1.5: 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	-1.00***	(0.35)	-0.47	(0.27)	-0.59*	(0.25)

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.

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. 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, 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.

## 1.5 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 de-masculinized 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 simulations, 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 educa-

tion 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.

## **Appendix C Evidence from a Second Quasi-Experimental Case Study**

While the estimates presented in this article provide a clear advantage over regression or matching based methods, they are limited to Germany or even a single German state. Using a unique, administrative dataset from the third largest school district in the US, this appendix provides preliminary evidence from a second quasi-experimental case study and extends the findings in important ways.

### **Data and Methods**

I use an extensive longitudinal database from the Chicago Public Schools (CPS) system assembled by the Consortium on Chicago School Research (CCSR) at the University of Chicago. With more than 400,000 enrolled students and about 600 public elementary and high schools in 2009-2010, CPS is currently the third largest school district in the US. The database consists of different components: (a) the administrative student records for every student enrolled in CPS from the school year 1993/94 to 2005/06. These records include the school and grade identifier for the fall and spring term as well as a limited number of standard demographic characteristics such as gender, date of birth, students' race, their eligibility for free lunch as a measure of parental background, and their special education status. (b) The test file records, which contain different reading and math tests administered over the years. Most noticeable are the Iowa Test of Basic Skills (ITBS) in reading and math, which was taken by almost all students in the spring of grades 3 through 8 over the whole period, and the Tests of Achievement and Proficiency (TAP) in reading and math, which was given to different high school grades up until 2002. (c) Data from a set of school, teacher and student surveys conducted by CCSR in the spring of 1994, 1997,

1999, 2001, 2003, and 2005. These surveys include a range of measures about school related attitudes and behavior, school climate, self-evaluation of abilities, student-teacher relations and other topics. All three components can be perfectly matched over time so that I can follow students in CPS as they move through grades, change school, and improve their reading and math skills.<sup>17</sup>

For the main analysis presented in this appendix, I restrict our sample to 4th grade students who participated in the ITBS on their grade level<sup>18</sup>, and did not change the school or grade within 4th grade. These restrictions reduce the sample size to about 330,000. The construction of our instrument (see below) makes it necessary to exclude certain years from the analysis so that our final sample consist of about 200,000 students in grade 4 from the school year 1996/97 to 2004/05.

### **Estimating Compositional Peer Effects: Analytic Strategy**

Similar to Hoxby (2000) and others (e.g. Hanushek and Rivkin 2009), I exploit the variation in peer composition of adjacent cohorts within a school within a grade to estimate the causal effect of SES composition and peer ability. This variation is illustrated in Figure A1.3a, which shows the peer SES (for the definition see below) of adjacent cohorts in an example school between 1993 and 2005. The line in Figure A1.3a refers to the school-grade specific mean so that the variation in SES composition of adjacent cohorts within a school is the deviation of the observed SES composition index from this line.

Hoxby (2000) uses data from Texas public schools and such cohort-to-cohort variation in gender composition to address the problem that students self-select into schools. She argues that these differences between adjacent cohorts in gender composition are largely the consequence of random variations in the

---

<sup>17</sup>Note that teachers can only be matched with schools but not individual students and can not be followed over time.

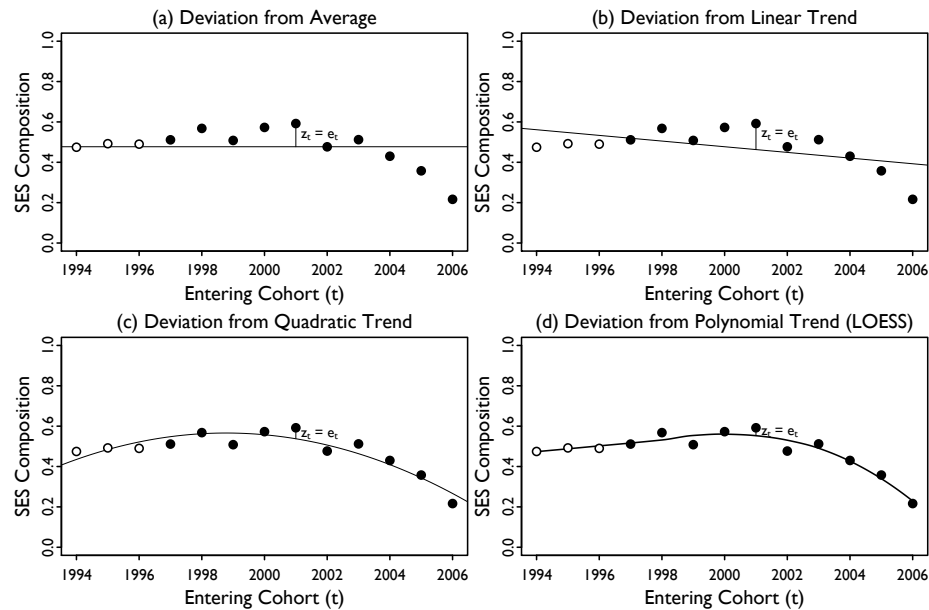
<sup>18</sup>A small number of students do not take the test on their grade level.

birthrate of boys and girls so that “*some* variation in adjacent cohorts’ peer composition *within a grade within a school* [...] is idiosyncratic and beyond the easy management of parents and schools” (Hoxby 2000, XY). As emphasized by Epple and Romano (2011), this strategy is particularly effective when applied to gender composition: “An appealing aspect of the study of the effects of gender variation is that one would not expect differences across cohort in the proportion female to be correlated with other observables. In the absence of endogenous changes in schools in response to observed changes in gender, the study of gender variation thus circumvents complexities involved in studying peer effects when the peer variable of interest may vary both for systematic and idiosyncratic reasons.” (Epple and Romano 2011, XY). Accordingly, the estimation of SES or ability peer effects based on cohort-to-cohort variations might be problematic considering systematic, non-random variations that are connected to changes in the school environment, parents reactions to the composition of certain cohorts or other factors. To circumvent these potential problems, I adopt an alternative strategy based on instrumental variables and the decomposition of the cohort-to-cohort variations into different components.

For a certain grade, the cohort-to-cohort variation in peer SES or ability as shown in Figure A1.3a can be decomposed in (a) the variation coming from differences between entering cohorts in first grade and (b) the variation coming from mobility of students between schools and grades. School mobility is pervasive in schools throughout the United States and particularly in urban, minority-dominated school districts (Rumberger 2003). Chicago is no exception in this regard. Only about 75% of students do not change their elementary school from one school year to the next and only 50% remain enrolled in the same elementary school over a three year period (Kerbow et al. 2003a, 158). While these rates have declined over the last decades, student mobility remains pervasive in CPS (Torre and Gwynne 2009). A majority of these school moves



Figure A1.3: Variation in SES composition of Adjacent Cohorts within a School within a Grade with School-Grade Average



Note: The figure shows the SES composition in 4th grade in one example school from fall 1994 to fall 2006. The lines refer to the school-grade specific mean, linear and quadratic trend respectively. The variation in SES composition of adjacent cohorts is the deviation of the observed mean (dots) from the school-specific mean (a) or trend (b and c). This residual variation is the crucial identifying information based on which the instrument  $z$  is defined for each cohort in a particular school. The white points indicate the years that are dropped from the final analysis because of the way in which the instrument is constructed.

are driven by residential mobility connected to housing costs or family instability but many also report problems or dissatisfaction with a particular school. Based on a survey of over 13,000 6th grade students in Chicago, Kerbow and colleagues document these different reasons for changing a school: “A majority of school changes (58%) were associated with a residential change. Beyond these factors, however, many students also cited school-related concerns. In fact, 42% of students cited only school-related concerns, such as lack of safety or limited academic opportunities, as reasons for changing schools. In sum, a significant element of student mobility was generated by dissatisfaction with

their current school or the possibilities of greater satisfaction at another school.” (Kerbow et al. 2003a, 159 also see Torre and Gwynne 2009, 4). Accordingly, the experience in school is an important factor that influences the mobility of students between schools. While many residential moves of families might be exogenous to a child’s school, Kerbow’s research documents that school mobility is also a school-related decision made by parents to increase the educational opportunities of their children or in response to problems at a particular school. Some of these school-related reasons for student mobility are connected to the school as a whole (e.g. school safety) so that a comparison of different cohorts within the same school controls for these factors. Torre and Gwynne (2009, 4), however, show that bad grade, and problems with other students or teachers also play an important role. These factors are largely cohort specific so that the focus on cohort-to-cohort variations in the composition of the student body does not address this problem. When students selectively change schools based on the experiences they made with other students and teachers in their classroom, cohort-to-cohort variations in the mobility of students are not a credible source of variation for the estimation of compositional peer effects. As a consequence of the initial assignment to a particular cohort, parents might selectively withdraw their children based on their experiences in a certain school. These sources of non-random selection might lead to biased estimates of peer SES when they are based on cohort-to-cohort variations within schools.

Instead, we focus on the variations coming from differences between entering cohorts in first grade. While these variations show a temporal pattern connected to the changing popularity of a school or shifts in the composition of the neighborhood (also evident in the school from Figure A1.3), these variations also have an important idiosyncratic component that is arguably beyond the control of parents, and school principals. As previously argued by Hoxby, “*some variation in adjacent cohorts’ peer composition within a grade within a*

*school* [...] is idiosyncratic and beyond the easy management of parents and schools" (Hoxby 2000, XY). Accordingly, my analysis are based on the variation coming from differences between entering cohorts in the fall of 1st grade after adjusting for some general temporal pattern and before any mobility between schools and grades occurred. We can obtain estimates based on this variation using instrumental variables. In particular, I treat 4th grade SES composition as endogenous and instrument the 4th grade peer SES in school  $s$  with the composition of this cohort when it entered the school 3 years earlier. For the fall of the school year 2004/05, for example, I instrument the 4th grade SES composition at school  $s$  with the composition of the 1st grade in fall 2001 - the year when this cohort entered the school. In the two stage framework, this can be described as

$$\widehat{D}_{sc}^{4.grade} = \alpha_s + \gamma_c + \delta D_{sc}^{1.grade} \quad (1.2)$$

$$y_{isc}^{4.grade} = \alpha_s + \gamma_c + \theta \widehat{D}_{sc}^{4.grade} + \epsilon_{isc} \quad (1.3)$$

where equation 1.2 represents the first stage regression of 4th grade SES composition on the school-cohort level on the instrument  $D_{sc}^{1.grade}$  together with school as well as cohort fixed effects. Equation 1.3 represents the second stage regression of some measure of performance in 4th grade on the fitted values of  $D_{sc}$  from the first stage regression together with school and cohort fixed effects. Both equations also control for a set of covariates on the individual and school-cohort level  $\mathbf{X}_{isc}\beta_1 + \mathbf{U}_{sc}\beta_2$ , which I omitted from the equation for simplicity.<sup>19</sup>

The problem with the specification so far is that it does not adjust for any trend in the composition of adjacent entering cohorts. It basically assumes that the deviations from the school-specific mean as illustrated for the 2001 cohort in Figure A1.3a are as good as random. As argued before, however, the composi-

---

<sup>19</sup>Note that the two variables of SES composition in 1st and 4th grade are measured in the fall of a school year whereas the dependent variable is measured in the spring term.

tion of adjacent entering cohorts is partly driven by naturally occurring differences between cohorts and partly by parents who select into specific schools in a specific year. Only the first source of variation is arguably random. Schools might, for example, exhibit a certain trend in the composition of their student body which is known to parents. The SES composition in the school in Figure A1.3, for example, is decreasing slightly over the years. Such a development might occur because of local changes in the composition of the neighborhood. It is reasonable to argue that involved parents are aware of this trend, which ultimately steers them away from this specific school. In general, the trend might not only influence parents schooling decisions but might be connected to some unobserved characteristics, which are also related to the dependent variable so that the estimate of the causal effect is biased. To address this problem, I model a group-specific time trend - i.e. a time trend that is specific to each school. This is illustrated in Figure A1.3b, and A1.3c, which show a linear and quadratic trend respectively. In these two cases, the relevant variation is the deviation of the observed SES composition (dots) from the linear or quadratic trend line so that the instrument is defined by the error term from a school-specific regression line as illustrated in the two figures for the 2001 cohort. Statistically, this can be implemented by adding  $\beta_s year$  for the linear trend or  $\beta_s^1 year + \beta_s^2 (year \times year)$  for the quadratic trend to both the first and second stage regression of the instrumental variable model reflecting the school-grade specific time trend.

This approach relies on the assumption that the time trend on the school-grade level is either linear or quadratic or at least that the trend known to parents and other actors who influence the school decisions follows this functional form. It could still be the case that high SES kids in a specific year are attracted to a specific school by some year-specific school resources (e.g., a popular teacher joined the school, or a popular principal joined the school, or some-

thing). It could also be the case that the functional form of the trend does not adequately reflect the way in which school change is perceived by parents. These possibilities, however, seem unlikely especially considering that we are talking about elementary schools. To a much higher extent than for high schools, children usually attend their neighborhood school, which makes natural occurring variations between schools the more important source of variation in the cohort-to-cohort differences in SES composition of entering cohorts within a school.

**Plausibility of Identification Strategy** The estimation of causal effects using instrumental variables can be a powerful estimation strategy but the results crucially depend on the quality of the instrument and rely on two core assumptions (Angrist and Pischke 2008, 116ff). First, the instrument must be correlated with the treatment variable conditional on the covariates. Second, the instrument is as good as randomly assigned conditional on the covariates and only related to the outcome variable through the treatment (*exclusivity assumption*).

The first assumption is easy to test statistically. Common criteria have been established in the literature. It implies that the instrument  $D_{sc}^{1,grade}$  in the first stage regression defined in equation (1.2) has a clear effect on the treatment indicator  $D_{sc}^{4,grade}$  conditional on the covariates. These covariates include the control variables on the individual and school-cohort level (see below) as well as the school and cohort fixed effects and the school specific time trend. Despite the pervasive nature of school mobility in Chicago, the correlation between the composition in 4th grade of a certain school in a certain year (the treatment indicator) with the SES composition of this particular cohort three years earlier when it entered the school (the instrument) is extremely high on the bivariate level. The results from the first stage regression show that the effect of first grade composition on fourth grade composition is highly significant at the

X percent level and substantial even after conditioning on the covariates, the fixed effects, and the school specific time trend (available from the author). The partial F statistic for the exclusion of the instrument(s) and the partial r-square value indicate a very strong instrument, and the F-statistic is clearly above the commonly used threshold of 10 (Staiger and Stock 1997).

The second assumption is critical for the estimation of causal effect with instrumental variables and can not easily be assessed with statistical methods. In general, this assumption is violated if the instrument has a direct relationship with the outcome or is related to an omitted variable that influences the outcome conditional on the covariates. This assumes, for example, that the school fixed effect and the school specific time trend adequately control for the selection of students into specific schools in a specific year. It implies that the remaining variation of the cohort-to-cohort differences in SES composition – i.e. the deviation of the observed composition from the trend line in Figure A1.3 – is as good as random. Given that most children in Chicago attend the elementary school in their neighborhood and that parents only have limited information about the cohort-to-cohort variation of entering cohorts in a certain school, this assumption seems plausible. Nonetheless, it remains possible that the cohort-to-cohort variation within a school conditional on the school-specific time trend is not as good as random. It might well be the case that the time trend does not sufficiently capture the ways in which parents perceive changes in their local neighborhood schools or that parents are attracted to particular schools in a certain year.

To evaluate whether the differences across adjacent cohorts are as good as random, I compare the observed variation across cohorts within a school with the variation under random assignment of students to cohorts. For some binary trait such as gender, the variance in the share of this trait across cohorts is determined by the frequency of the trait in a school and the size of different

cohorts. In particular, for some trait with probability  $\pi_j$  in school  $j$ , the variance in the share of this trait across cohorts is

$$\pi_j(1 - \pi_j) \frac{1}{K_j} \sum_{k=1}^{K_j} \frac{1}{n_{kj}} \quad (1.4)$$

where  $n_{kj}$  is the size of cohort  $k$  and in school  $j$ , and  $K_j$  is the number of cohorts in school  $j$  (Epple and Romano 2011, 1127). All terms in this formula are known for observed characteristics so that we can easily compare the observed variance with the variance under randomness calculated based on Equation 1.4 and determined by the share of the trait in the school  $\pi_j$  and the size of the different cohorts  $n_{kj}$ . Such a comparison allows us to precisely evaluate whether the variation across cohorts within a school are as good as random. If, for example, students select into schools in specific years based on the composition of the entering cohort in that year, the movement of popular teachers and principles, or other cohort-specific factors, we would expect that the variance across cohorts is larger than the variance under random assignment. If the variations are purely driven by idiosyncratic variation across cohorts related to the frequency of births or other factors, the observed variance should be roughly the same as the analytically derived variance.

Figure A1.4 and A1.5 compare the observed variance with the analytical variance under random assignment across the 1994 to 2006 entering cohorts of public elementary schools in the Chicago Public School district.<sup>20</sup> The figures plot the analytical variance calculate for each school based on Equation 1.4 on the x-axis against the observed variance on y-axis. The scales on both axes are the same so that the straight line through the plot indicates equal variance or a variance ratio of one. Hollow circles represent schools with a significant

---

<sup>20</sup>Schools with less than five observed cohorts are omitted from these graphs as well as the later analysis.

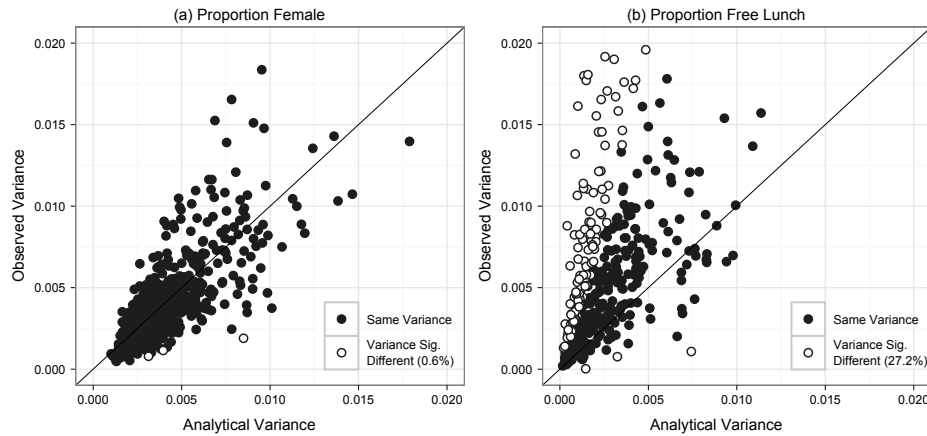


Figure A1.4: Variance in the Share of Female Students across Cohorts in CPS Schools

difference between the observed and analytical variance.<sup>21</sup> Figure A1.4a first shows the variance in the share of girls across entering cohorts. The schools are evenly distributed above and below the line indicating that in some schools the observed variance is larger than expected under randomness and in others it is smaller. In less than one percent of the schools, we observe a significant difference between the observed and analytical variance. The finding strongly supports Hoxby's (2000) use of idiosyncratic variation in the gender composition of cohorts within schools to estimate the effect of the proportion of girls among peers. As previously argued (also see Epple and Romano 2011, 1127), this strategy is less convincing for other traits such as the composition by family background or performance. Figure A1.4b supports this argument. It shows that for the proportion of students who receive free lunch, the observed variance across cohorts is mostly larger than the variance based on randomness (as indicated by the fact that most points are above the line). Indeed, in over 27% of the schools the observed variance is significantly different (and mostly

<sup>21</sup>The significant tests is based on a two-tailor, one-sample F-test for the equality of variance using 0.05 as a threshold.



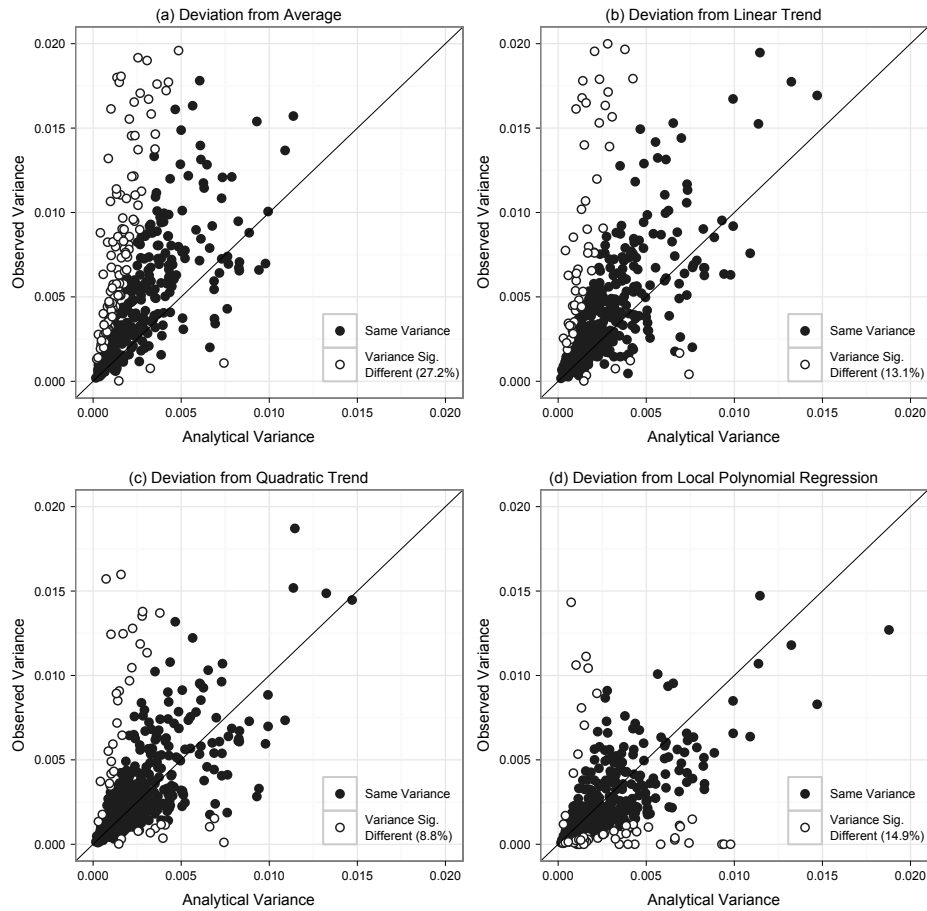


Figure A1.5: Variance in the Share of Students with Free Lunch across Cohorts

greater than) the analytical variance (hollow circles). This pattern indicates that the socioeconomic composition of cohorts within schools is not as good as random. It shows that the variation across cohorts within schools is larger than the variation obtained under randomization.

In the last section, I have argued that trends in the composition of cohorts over the years play an important role as a potential source of bias that drives the selection of student into certain schools. These systematic changes in the composition of entering cohorts over the years might also be responsible for the

high number of schools with a larger than expected variation across cohorts. Figure A1.5 evaluates the proposed strategy to control for school-specific time trends and compares the analytical variance with the observed variance after adjusting for different time trends. Figure A1.5a first reproduces A1.4b with the high number of schools that have a larger than expected variance (27.2%). As shown in Figure A1.5b and c, adjusting for a linear or quadratic time trend clearly improves the situation. The schools are more evenly distributed below and above the line (particularly after adjusting for the quadratic time trend) and the proportion of schools with a significant difference between the analytical and observed variance reduces from 27.2% to 13.1% for the linear and 8.8% for the quadratic trend. While the proportions are still slightly above 5% (the threshold used in the statistical test), the figures show that the variation in SES composition across entering cohorts within schools can be considered as good as random after adjusting for some general time trend considering that the test is very precise. To rule out the possibility that the estimated effects are driven by the small number of schools with a larger than expected variation across cohorts, I also perform the same analysis based on a restricted sample that omits the schools with a significantly larger than expected variance across cohorts.

Finally, A1.5d compares the analytical and observed variance after adjusting for a school-specific, local polynomial regression (LOESS). The polynomial time trend has a highly flexible functional form and the fitting procedure gives higher weight to closer observations (Figure A1.3d illustrates a LOESS curve for an example school). Such a time trend, however, seems to over-adjust insofar as a high number of schools have a lower than expected variance and the number of schools with a significant difference in the variance increases again to nearly 15%. Accordingly, a linear and in particular a quadratic time trend seems to be more appropriate to adjust for the changing popularity of schools or shifts in the composition of the neighborhood.

The comparison of the observed variance across cohorts with the analytically derived variance under completely randomized assignment provides a precise test of the proposed estimation strategy. The results show that, in contrast to the share of female students, the variance for the share of students who receive free lunch across cohorts is slightly larger than expected. Adjusting for a linear and particularly a quadratic time trend, however, largely solves the problem. After taking such a general trend into account, the observed variance closely resembles the analytical variance for most schools, which strongly supports the argument that cohort-to-cohort variations in the SES composition can be used for a quasi-experimental estimation strategy.

## **Variables**

My analysis uses reading and math test scores in 4th grade from the Iowa Test of Basic Skills (ITBS) as the main outcome variables (see table A1.6 for descriptive statistics). The ITBS was designed by Riverside Publishing and was given to all students in grade 3 through grade 8 up until the spring of 2005. 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.

The focal treatment variable is the socioeconomic (SES) composition of the student body, which is measured at the school-grade-year-level as the average social status on a scale constructed from three indicators. The most common measure of family background used in administrative datasets is the student's eligibility for free or reduced-price lunch as defined by the federal state. In the context of CPS, however, this measure is not satisfying because the extraordinary high concentration of poverty. In more than 50% of the schools, for example, the proportion of students who receive free lunch is above 85% so that the measure does not show a lot of variation across schools. To address this problem, I constructed a SES measure from a set of indicators and averaged these

Table A1.6: Variables in Main Analysis at Individual and Class Level

Variables	Description	Mean	SD	Mean (Male)	Mean (Female)	Std. Diff. in Means
<i>Dependent Variables</i>						
Reading Test Scores	4th grade reading test scores	196.19	24.15	194.46	197.86	.14
Math Test Scores	4th grade math test scores	198.77	19.32	198.58	198.95	.02
<i>Independent Variables (Individual Level)</i>						
Prior Achievement	3rd grade reading test scores	182.12	21.13	180.74	183.45	.13
Prior Achievement	3rd grade math test scores	184.08	16.44	184.11	184.05	0
Female	0 - Male; 1 - Female	0.51	0.50			
Family Background	0 - not eligible for free lunch; 1 - eligible for free lunch	0.78	0.41	0.78	0.79	.03
	avg. free lunch across years	2.58	0.54	2.57	2.58	.04
	mother's education (7-point scale)	3.39	1.75	3.42	3.35	-.04
Race	1 - White	0.10	0.30	0.10	0.10	-.01
	2 - African-American	0.53	0.50	0.52	0.54	.03
	3 - Native American	0.00	0.04	0.00	0.00	.01
	4 - Asian	0.03	0.17	0.03	0.03	-.01
	5 - Latino	0.34	0.47	0.34	0.33	-.02
Mover	0 - did not change school; 1 - changed school	0.30	0.46	0.31	0.30	-.01
Age	continuous variables					
Interaction Terms	School x Year (School specific time trend)					
Fixed Effects	School Fixed Effects					
	Year Fixed Effects					
<i>Independent Variables (School-Grade-Year Level)</i>						
	4th grade (fall) avg. SES at school-grade-year level (aggregated)	0.65	0.16	0.65	0.66	.02
	1st grade (fall) avg. SES at school-grade-year level (aggregated)	0.65	0.17	0.65	0.65	.03

Source: CPS Data data; n=XXXX; \* p < 0.05, \*\* p < 0.01

Note: The descriptive statistics show in this table are based on the same sample as the final analysis. The difference in means refers to the mean for boys minus the mean for girls divided by the pooled standard deviation. Note that all the continuous variables are standardized for the final analysis.

separate indicators weighting each one equally. For the first indicator, I averaged the free lunch eligibility of a student across the different years the student is observed in order to get a more stable measure of family background. Accordingly, a student who received free lunch in two years but not in the third gets a value of 0.66. I then averaged this more stable indicator of family background on the school-grade-year level. The second indicator is the proportion of minority students in a school-grade-year defined as the proportion of black students. The third indicator is a seven point scale for mothers education obtained from the student questionnaires. This variable only exists for a subset of the students in our sample and was imputed for the other cases using the three other indicators. Before averaging, all of the four indicators were scaled from 0 to 1 and oriented so that a higher value reflects an lower average status.

In addition, we use a number of control variables both at the individual and at the school-grade level. These variables are described in Table A1.6 together with some descriptive statistics. All independent, continuous variables are standardized to have a mean of 0 and a standard deviation of 1 across the combined sample of males and females in both dataset.

## **Results**

Table A1.7 presents the results for the instrumental variable-fixed effect regressions of 4th grade reading test and math scores on school-grade level SES composition, gender and other control variables. The table also shows separate models with and without a control variable for prior performance. Across the four models, the results for SES composition indicate a positive effect of about 0.15 standard divisions for the raw score and of about 0.08 (reading) and 0.11 (math) for the gain scores. These results are in line with the findings from the Berlin case and also other studies such as Crosnoe (2009). More importantly, the interaction between peer SES and female is negative and significant for all four

Table A1.7: Gender Differences in the Effect of Peer SES

Model	Prior Perf.	Female		SES Comp.		SES Comp. x Female	
		coef.	(se)	coef.	(se)	coef.	(se)
1. IV/FE - Estimate (Outcome: Reading)	yes	0.055***	(0.00)	0.077***	(0.02)	-0.010***	(0.00)
2. IV/FE - Estimate (Outcome: Reading)	no	0.139***	(0.00)	0.157***	(0.03)	-0.021***	(0.00)
3. IV/FE - Estimate (Outcome: Math)	yes	-0.023***	(0.00)	0.105***	(0.02)	-0.024***	(0.00)
4. IV/FE - Estimate (Outcome: Math)	no	0.026***	(0.00)	0.140***	(0.03)	-0.052***	(0.00)

n=192,014; Standard errors in parentheses; \* p < 0.05, \*\* p < 0.01, \*\*\* p < 0.001

*Note: The first stage results show that the two instruments are highly correlated with SES composition as the treatment. The F-statistics are over XXX (highly significant), which is far above the commonly used threshold of 10. The additional control variables are described in table A1.6. The number of cases is 192,014.*

models. Compared to the German results, the size of the interaction effect is slightly smaller but still substantial considering the the gain scores accumulate over the years. Overall, the estimates validate my findings and provide strong evidence that boys are more sensitive than are girls to the important school resource of classroom SES composition.

## Conclusion

In this appendix, I have presented results from a second quasi-experimental case study for the gender differences in the effect of peer SES. The results presented here reconfirm my findings from the Berlin case and provide further evidence for the argument. Using data from the Chicago Public Schools (CPS), they reveal the same pattern for the US and also extend them to math test scores.

The results from the two quasi-experimental case studies presented here

show that boys benefit particularly from peers in their environment indicating that boys are more sensitive to school resources that create a learning oriented peer culture. This observed gender difference in the causal effect of peer SES is at least partly responsible for the variations of the gender gap across schools, provides support for our theory by examining the main macro level implication of the argument, and has important policy implications.





## Article 2

# Disruptive Change: Peer Effects and the Social Adjustment Process of Mobile Students <sup>1</sup>

*Student mobility is pervasive across school districts in the U.S. and has important implications for the social integration of students in peer groups. Yet previous research on neighborhood, school, and peer effects largely ignores the role of changes in family residence and related school transitions, despite the fact that most theories of peer influence attribute a critical role to the social integration of students and the influence of their peer networks. In this article, we integrate the literatures on student mobility*

---

<sup>1</sup>This research was supported by a grant from the American Educational Research Association which receives funds for its "AERA Grants Program" from the National Science Foundation under Grant #DRL-0941014. Opinions reflect those of the author and do not necessarily reflect those of the granting agencies.

*and peer effects, arguing that student mobility not only has a temporary negative effect on test-score growth, but that it also alters context effects and the influence of peers in the years after students school change. To study this temporal adjustment process, we use a large-scale administrative dataset and a quasi-experimental research design based on a difference-in-difference, matching approach. We find that the effect of peers is substantially smaller for mobile students with a clear temporal adjustment process and that effects are more pronounced for boys. These results have important implications for our understanding of context effects and reconcile opposing findings in previous research.*

## **2.1 Introduction**

Ever since the publication of landmark studies such as William Julius Wilson's *The Truly Disadvantaged* (1987) or the Coleman report (Coleman 1966), neighborhood and school effects have been a central research topic across the social sciences. At the core of this agenda is the argument that exposure to high poverty neighborhoods places teens at risk (Sampson et al. 2002) and peers in school shape educational outcomes (Epple and Romano 2011; Sacerdote 2010). Previous research on neighborhood, school, and peer effects, however, largely ignores the fact that students frequently change school and families move, processes that have important implications for the social integration of students and the potential influence of their peers. Along these lines, the empirical evidence for the success of policy interventions that aim to place families in low-poverty neighborhoods or move students to better schools is mixed (Deluca and Dayton 2009; Sampson 2008; Angrist and Lang 2004; Cullen et al. 2005).

In this article, we integrate the literatures on student mobility and peer effects to broaden our understanding of context effects and to address some of

the contradictory findings in the literature. Most theories of context effects describe mechanisms that are based on social integration and relations to peers in the local environment as well as knowledge about available resources. Student mobility arguably disrupts existing relations and exposes students to an entirely new environments. Considering that student mobility is pervasive across urban school districts in the U.S., this disruptive nature of student mobility has fundamental implications for our understanding of context effects. Based on this argument, we assert that student mobility not only has a temporary negative effect on test-score growth itself, but that it also alters context effects in the years after students move to a new school. Accordingly, this temporal adjustment process initially reduces the benefits of transferring to a school with higher achieving peers. It is only with time that students begin to experience the positive effect that is commonly associated with higher quality schools. This temporal perspective emphasizes the adjustment process inherent within school change and broadens our understanding of peer effects. It also helps us to reconcile the established findings in the literature on neighborhood and school effects with the mixed evidence from policy interventions that aim to place families in low-poverty neighborhoods or move students to better schools such as voucher programs or the Moving to Opportunity experiment. Throughout our discussion, we also build on recent work that highlights important gender differences in exposure and mobility effects (Kling et al. 2005; Clampet-Lundquist et al. 2011; Legewie and DiPrete 2012) and carefully delineate how boys and girls may adopt differently to a new school.

To evaluate our argument about the temporal adjustment process of mobile students, we use a large-scale administrative dataset from Chicago Public Schools (CPS) that includes all students between 1993 to 2006 so that we can track students when they transfer between public schools within the Chicago school district. Using a quasi-experimental research design based on a com-

bined difference-in-difference, matching approach, we find that the effect of peers is substantially smaller for mobile students with a clear temporal adjustment process, and that these effects are more pronounced for boys. Boys who transfer to a school with higher achieving peers initially learn at the same rate as their peers in the previous school, despite the improved learning environment. Only after several years do they begin to experience the benefits associated with higher performing peers. For girls, this temporal adjustment process is less pronounced, with a positive effect almost immediately after transferring to a higher quality school but only a small increase of this effect over time.

These findings support our central argument that temporal adjustment processes affect boys more than girls and contribute to several research areas. First, the study broadens our understanding of context effects by highlighting the important role of student mobility and exposure dynamics. It also reveals a possible problem in previous research that has ignored mobility as a potential confounder of context effects. Second, understanding the temporal adjustment process helps us to reconcile the seeming contradiction between the literature on neighborhood and school effects on the one hand, and the mixed evidence for the effectiveness of certain policy interventions on the other. Finally, our findings speak to the recent debate on gender differences in context and policy effects.

## **2.2 Peer Effects and Educational Outcomes:**

### **Theory and Research**

In the 1954 *Brown vs. Board of Education* landmark decision, the Warren Supreme Court unanimously ruled that “separate educational facilities are inherently unequal” declaring the *de jure* racial segregation of public schools for unconstitutional. The decision marked a watershed in the ongoing political and public

tensions connected with a simple question that has concerned parents, policymakers and scholars alike: *“How do peers in school influence educational outcomes?”*.

Over the last decades, this question has been an underlying issue in many important debates on educational policy ranging from the desegregation of schools, to school choice and the tracking of students within schools. At the same time, peer effects in school have played an important role in educational research as a major determinant of both educational and non-educational outcomes. Broadly defined, peer effects are based on any influence class- or school-mates exert on a particular student (Epple and Romano 2011, 1054f). They include both direct effects that work without changing the behavior or attitudes of students as well as indirect effects. Previous estimates of peer effects generally focus on the composition of the school or classroom in terms of gender, class-background, race, or ability and report modest and statistically significant effects. Far more than appreciated in early research, a number of recent studies acknowledge that the estimation of peer effects is challenging (Angrist and Pischke 2008, 193-97) because students select into schools and unmeasured confounding variables such as teacher quality affect student outcomes. In fact, a series of experimental (Duflo et al. 2011; Carrell et al. 2009) and quasi-experimental (Legewie and DiPrete 2012; Hoxby and Weingarth 2005; Imberman et al. 2012b) studies use a variety of estimation strategies to overcome the limitations of previous work and largely report that peers play an important role in shaping education outcomes (for a broad overview of the literature see Epple and Romano 2011; Sacerdote 2010).

At the same time, the empirical evidence for the success of policy interventions that aim to place families in low-poverty neighborhoods or move students to better schools is mixed (Deluca and Dayton 2009). Most prominently, the Moving to Opportunity (MTO) experiment, which randomly gave

low-income families in high-poverty housing projects the opportunity to move to low-poverty neighborhoods, provoked a debate about the lack of neighborhood effects for many of the outcome measures (Kling et al. 2007; Sampson 2008) with the most recent findings showing long-term effects on subjective well-being but not economic self-sufficiency (Ludwig et al. 2012). Similarly, the evidence for positive effects of school desegregation programs that rely on vouchers or bus students to schools across districts is controversial. Angrist and Lang (2004), for example, find “modest and short lived” peer effects from a desegregation program in Boston that sends student from inner-city schools to more affluent suburbs. Recent findings in this literature also indicate that girls but not boys benefit from moving to a higher-resource environment (Hastings et al. 2006; Clampet-Lundquist et al. 2011; Kling et al. 2007) while other research suggests that boys are more sensitive to peers in their context (Legewie and DiPrete 2012; Carrell and Hoekstra 2010).

In this paper, we examine the role of student mobility and exposure dynamics for the effect of peers on education outcomes carefully delineating by gender. Student mobility is pervasive across urban school districts in the U.S. so that students regularly change school. These transfers between schools shape the experience of students and how they are influenced by peers so that these dynamics have important implications for our understanding of context effects. To elaborate this argument, we first review mechanisms of peer effects and then discuss how the social adjustment process inherent in changing school alters peer effects.

### **2.2.1 Mechanisms of Peer Effects**

A number of theories describe mechanisms that explain the effect of peers in school on educational outcomes. The most prominent account is based on peer socialization or contagion mechanisms (“socialization effects”). In gen-

eral, these processes refer to any situation in which the behaviors or attitudes of peers spread to other students and increase the likelihood that they adopt similar attitudes or behaviors (Crane 1991; Heilbron and Prinstein 2008, 169-73). “The mechanism that explains this socialization effect may vary, however. For instance, these processes could include explicit reinforcement from peers, social modeling (i.e., vicarious learning whereby individuals model behaviors based on their observations of others), and/or responses to perceived norms of members of a given peer group” (Heilbron and Prinstein 2008, 170). As such, peer socialization processes encompass many of the popular sociological accounts such as Coleman’s early focus on the value and reward system among students (Coleman 1960, 1961). In his classical study on the adolescent society, he described the value and rewards attached to certain attitudes and behavior in school and argued that the attention and effort in school are shaped by this status system established in peer groups. Other explanations that build on peer socialization processes refer to a learning-oriented peer culture created by highly motivated and capable students, which are attributes that are more common among students from higher SES background or among higher ability students (Jencks and Mayer 1990b; Rumberger and Palardy 2005b; Legewie and DiPrete 2012; Goldsmith 2011).

Other theories focus on institutional mechanisms (Lee et al. 1991) or reference group effects (Goldsmith 2011; Dai and Rinn 2008). Institutional mechanisms refer to processes that work through the teacher or school in general (Lee et al. 1991). The behavior of other students in the classroom, for example, might influence the expectations of the teacher, teaching methods, or even the curriculum. The parents of other students might be heavily involved in the school influencing the learning experience of all students. Reference group processes, in contrast, emphasize social comparisons with other students and argue that high performing students in the classroom might actually deterio-

rate the motivation of other students who see themselves as inferior.

## **2.3 Student Mobility, Social Integration and the Influence of Peers**

Peer socialization as the most prominent account in the literature operates through social dynamics between students and crucially relies on the integration of students in the peer group. Students who recently transferred to a school, however, lack this social integration and the friendship networks that are at the core of peer influence. Based on this reasoning, we argue that student mobility and the dynamics of exposure to different school contexts play an important role for peer effects especially considering that student mobility is pervasive across urban school districts in the the U.S. To evaluate this argument, we first review the literature on the effect of student mobility with a particular focus on the consequences for the social integration and friendship networks of mobile students and then turn to the implications of changing school for the influence of peers on educational outcomes.

### **2.3.1 The Effect of Student Mobility**

Recent studies have documented significant effects of residential and school mobility on a number of outcomes including academic performance (Grigg 2012), high school dropout (Rumberger and Larson 1998; South et al. 2007), delinquent behavior and violence (Haynie and South 2005; Hoffmann and Johnson 1998). But some authors also challenge this finding reporting non-significant or even positive effects of student mobility (Alexander et al. 1996) pointing at fundamentally different reasons for changing school. While a large proportion of moves are driven by divorce, job loss or similar family events, other



school changes are motivated by the desire to attend higher quality schools or resolve conflicts at the current school. Using data from a survey of more than 13,000 students in Chicago, Kerbow et al (2003b) report that about 58% of school changes were associated with residential mobility and 42% of students listed school-related concerns alone as reasons for their move such as school quality or safety issues. Despite these different motivations and the importance of school choice and market-driven school reform in many policy debates (Nechyba 2000), few studies focus on the change in school quality following a move. Instead, most research conflates the effect of mobility itself with context effects related to changes in school quality (for some exceptions see Hanushek et al. 2004; Cullen et al. 2005, 2006). As part of our focus on the role of student mobility for the influence of peers, we separate mobility from context effects using an innovative design and thereby also contribute to the literature on student mobility, which largely ignores the change in school quality after students change school.

Across the literature on student mobility, a common explanation of mobility effects is that students who move to a new neighborhood or change school are not as well socially integrated and accepted among their peers. Early formulations of this argument go back to Robert E. Park's "Human Migration and the Marginal Man" (Park 1928), in which he described migrants as outsiders that are isolated both from their origin and receiving community. While this perspective on migrants has fundamentally changed, research continues to emphasize the negative effect of residential and school mobility on social relations and the integration in peer groups (Hagan et al. 1996). A number of recent studies describe this social adjustment process. Clampet-Lundquist et al (2011), for example, use data from 86 in-depth interviews with teens to explore the gender differences in the effect of moving to a low-poverty neighborhood documented in the Moving to Opportunity (MTO) study. Using an experimental de-

sign, MTO revealed that the benefits of switching from low-income to middle-income neighborhoods were limited to girls (Kling et al. 2005). To address this puzzling finding, Clampet-Lundquist et al examine how boys and girls adjust differently to the new context. Their findings show that boys who moved to better neighborhoods were not accepted in higher status peer groups and ended up gravitating to the more delinquent kids at the margins of the network, while the girls made a more successful transition that allowed them to take advantage of the available resources. Building on Cartner's work (2003) about the dominant and "non-dominant" cultural capital of low-income African-American students in high school, the authors interpret their findings in terms of cultural conflict and argue that boys from low-income neighborhoods imported non-dominant attitudes and behavior into lower-poverty neighborhoods, which provoked negative reactions. Using quantitative data from friendship nominations in school, South and Haynie (2004), Lubbers et al (2011) and Vernberg (1990) confirm the general pattern that students who recently moved to a school have smaller networks and their position is less central and prestigious for several years after the school change. Contrary to Clampet-Lundquist et al's findings, South and Haynie (2004) report "some slight indication" that the impact of school mobility on friendship networks is larger for girls than boys with a small but statistically significant difference for two out of nine network related outcome measures. Lubbers et al (2011), however, do not confirm this finding and show that girls and boys experience a similar reduction in the number of friends, and Vernberg (1990) reports that boys who moved experienced more frequent rejection. Clampet-Lundquist et al and Vernberg's conclusion about gender differences is in line with common findings about the network structure and group processes among boys and girls. Research shows that boys tend to enforce boundaries more strongly and are embedded in more hierarchically structured networks (Thorne 1993). As a consequence, rejecting new

students might be more common among boys so that boys who change school might have more problems than girls to integrate into the peer groups at the new school.

Overall, student mobility plays an important role for friendship relations and the acceptance among peers. Students who move to a new neighborhood or change school tend to be less integrated in the peer networks, socialize with other students that are at the margins of the network, and are not as socially accepted as students who have been at the school for a longer time. Previous research also indicates that this social adjustment process is more pronounced among boys.

### **2.3.2 Peer Influence and Student Mobility**

The effect of student mobility on the social integration in peer networks documented by Clampet-Lundquist et al, South and Haynie and others has important implications for the influence of peers. Many of the mechanisms through which peers affect educational outcomes operate through social dynamics between students and crucially rely on the integration of students in the peer culture. Students that are part of densely connected social groups with close friends are more sensitive to peer influence for various reasons. First, densely connected peer groups and close friends imply a higher level of exposure in terms of interactions and communication so that students within these peer groups are more likely to be influenced by their peers. Second, closely knit peer groups can develop social identities when their members' self-concept is connected to the membership in the social group, which increases the likelihood that students adopt the norms that are central to their social group (Terry, Hogg, and White 2000). Accordingly, students are more susceptible to the influence of peers when they are embedded in closely knit friendship networks that have a social identity function. And finally related to the second point, peer

groups with a dense network structure and a social identity based on group membership are more likely to exert peer pressure and enforce group boundaries so that members of such groups tend to adhere more closely to group norms (Brechwald and Prinstein 2011; Haynie 2001).

Along these lines, previous research by a number of social psychologists generally supports this argument. Urberg et al (2003), for example, show that adolescents who are accepted in their peer group and have close friends are more likely to adapt their friend's smoking behavior (for similar findings on depression and pro-social behavior see Prinstein 2007; Barry and Wentzel 2006). Other studies go beyond the focus on dyadic friendship characteristics and instead look at structural properties of friendship networks. Haynie (2001), for example, examines how the popularity and centrality of students as well as the density of their peer group is related to the association between their own and their friends' delinquency. Her findings show that popular student in central network positions and particularly within dense friendship networks are more susceptible to the influence of peers. Accordingly, friendship relations and the structural position of students in the friendship network have important implications for the influence of peers.

In summary, student mobility alters the nature of exposure effects because many of the mechanisms that explain school and peer effects focus on social integration as essential for benefiting from higher performing peers or a learning orientation in the peer culture. Mobile students, however, are not as well integrated in the friendship network at the new school (fewer friends, less central and less prestigious position) and as a consequence are less susceptible to the influence of peers. This social adaptation process seems to be particularly pronounced among mobile boys who are not accepted in higher status peer groups at the new school and gravitate to other students at the margins of the friendship network. Girls, in contrast, make a more successful transition that

allows them to take advantage of the available resources. Accordingly, exposure dynamics related to the frequent school transfers of students are essential for our understanding of peer and more generally context effect. Based on our argument, we expect that student mobility temporarily reduces peer effects particularly for boys and to a smaller extent for girls. Accordingly, this temporal adjustment process initially dampens the benefits of transferring to a school with higher achieving peers but over the years students begin to experience the positive effect that is commonly associated with the improved environment (higher performing peers).

## 2.4 Data and Methods

We evaluate our argument using a quasi-experimental research design based on an extensive longitudinal database from the Chicago Public School (CPS) system assembled by the Consortium on Chicago School Research (CCSR) at the University of Chicago. With more than 400,000 enrolled students and about 600 public elementary and high schools, CPS is currently the third largest school district in the US. The database consists of the administrative student records for all CPS students in grade 1 through 8 from the school year 1993/94 to 2005/06. These records include the school and grade identifier for the fall and spring term, a limited number of standard demographic characteristics such as gender, date of birth, students' race, their eligibility for free lunch as a measure of parental background, and their special education status. The database also includes the test file records, which contain different reading and math tests administered over the years. Most noticeable is the Iowa Test of Basic Skills (ITBS) in reading and math, which was taken by almost all students in the spring of grade 3 through 8.

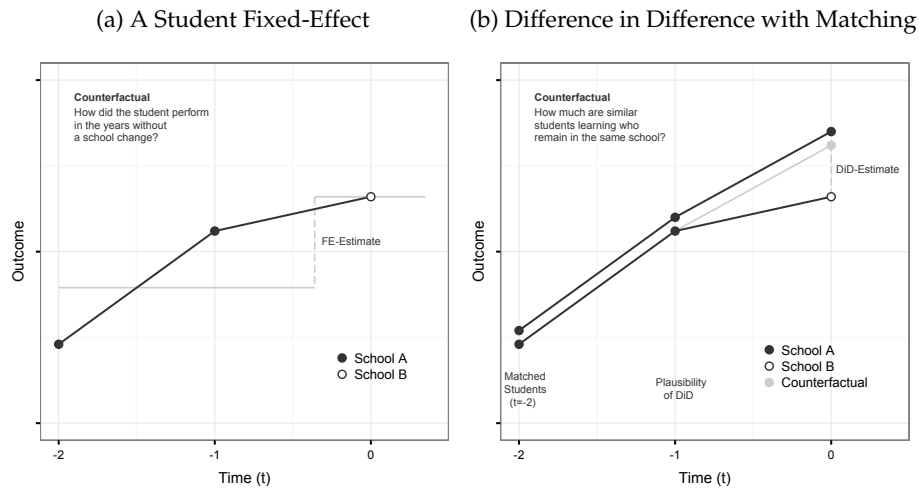
The administrative student records from CPS allow us to track students

when they transfer between public schools within the Chicago school district so that we can follow the development of students not only when they remain in the same school and are exposed to certain peers but also when they change school. This comprehensive nature of the dataset provides the basis for a quasi-experimental research design based on a difference-in-difference, matching approach. In particular, we compare the test-score growth of students who transfer from school  $a$  to  $b$  (treatment) with the test score growth of matched students in both schools in the years before and after the transfer. The core assumption of this design is that in the absence of the treatment (i.e. without school transfer) the growth in test-scores for the two groups would have been the same – an assumption that can be evaluated almost directly. Using this approach, we first estimate the overall effect of changing school over several years after the transfer subsuming both the effect of mobility itself and the context effect related to the new environment. In the second step of our analysis, we carefully separate these two components and examine how student mobility alters peer effects.

#### **2.4.1 Difference in Difference Estimation Strategy**

To evaluate our theoretical argument, we follow students who change to schools with lower/higher peer ability in the years after the transfer and compare them to non-mobile students who remain in the same school. A simple comparison of students who recently transferred with those who did not is problematic because the two groups of students might be different in other ways that are related to their performance. Even a comprehensive set of control variables is unlikely to solve this problem as long as unobserved factors are related both to the treatment and the outcome. An alternative approach are fixed-effect models, which use within student variation to estimate the causal effect and commonly also condition on some time specific effect. Such a within-student, before-after comparison is common in the literature on student mobility and essentially

Figure 2.1: Illustration of Alternative Estimation Strategies



compares a student's test score growth to the average growth of other students in the same grade in the same year. As discussed by Grigg (2012), such a student and year-by-grade fixed effect model is a powerful way to rule out many unobserved confounding factors. The approach, however, ignores contextual factors as an important determinant of test-score growth. The year-by-grade fixed effects assume that students who change school would have learned as much as *all* non-mobile students regardless of the school they attend, which might introduce bias because of an implausible comparison group.

To circumvent this problem, we use a difference-in-difference approach that is based on the comparison of students who change school to non-mobile students who remain *in the same school*. As illustrated in Figure 2.1a, the core assumption of this design is that in the absence of the treatment (i.e. without school transfer) the test-score growth for students who change from school *a* to *b* would have been the same as the actually observed growth for students who remain in school *a*. In other words, a mobile student would have learned as much as his/her non-mobile classmates if s/he had remained in the same school. This approach overcomes the limitations of the fixed-effect model dis-

cussed above. It adjust for time-specific factors that are constant across individuals from the same school and clearly defines the counterfactual as what a student would have learned if s/he had remained in the same school. In addition, the plausibility of this difference-in-difference approach can be evaluated almost directly by comparing the test score growth of treatment and control units before the actual transfer. As illustrated in Figure 2.1a, control students on average might have higher growth in test scores even in the years before the actual transfer, which questions the plausibility of the common trend assumptions. To address this problem, we match students who change from school *a* to *b* to similar students in school *a* based on characteristics that are observed before the mobile student actually change school such as parental background, gender, race and test-scores.

**Sample construction and treatment indicator** To implement our difference-in-difference approach, we construct a sample of students that allows us to study the dynamics of exposure effects, estimate the relevant effect over several years after the school change and examine the sensitivity of our results. For this purpose, we focus our analyses on *Chicago students* - i.e. students who begin school in a Chicago public school and remain enrolled in the same school from 1st to 3rd grade. This restriction excludes students who transfer into and out of the Chicago public school system so that we are unable to fully observe their exposure dynamics. The restriction to students who remain in the same school between 1st and 3rd grade is necessary because the first test is administered to students in 3rd grade so that we are unable to estimate the effect of school transfers before that. Our final sample includes 187,408 students and 760,097 student-year observations with valid test-score information (on average, 4.1 observations per student).

Based on this sample, our control group includes all “stable” students who



never change school and our treatment group all student who change school at least once after 3rd grade. Accordingly, our analysis compares mobile students to students who remain in the same school for a longer period. To estimate the relevant effect over several years, we use different treatment indicators  $D_{it}^{(1)}$ ,  $D_{it}^{(2)}$ ,  $D_{it}^{(3)}$ , and  $D_{it}^{(4)}$  for student  $i$  in year  $t$  to indicate the timing of the transfer.  $D_{it}^{(1)}$  captures transfers that occurred over the last year so that it is coded as one for student who changed school from the spring term at  $t - 1$  to  $t$ . The additional indicators refer to the years since the last transfer so that  $D_{it}^{(2)}$  indicates that a student changed school between  $t - 2$  and  $t - 1$  but remained in the same school from  $t - 1$  to  $t$ .

This sample together with the series of treatment indicators allow us to estimate the effect of student mobility for several years after the transfer and observe how students adjust to the new environment. It also allows us to conduct sensitivity analysis by comparing the performance growth of students in the treatment and control group before the actual transfer occurs.

**Model specification** The specification of difference-in-difference models is based on two fixed effect terms – one for the time-invariant group effect and one for the time-specific effect that is constant across all groups – so that in the absence of the treatment, the outcome is determined by the sum of these two effects. To estimate the treatment effect, difference-in-difference models add a dummy variable for group-by-time observations that are in the treatment group. For the panel data at hand, these models can be specified as

$$y_{isgt} = \alpha_i + \mu_{s*gt} + \delta_1 D_{it}^{(1)} + \delta_2 D_{it}^{(2)} + \delta_3 D_{it}^{(3)} + \delta_4 D_{it}^{(4)} + \varepsilon_{isgt} \quad (2.1)$$

Here,  $i$ ,  $s$ ,  $g$ , and  $t$  are the indexes for student, school, grade, and year respectively.  $\alpha_i$  is the time-constant, student-specific effect, which captures all observed and unobserved factors on the individual level that are constant across

time such as family background or prior performance.  $\mu_{s^*gt}$  is an original school-by-grade-by-year fixed effect term and captures all grade-by-year specific factors that are constant across all students from the same school. Note that the star for the index  $s$  indicates that the term  $\mu_{s^*gt}$  does not refer to the current but to the original school so that the performance growth of treatment student is compared to the growth of students who remain in the same school. Together, these two fixed effect terms model the performance of students as the sum of an individual-specific component and a school-by-grade-by-year component.  $D_{it}^{(x)}$  are dummy variables for the treatment indicator described above. They indicate whether a student changed school for the four years after the transfer so that the corresponding coefficients  $\delta_{(x)}$  estimate the difference in performance growth for mobile students compared to students in their original school in the years after the transfer.

**Plausibility of Estimation Strategy and Matching Procedure** The causal interpretation of these effects crucially relies on the common trend assumption, which implies that in the absence of the treatment (i.e. without school transfer) the test-score growth of mobile students would have been as large as the growth of non-mobile students in the same school. In order to evaluate this assumption, we compare the test-score growth before mobile students actually change school ( $t - 2$  to  $t - 1$  in the simplified set-up in Figure 2.1a). As documented in Appendix B, the findings from this sensitivity analysis show that mobile student *do not* learn at the same rate as their non-mobile peers from the same school. Accordingly, all non-mobile students are not a good comparison group and the estimates based on the raw sample are most likely biased. To circumvent this problem, we match mobile students to similar students in their original school using a recently developed matching procedure called Mixed Integer Programming for Matching (MipMatch). This approach

automatically balances multiple criteria of the empirical distribution of the observed covariates (Zubizarreta 2012). The details about the matching procedure are described in Appendix A. The matched sample substantially increases the balance between the control and treatment group. Most importantly, the sensitivity analysis for the matched sample shows that matched students are an excellent comparison group (Appendix B). The test-score growth of mobile students (treatment group) is the same as the growth for similar – i.e. matched – students in the same school before they actually change school. This finding provides strong support for the common trend assumption and as such makes a compelling case for the validity of our estimation strategy.

**Separating Mobility and Exposure Effects** The effect of changing school estimated with the model described in the last section conflates both the effect of mobility itself and the context effect related to the environment at the new school. In the second step of our analysis, we separate these two effects by adding additional fixed-effect terms for the current school. Formally, the extended models can be expressed as

$$y_{isgt} = \alpha_i + \mu_{s^*gt} + \gamma_{sg} + \lambda_{st} + \eta_{gt} + \delta_1^* D_{it}^{(1)} + \dots + \delta_4^* D_{it}^{(4)} + \varepsilon_{isgt} \quad (2.2)$$

Here,  $\gamma_{sg}$ ,  $\lambda_{st}$ , and  $\eta_{gt}$  are fixed effect terms for school-by-grade, school-by-year, and grade-by-year factors that all refer to the current school and not the original school as the term  $\mu_{s^*gt}$ . In these modified models,  $\delta_x^*$  does not estimate the *overall* effect of changing school, which includes both the mobility and context effects, but isolates the effect of student mobility from any context effects related to the new school. Accordingly, mobile student are compared to non-mobile students *both* from their current *and* their previous school so that

our estimates now reflect the effect of mobility above and beyond any contextual factors related to the current and previous context. Although the actual implementation is completely different, this understanding of mobility effects resembles the underlying idea of Sobel’s diagonal mobility models (Sobel 1981, 1985) insofar as both approaches define mobility effects as distinct from effects relates to the previous and current context.

In addition to isolating the effect of student mobility, these extended models allow us to thoroughly estimate the effect of peer performance and its interaction with student mobility so that we can evaluate how students who recently transferred to a school are affected differently by their peers. These additional variables are integrated in the model described in formula 2.2 as  $\bar{y}_{sgt}^{(-i)}$ , which is the average performance in the previous year of all student in a certain school, grade and year except individual  $i$  (indicated by the superscript  $-i$ ), and with a number of interaction terms between peer ability and the treatment indicators  $D_{it}^{(x)} \times \bar{y}_{sgt}^{(-i)}$ . The additional fixed-effect terms for multiple combinations of (current) school, grade, and year ensure that the estimation of the peer effects uses cohort-to-cohort variations within schools, which has been described as a “large number of quasi-experiments” for each of the available schools by Hanushek and Rivkin (2009, 379; also see Hoxby 2000). Appendix C contains a more detailed description of this approach and also discusses potential problems with the estimation strategy.

Together, these models evaluate our argument about the temporal adjustment process of mobile students and its influence on peer effects over several years after the transfer using an innovative quasi-experimental research design based on a difference-in-difference, matching approach. Based on this design, we first estimate the overall effect of changing school (mobility and exposure effect) for different types of transfers (up vs down mobility in terms of change in peer ability). In the second step of our analysis, we separate the effect of stu-

dent mobility from exposure effects and examine both the direct mobility effect and how it interacts with peer effects.

## 2.4.2 Variables and Missing Data

Our analyses use reading test scores from the Iowa Test of Basic Skills (ITBS) as the main outcome variable. The ITBS was designed by Riverside Publishing and was given to all students in grade 3 through 8. 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.

The focal independent variables are (a) dummies for the first, second, third, and fourth year after a student changes school and (b) the ability of peers. The school changes are all non-promotional and include both changes within and between school years.<sup>2</sup> Peer ability is defined as the average performance in terms of test-scores of all students in a certain school, grade and year – i.e. all student in a certain cohort.

In the first step of the analysis, I estimate the effect of school changes separately for various levels of change in school quality that reflect different types of mobility. In particular, the change in school quality associated with the school transfer is based on peer achievement and categorized in three groups, namely below -0.25, -0.25 to 0.25, and above 0.25. With these categories, a change between +/- 0.25 standard deviations in peer performance is defined as a transfer to a similar school and the other two categories as a transfer to a better/worse school in terms of peer ability. These three categories reflect important differ-

---

<sup>2</sup>Promotional school changes or changes associated with the opening of new schools are excluded from the analysis so that all school changes are non-promotional. Because our data does not include information about the grade range of each school (in general, Chicago elementary schools include grade 1 through 8), we exclude cohorts (school-year-grade) for which over 50% of students change school in a particular year or enroll in a new schools. Overall, this restriction excludes less than 2% of cases. Separate analysis confirm earlier findings showing that school changes during the school year have larger negative effects (available from the authors) but the overall pattern of findings is the same across the different types of school changes.

ences in school mobility and peer performance that allow us to examine how school changes and peer ability jointly influence test-score growth. Importantly, the effect estimates in this first step absorb *both* the effect of mobility itself *and* the exposure effect connected to the new environment. In the second step of our analysis, we separate these two components and estimate the effect of peer ability directly and how it interacts with student mobility.

Table 2.1: Sample Characteristics

Variable	Mean	St. Dev.	Mean (Male)	Mean (Female)
Female	0.51	0.50		
Free/Reduced lunch	0.78	0.41	0.78	0.79
<i>Race</i>				
Black	0.49	0.50	0.48	0.50
White	0.10	0.30	0.10	0.10
Hispanic	0.37	0.48	0.38	0.36
Asian	0.03	0.17	0.03	0.03
<i>Student Mobility</i>				
Number of transfers	0.51	0.77	0.52	0.50
At least one transfer	0.39	0.49	0.39	0.38

$N = 187,408$

In addition, the dataset includes a number of useful control variables such as race, gender, and family background but these variables are all captured with the student fixed effect term. Table 2.1 presents descriptive statistics for some of these student characteristics.

## 2.5 Results

### 2.5.1 Patterns of Student Mobility

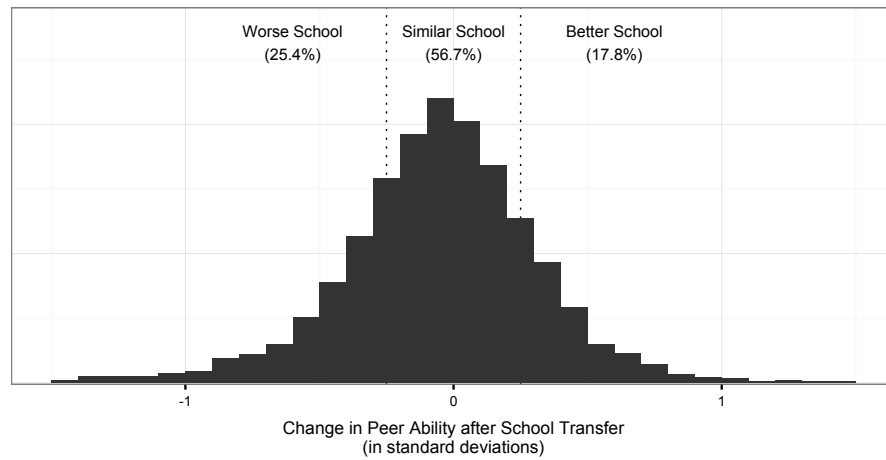
Non-promotional student mobility is pervasive in Chicago public schools. On average, students in our sample change school 0.5 times between 3rd and 8th grade. 39.4% of students transfer at least once with large variations by ethnicity and family background (for a more detailed report on student mobility in

Chicago see Kerbow 1996; Torre and Gwynne 2009). This high rate of student mobility is comparable to other urban school districts and slightly larger compared to less urban settings. In Texas, for example, about one third of students move at least once over a three year period (Hanushek et al. 2004).

In terms of change in school quality, most students transfer to schools with similar or slightly lower performing peers. Figure 2.2 shows the standardized change in peer ability after students transfer from one school to another within the Chicago school district. A one unit change indicates that a student moves to a school in which peer ability, on average, is one standard deviation higher compared to the former school. The center of the distribution is slightly below zero and over 82% of students change to similar or worse schools. On the other hand, most neighborhood and school transfers induced by policy initiatives such Moving to Opportunity or the many voucher programs in school districts across the country modestly and sometimes significantly improve the environment in terms of poverty rate or other characteristics. Accordingly, the pervasive student mobility in urban school districts mainly occurs between similar or worse schools whereas many policy interventions try to modestly or significantly improvement the school environment for students from disadvantaged backgrounds.

Previous research on student mobility almost exclusively conflates this change in school quality with the effect of student mobility itself (for an exception see Hanushek et al. 2004). In the following analysis, we first estimate the overall effect of changing school for the three categories of transfers indicated by the dotted lines in Figure 2.2 and over several years after students change school. In the second step, we isolate the effect of student mobility from context effects and examine how student mobility alters the influence of peers on educational outcomes. Together, these two steps of our analysis allow us to examine the temporal adjustment process of mobile students and highlight the importance

Figure 2.2: Change in School Quality in terms of Peer Ability for mobile Students across Chicago Public schools



*Note: Distribution of change in peer ability as a measure of school quality for 89,380 non-promotional school changes between Chicago Public schools.*

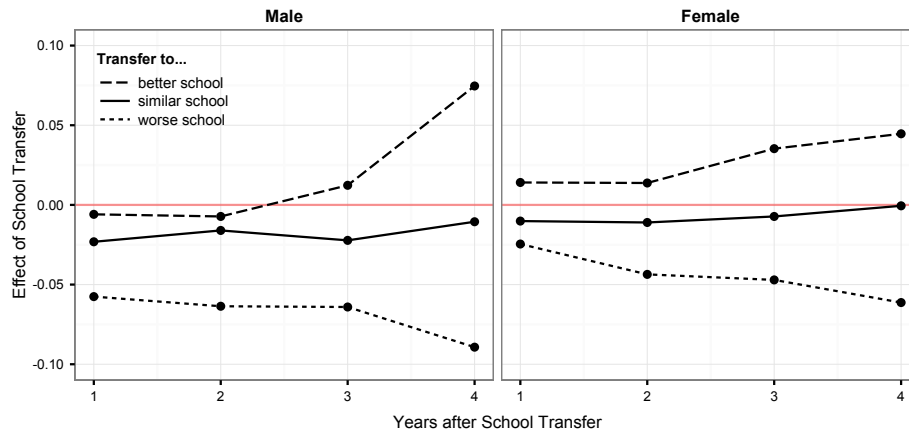
of student mobility and exposure dynamics for our understanding of peer effects.

## 2.5.2 The Effect of Changing School

We begin our analysis with a set of regression models based on a difference-in-difference design that estimates the overall effect of changing school on test-scores by different types of transfers and by gender. These models compare the performance growth of students who change school to the growth of students who remain in the same school conflating the effect of mobility itself with exposure effects related to the new environment. The point estimates of the standardized effect size from the matched sample for student mobility in the years after the transfer are reported in Figure 2.3. With over 250,000 observations for each of the three regression models (one for each type of transfer), the sample size is large and even effects as small as 0.02 standard deviations are highly significant. Accordingly, our interpretation focuses on substantive difference in



Figure 2.3: Effect of Student Mobility on Test-Scores by Type of Transfer and Gender



effect size.

For boys, the results show a clear temporal adjustment process for students who transfer to schools with higher performing peers. In the first three years after a transfer, the effect size is small but gradually increases and finally shows a substantial positive effect in the fourth year after changing to a higher quality school. Accordingly, boys who transfer to better schools with higher peer ability initially perform at the same level as their matched peers in the previous school but begin to experience a positive effect after several years in the new school. For girls, we observe a positive effect even in the first year after the transfer. This effect, however, increases at a slower rate compared to boys and reaches a lower level after four years at the higher quality school. Accordingly, girls initially benefit more strongly from such a transfer but after several years at the new school this pattern reverses showing a larger benefit for boys. This finding supports our theoretical argument about gender differences in the adjustment process indicating that boys have more problems adjusting to the new school. For transfers to similar or worse schools, we observe a similar but less

pronounced temporal adjustment process. Both boys and girls who change to similar schools initially experience a small negative effect that slightly declines over time reaching zero after four years in the new school and summing up to a modest negative effect over the years. For transfers to worse schools, the pattern indicates a negative effect right in the first year that further declines over time with overall less strong effects for girls compared to boys.

These findings with a pronounced temporal adjustment process for boys who change to higher quality schools and a present but less distinct temporal pattern for other transfers are highly informative about the overall effect of changing school. They directly speak to critical policy debates about school choice and student mobility. But they also conflate mobility with context effects potentially hiding important temporal patterns that speak to our theoretical argument about the role of student mobility for the influence of peers on educational performance.

### **2.5.3 Student Mobility, Exposure Dynamics and the Effect of Peers**

The results for the effect of changing school show a pronounced temporal adjustment process for boys but not for girls with large variations across the different types of transfers. These findings confirm our argument about the temporal adjustment process but our analysis so far focuses on the overall effect of changing school subsuming mobility and exposure effects. To separate the two components and study how mobility alters exposure effects in the years after the transfer, we extend the models used in the last section with a number of fixed-effect terms to thoroughly estimate peer effects. As described in the Data & Methods section, these extended models capture the effect of mobility above and beyond any contextual factors related to both the current and previ-

ous context and allow us to examine the interaction between student mobility and peer effects.

Model I in Table 2.2 first shows the isolated effect of student mobility above and beyond exposure effects related to the current or previous context. The findings show a very small but statistically significant effect of mobility itself of -0.01 standard deviations for boys and girls in the first year after the transfer that gradually declines to zero over the subsequent years. These estimates are smaller compared to Grigg's (2012) recent study, who reports that changing school lowers achievement growth by about 6 percent (our comparable estimate are 1.5 percent). This difference is probably related to the fact that we isolate mobility effects accounting for factors related to both the current and previous context. Given that students on average transfer to slightly worse schools, Grigg's and other estimates in the literature pick-up part of the effect related to the lower performance of peers in the new school indicating that it is important to carefully distinguish context from mobility effects. Despite these very small effects, student mobility can have consequences. First, student mobility itself influences performance in the years after the transfer, which adds up to cumulative effect of 0.03 standard deviation. Second, many students transfer multiple times so that the effect accumulates, which is problematic considering that students from disadvantaged backgrounds transfer most frequently (particularly, low SES and black students). Importantly, the effect of mobility itself appears to be the same for boys and girls with only small and insignificant differences in the estimates after the second digit.

Model II in Table 2.2 continues by showing the effect of peer ability on reading test-scores alone. The theoretical mechanisms discussed above as well as previous evidence from observational, quasi-experimental, and experimental studies suggest that peer ability has a positive effect on performance (Sacerdote 2010; Epple and Romano 2011) but some recent studies also argue that

Table 2.2: Effect of Peer Ability and Student Mobility

	Model II		Model II		Model III					
	Male	Female	Male	Female	Male	Female				
	Coef	(se)	Coef	(se)	Coef	(se)				
<i>Mobility Effects</i>										
Mobility (Year 1)	-0.01**	(0.00)	-0.01**	(0.00)	-0.03***	(0.01)	-0.03***	(0.01)		
Mobility (Year 2)	-0.01*	(0.00)	-0.01*	(0.00)	-0.02***	(0.01)	-0.02***	(0.01)		
Mobility (Year 3)	-0.01	(0.00)	-0.01	(0.00)	-0.01*	(0.01)	-0.01*	(0.01)		
Mobility (Year 4)	-0.00	(0.00)	-0.00	(0.00)	-0.01*	(0.01)	-0.00	(0.01)		
<i>Peer Effects</i>										
Peer Ability (PA)			0.12***	(0.00)	0.11***	(0.00)	0.15***	(0.00)	0.11***	(0.00)
<i>Interaction Terms</i>										
PA x Mobility (Year 1)					-0.06***	(0.01)	-0.02**	(0.01)		
PA x Mobility (Year 2)					-0.05***	(0.01)	-0.01*	(0.01)		
PA x Mobility (Year 3)					-0.03***	(0.01)	-0.01	(0.01)		
PA x Mobility (Year 4)					-0.02*	(0.01)	-0.00	(0.01)		

N=392,576; \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ ; standard errors in parenthesis

Note: All models includes student, grade-year, school-by-grade, school-by-year, and original school-by-grade-year fixed-effect terms.

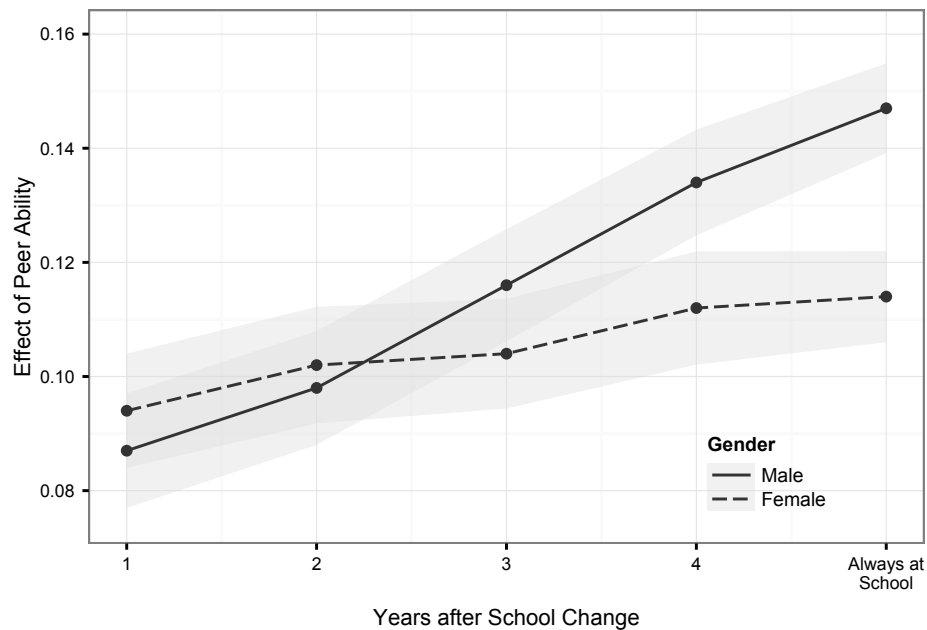
peer effects are driven by selection processes (Lauen and Gaddis 2012). The findings from our own models indicate that peer ability has a clear effect on test-scores of 0.12 standard deviations for boys and 0.11 for girls. The size of these effects is modest and comparable to some of the other most reliable estimates in the literature. This finding reaffirms previous studies and indicates that peers in school play an important role for educational performance. They also show relative small gender difference in peer effects.

Finally, Model III in Table 2.2 adds interaction terms between student mobility and peer performance. These models allow us to examine our main argument about the role of school changes for peer effects. The main effect of peer ability now refers to students *who have always been* at their respective school and the estimates for the interaction indicate the extend to which peer effects differ for students who recently transferred to a school. The findings show that the effect of peers for students who have always been at a school (main effect) is larger compared to the estimates for all students reported before and we now observe a clear gender difference in the size of the effect. Accordingly, peer effects are larger for boys than girls among student who did not recently change school.

The estimates for the interaction terms are negative and highly significant. They indicate that student mobility substantially alters peer effects after students change school. In the first year after the transfer, the effect of peer ability is reduced by about 40% for boys and nearly 20% for girls. In the subsequent years, this reduction gradually declines for boys and girls but the temporal pattern is less pronounced for girls. Figure 2.4 illustrates this finding and shows the size of peer effects together with the 95% confidence interval over several years after students change school. The figure again highlights the significant gender differences in the temporal adjustment process.

Overall, these findings indicate that student mobility itself has a very small

Figure 2.4: Effect of Peer Ability by Years since Student Changed to School



direct effect on the performance of students but instead reduces the influence of peers in important ways. They support our theoretical argument about the temporal adjustment process for boys but not for girls, and indicate that student mobility plays a critical role for peer effects considering that students across the U.S. frequently change school.

## 2.6 Conclusion

Neighborhood, school and peer effects have been an important element of contemporary stratification research. At the core of this agenda is the argument that exposure to high poverty neighborhoods places teens at risk and peers in school shape educational outcomes. Yet, previous research on neighborhood, school, and peer effects largely ignores that students frequently change school

and families move despite the fact that most theories of peer influence attribute a critical role to the social integration of students and their peer networks. At the same time, the empirical evidence for the success of policy interventions that aim to place families in low-poverty neighborhoods or move students to better schools is mixed.

To address this question, this article integrates the literatures on student mobility and peer effects focusing on the temporal adjustment process of students who change school. This adjustment process implies that student mobility not only has a temporary negative effect on test-score growth itself but also alters context effects in the years after a student moves. Results from our analysis of a large-scale administrative dataset and a quasi-experimental research design show that the effect of peers is substantially smaller for mobile students with a clear temporal adjustment process that is more pronounced for boys. Boys who transfer to a school with higher achieving peers initially learn at the same rate as their peers in the previous school despite the improved learning environment. Only after several years, they begin to experience the benefits associated with higher performing peers. For girls, this temporal adjustment process is less pronounced with a positive effect right after transferring to a higher quality school (in terms of peer performance) but only a small increase in this positive effect over time. A careful separation of mobility and peer effects shows that student mobility itself has only a marginal effect on performance but changes the ways in which students are influenced by their peers in important ways. In particular, peer effects are reduced by as much as 40% for boys who transferred to a school over the last year and only gradually begin to experience the same peer effects as students who have always been at a certain school. The change in peer effects for girls, however, is substantially smaller (about 18%) and the adjustment process much quicker.

These findings indicate that student mobility itself has a very small direct

effect on the performance of students but instead plays a critical role for the influence of peers and as such for neighborhood and school effects more broadly. They also contribute to several research areas. First, the findings broaden our understanding of peer effects by highlighting the ways in which student mobility alters exposure effects. Given that student mobility is pervasive across urban school districts in the U.S., the finding that the influence of peers on education performance is different for students who recently transferred to a school has important implications for the peer effects literature. It builds on and extends the recent literature on exposure duration (Wodtke et al. 2011; Sharkey and Elwert 2011; Crowder and South 2011; Jackson and Mare 2007; Sampson et al. 2008). Focusing on neighborhood and not school effects, this literature shows that long-term exposure to disadvantaged neighborhoods has severe consequences that go beyond the effects previously documented in the literature. In contrast to our argument, these studies approach mobility merely as a methodological challenge and measure long-term exposure with a summary measure of concentrated disadvantage disregarding the role of moves between different contexts. Accordingly, our argument extends the literature on the temporal dimension of exposure effects by emphasize mobility as an important aspect of this process.

Second, understanding the temporal adjustment process helps us to reconcile the seeming contradiction between the literature on neighborhood and school effects on the one hand and the mixed evidence for the effectiveness of certain policy interventions on the other hand. Over the last decades, neighborhood and school effects have been a central research topic across the social sciences. Yet, the empirical evidence for the success of policy interventions that aim to place families in low-poverty neighborhoods or move students to better schools is mixed. Most prominently, the Moving to Opportunity (MTO) experiment, which randomly gave low-income families the opportunity to move to



low-poverty neighborhoods, provoked a debate about the lack of neighborhood effects for many of the outcome measures (Kling et al. 2007; Sampson 2008) and school voucher programs similarly show mixed evidence for positive effects (Angrist and Lang 2004; Cullen et al. 2005, 2006). Our research suggests that these surprising findings can be explained by the temporal adjustment process of student mobility, which implies both a negative effect itself and temporarily altered exposure effects so that students only begin to benefit from an improved environment several years after the school change. Accordingly, policy interventions that offer vouchers to move student to better schools juxtapose mobility and exposure effects, which partly undermines the purpose of the intervention. A further complication is the fact that many students do not remain in a new school for an extended period so that mobile students might never experience the benefits of higher achieving peers or other resources in the improved environment. From this perspective, context effects play an important role in shaping educational outcomes but moving students across school districts might still not provide a tangible solution. If the consequences of moving can not be alleviated, an alternative would be to focus resources on improving neighborhood schools that provide benefits for all students.

Finally, our findings speak to the recent debate on gender differences in context and policy effects (Kling et al. 2005; Clampet-Lundquist et al. 2011; Legewie and DiPrete 2012). A number of studies indicate that girls but not boys benefit from moving to a higher-resource environment (Hastings et al. 2006; Clampet-Lundquist et al. 2011; Kling et al. 2005) while other research suggests that boys are more sensitive to peers in their context (Legewie and DiPrete 2012; Carrell and Hoekstra 2010). Our findings contribute to this debate by showing that the temporal adjustment process is particularly pronounced for boys. Accordingly, boys might well be more sensitive to peer effects but after changing school they have more problems adjusting to the new environment, which temporary re-

duces the effect of peers and explains the contradictory findings in previous research.

## Appendix A Matching procedure

As documented in Appendix B, mobile students on average have lower test-score growth compared to their non-mobile peers in the same school even before they actually transfer. This finding challenges the common trend assumption and suggests that *all* students who remain in the same school are not a good comparison group. To address this problem, we match mobile students (treatment group) to similar, non-mobile students (control group) in the same school so that our estimates compare the test-score growth of mobile students to similar – i.e. matched – students from their original school. Given the sample used in our analyses, we match students based on their 3rd grade characteristics so that we can still evaluate the common trend assumption based on test-score growth before the actual transfer.<sup>3</sup>

The matching is based on a recently developed matching procedure called Mixed Integer Programming for Matching (MipMatch), which automatically balances multiple criteria of the empirical distributions of the observed covariates (Zubizarreta 2012). In contrast to the more widely used technique of propensity score matching, MipMatch circumvents the problem of finding the best propensity score model by optimizing predefined criteria such as the differences in univariate moments (means, variances, and skewness), entire distribution characteristics (quantiles or Kolmogorov-Smirnov (K-S) statistic) or features of joint distributions (e.g. multivariate moments). This optimization procedure based on mixed integer programming maximizes bias reduction on the observed covariates and can also impose constraints for exact and near-exact matching of selected covariates. As a further step to optimize balance, we

---

<sup>3</sup>We conducted the same analysis by matching students based on all information that is available before treatment students actually transfer (particularly, all pre-treatment test-scores). This matching based on an extended set of pre-treatment variables might improve the balance between the treatment and control group and therefore the common trend assumption but it does not allow us to conduct the sensitivity analysis described in Appendix B. The findings are almost identical to the ones reported here.

perform the same matching procedure for each school using 1:1, 1:2, 1:3, and 1:4 ratios between control and treatment observations and select the best result for each school.

The actual variables used for the matching are mother's education, free lunch status, and 3rd grade reading and math test scores. For each of these variables, we find matches that optimize balance in terms of the univariate moments as well as the K-S statistic for the continuous variables. In addition, we enforce exact matching for gender, year, and school so that each treatment student is matched to control students with the same gender, similar performance and similar parental background who attended the same school in the same year before our treatment student changed school.<sup>4</sup> The actual set of variables is relatively small compared to other studies but it includes multiple pre-treatment measures of the outcome variable, which is often described as the most important covariate (Shadish et al. 2008a; Steiner et al. 2010a). Our two measures of parental background are also related to both the treatment and the outcome and as such a crucial covariate. More importantly, our estimation strategy uses matching to construct a better comparison group for our difference-in-difference approach and not as part of a design that simply conditions on observable covariates.

The matching is based on a recently developed matching procedure called Mixed Integer Programming for Matching (MipMatch), which automatically balances multiple criteria of the empirical distributions of the observed covariates (Zubizarreta 2012). In contrast to the more widely used technique of propensity score matching, MipMatch circumvents the problem of finding the best propensity score model by optimizing predefined criteria such as the differences in univariate moments (means, variances, and skewness), entire dis-

---

<sup>4</sup>In practical terms, we run our matching procedure separately for each school, which corresponds to exact matching for the original school and circumvents computational problems that arise when the matching is performance on the whole sample.

tribution characteristics (quantiles or Kolmogorov-Smirnov (K-S) statistic) or features of joint distributions (e.g. multivariate moments). This optimization procedure based on mixed integer programming maximizes bias reduction on the observed covariates and can also impose constraints for exact and near-exact matching of selected covariates. As a further step to optimize balance, we perform the same matching procedure for each school using 1:1, 1:2, 1:3, and 1:4 ratios between control and treatment observations and select the best result for each school.

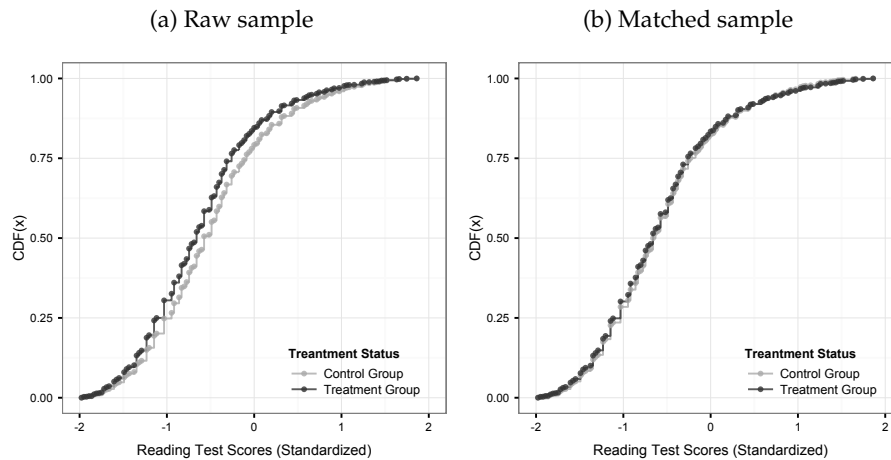
The actual variables used for the matching are mother’s education, free lunch status, and 3rd grade reading and math test scores. For each of these variables, we find matches that optimize balance in terms of the univariate moments as well as the K-S statistic for the continuous variables. In addition, we enforce exact matching for gender, year, and school so that each treatment student is matched to control students with the same gender, similar performance and similar parental background who attended the same school in the same year before our treatment student changed school.<sup>5</sup> The actual set of variables is relatively small compared to other studies but it includes multiple pre-treatment measures of the outcome variable, which is often described as the most important covariate (Shadish et al. 2008a; Steiner et al. 2010a). Our two measures of parental background are also related to both the treatment and the outcome and as such a crucial covariate. More importantly, our estimation strategy uses matching to construct a better comparison group for our difference-in-difference approach and not as part of a design that simply conditions on observable covariates.

Our matching procedure substantially increases the balance between the control and treatment group. Figure A2.5 shows the difference in the cumu-

---

<sup>5</sup>In practical terms, we run our matching procedure separately for each school, which corresponds to exact matching for the original school and circumvents computational problems that arise when the matching is performance on the whole sample.

Figure A2.5: Balance between Treatment and Control Group



lative distribution function for reading test scores of the treatment and control group for the raw and the matched sample. In the raw sample, the test score distribution for students who change school (treatment students) is clearly shifted towards the left indicating that treatment students tend to perform on a lower level compared to students in the control group (the standardized difference in means is 0.12). Similar differences exist for other observed covariates. Overall, this imbalance in the test score distributions in the raw sample is relatively small but consistent across other observed covariates. In the matched sample, this imbalance is reduced substantially and the cumulative distribution function for the control and treatment group closely resemble each other (the standardized difference in means is 0.02). We observe the same pattern for other covariates indicating that the matching procedure clearly reduces the imbalance between the two groups. The results presented in this article are based on the matched sample but similar results with the same substantive findings were obtained with the raw data (the main difference is that the estimates for student mobility are larger in the analyses with the raw data).

Table A2.3: Evaluation of Common Trend Assumption

<i>Type of Transfer</i>	Male				Female			
	Raw		Matched		Raw		Matched	
	Coef	(se)	Coef	(se)	Coef	(se)	Coef	(se)
much better schools	-0.07***	(0.01)	0.00	(0.03)	-0.06*	(0.02)	-0.08**	(0.03)
better school	-0.10***	(0.02)	-0.01	(0.02)	-0.05**	(0.02)	0.01	(0.02)
similar school	-0.09***	(0.02)	0.03	(0.02)	-0.13***	(0.02)	0.02	(0.02)
worse school	-0.03	(0.02)	0.01	(0.03)	-0.02	(0.02)	-0.02	(0.02)
much worse school	-0.06***	(0.01)	-0.05	(0.05)	-0.06*	(0.03)	-0.06*	(0.03)

*Note: Each cell shows the estimate and standard error from a separate model for the difference in test-score growth between mobile and non-mobile students (treatment and control group) before the mobile student actually change school. Sample size varies between 2,000 and 10,000 for the matched sample and between 80,000 and 150,000 for the raw sample.*

## Appendix B Plausibility of estimation strategy

The core assumption of our difference-in-difference approach is the common trend assumption, which implies that in the absence of the treatment (i.e. without school transfer) the test-score growth of mobile students would have been as large as the growth of non-mobile students in the same school. This assumption can be evaluated almost directly for a subset of our sample by comparing the test-score growth before mobile students actually change school ( $t - 2$  to  $t - 1$  in the simplified set-up in Figure 2.1a). The sensitivity analysis excludes students who transfer between 3rd and 4th grade simply because test-scores are observed for the first time in 3rd grade.

Table A2.3 presents estimates for the difference in test-score growth between the treatment and control group before the treated students change school. The models resemble the specification from equation 2.1 discussed in the next section but omit all observations after the transfer. The (placebo) treatment indicator is defined as mobile students in the last year *before* they change school so that the estimates test the common trend assumption. Each cell in the table presents estimates from a separate model showing the difference in test-score

growth between the control and treatment groups for the raw data and the matched sample by type of transfer (up vs. down mobility in terms of change in peer ability) and by gender. For the raw sample, we observe statistically significant and partly sizable differences in test-score growth before treatment students transfer to another school for most of the different types of transfers and both for boys and girls. This finding indicates that treatment and control students in the raw sample are not only different in terms of their level of performance (as documented in Figure A2.5) but also learn at different rates, which challenges the common trend assumption and undermines a difference-in-difference approach as a valid estimation strategy. For the matched sample, however, the differences in test-score growth between the treatment and control group before treated students actually change school are mostly negligible. The differences in performance growth are as small as 0.02 standard deviation (or smaller) across the different types of transfers and for both boys and girls. Accordingly, the test-score growth of mobile students (treatment group) is the same as the growth for similar – i.e. matched – students in the same school before they actually change school. This finding indicates that matched students in the same school are a better comparison group and overall provides strong support for the key assumption of our estimation strategy. The only exception are girls who transfer to either much better or much worse schools. For these extreme transfers with a change of over  $\pm 1$  standard deviations in peer ability, the test-score growth for mobile students is smaller compared to the matched non-mobile peers. These extreme transfers are relatively rare (below 10% of all school changes) and it appears to be difficult to find similar students who attend the same school. Excluding these students from our analyses does not change the results.



## Appendix C Estimation of Peer Effects

The overall effect of changing school estimated with the difference-in-difference approach described in equation 2.1 on page 85 subsumes mobility and context effects. As mentioned in the Data and Methods section, we extend this model so that we can separate the mobility effect from the effect of the current context in terms of peer ability. Using this approach, we are able to isolate the effect of student mobility itself, estimate the effect of peer ability and examine how student mobility alters peer effects. In this appendix, we further describe our extension of the model with a particular focus on the estimation of peer effects using the additional fixed-effect terms.

Estimating the effect of peers on achievement is challenging considering that parents deliberately choose schools for their children so that the composition of schools is far from random. Even a comprehensive set of control variables is unlikely to solve this problem as long as unobserved factors are related both to the treatment and the outcome. To circumvent this problem, we use a sophisticated fixed-effect approach. The administrative student records from CPS include repeated observations for all students from multiple grades, years and schools. The comprehensive nature of the dataset allows us to extend our models with different levels of (current) school, grade and year fixed effect terms. This approach not only provides reliable estimates of peer effects but also isolates the mobility effect from both the current and the previous context, which has not been done in previous studies of student mobility. It compares the test-score growth of students who change from school  $a$  to  $b$  with the test score growth of students in *both* of these schools.

To implement this approach, we measure peer ability on the school-grade-year (cohort) level and supplement the common student fixed-effect approach used in the literature (e.g. Lauen and Gaddis 2012; Grigg 2012) with school-

by-grade, school-by-year, and grade-by-year fixed effect terms. Together, these terms not only account for individual heterogeneity but also for unmeasured school characteristics that are constant across time and grades and many other factors (for a similar approach see Hoxby 2000; Hanushek and Rivkin 2009). This estimation strategy is based on the assumption that the remaining variation in peer characteristics is unrelated to unobserved factors predicting student performance (i.e. the error term). To develop a better understanding of the remaining variation, it is important to consider each of the different fixed-effect terms.

First, *school-by-grade* fixed effects use cohort-to-cohort variations to identify the effect of peer ability. These variations refer to differences in peer composition and performance over the years for a single grade within a particular school such as the difference in peer ability between 1995 and 1996 in school  $s$  and grade  $g$ . This cohort-to-cohort comparison within schools accounts for all school-grade-specific factors that are stable over the years such as the school leadership, curriculum, teachers, neighborhood and many others. Changes in these factors such as a new school principal, teacher or a revised curriculum, however, are not taken into account and pose a challenge when they are systematically related to our outcome.

*School-by-year* fixed effects as the second term address this problem and account for time-varying school and neighborhood factors as long as they are constant across all grades within a school such as a new school principle, changes in school policies, or changing characteristics of the neighborhood. Specifically, school-by-year fixed effects identify peer effects based on variations across grades within a school for a particular year such as the differences between 4th and 5th grade in school  $s$  and year  $t$ . They ignore, however, changes that are related to the age or grade of students such as increasing enrollment of students from a particular background and with a higher or lower performance over the

past years or a very effective teacher that only teaches certain grades. Finally, we include *grade-by-year* fixed effects, which account for grade-specific curriculum, test difficulty or policy changes that apply to all students from a particular grade across the district.

Together, these three terms comprehensively captures almost all confounding factors and amount to what Hanushek and Rivkin (2009, 379) describe as a “large number of quasi- experiments” for each of the available schools. We further combine the three factors with student fixed-effects that also capture all observed and unobserved characteristics on the students level. This combination of individual with multiple levels of school, grade, and year fixed effects goes beyond the approach recently developed by a number of economists who simply control for prior performance and other individual characteristics (Hanushek and Rivkin 2009) or use aggregate data on the school-grade-year level (Hoxby 2000). Formally, this extension of our models can be expressed as

$$y_{isgt} = \alpha_i + \mu_{s*gt} + \gamma_{sg} + \lambda_{st} + \eta_{gt} + \theta \bar{y}_{sgt}^{(-i)} + \delta_1^* D_{it}^{(1)} + \dots + \delta_4^* D_{it}^{(4)} + \varepsilon_{isgt} \quad (2.3)$$

This models is the same as equation 2.2 on page 87 with an integrated term  $\bar{y}_{sgt}^{(-i)}$  for peer ability, which is the average performance in the previous year of all student in a certain school, grade and year except the individual  $i$  (indicated by the superscript  $-i$ ). In addition to isolating the effect of student mobility, these extended models allow us to thoroughly estimate the effect of peer ability and its interaction with student mobility so that we can evaluate how students who recently transferred to a school are affected differently by their peers (the interaction term  $D_{it}^{(x)} \times \bar{y}_{sgt}^{(-i)}$  is omitted from the formula). For peer ability, the coefficient  $\theta$  is the crucial statistic and represents the difference in performance associated with a one standard deviation change in peer ability for a particular

cohort (conditional on the whole set of fixed effect terms). Under the assumptions discussed above, this estimate can be interpreted causally as the effect of peer ability on student achievement.

**Remaining bias** - While our approach accounts for a multitude of factors, changes that are particular to a specific cohort and systematically related to the outcome across different schools remain a challenge. Such processes are difficult to envision. Neighborhood changes and a new school principal, for example, are constant across all students from a particular school and not grade-specific so that they are subsumed in the school-by-year fixed-effect term. Excellent teachers for a particular grade affect multiple cohorts of students and as such are captured by our school-by-grade fixed-effect term. A potential problem, however, are students who change school in response to a particular cohort. Such cohort-specific mobility processes refer to students who transfer to a certain cohort not because of the school in general as well as students who leave a school because of their particular cohort. These processes, however, are only problematic if they are common across schools and systematically related to the outcome variable. We evaluate this potential thread by modeling the in- and out-mobility defined as the ratio of transfers into or out of a particular cohort (school-grade-year level) as a function of our three fixed effect terms and a number of cohort characteristics such as cohort size, and composition in terms of ability, race, and socioeconomic background. The findings indicate that neither cohort size nor student composition are related to in- and out-mobility conditional on the fixed effect terms. It remains possible, however, that certain teachers attract students to a particular cohort in certain years. Unfortunately, we are unable to evaluate the role of teachers for the in- and out-mobility of certain cohorts. Instead, we examine whether the deviations in in- and out-mobility from the fixed effect terms (the error structure) are larger than

expected for certain cohorts, which would indicate that mobility patterns are based on something else that is not captured in our model such as teachers attracting students to or repelling from particular cohorts. Our findings show no evidence for such a pattern indicating that student mobility is not cohort based but instead follows patterns that are captured with our fixed effect terms.



## Article 3

# High School Environments, STEM Orientations, and the Gender Gap in Science and Engineering Degrees

*Despite the striking reversal of the gender gap in education, women pursue science, technology, engineering, and mathematics (STEM) degrees at much lower rates than their male peers do. This study extends existing explanations for these gender differences and examines two important and related dimensions: the life-course timing of a stable gender gap in STEM orientation, and variations across high schools. We argue that the high school years play an important role for gender differences in orientation towards STEM fields as students develop a more realistic and cognitively grounded understanding of their future work lives. During this period,*

*the gender-specific formation of career aspirations is not only shaped by widely shared and hegemonic gender beliefs but also by the local environment in school. Together these two dimensions extend existing explanations of the gender gap in STEM degrees and open concrete avenues for policy intervention. Using the National Education Longitudinal Study (NELS), we then decompose the gender gap in STEM bachelor degrees and show that the solidification of the gender gap in STEM orientations is largely a process that occurs during the high school years. Far from being a fixed attribute of adolescent development, however, we find that the size of the gender gap in STEM orientation is quite sensitive to local high school influences; going to school at a high school that is supportive of a positive orientation by females towards math and science can reduce the gender gap in STEM bachelor degrees by 25% or more.*

### **3.1 Introduction**

When then-Harvard President Lawrence Summers pointed at innate differences between men and women as a possible explanation for women's under-representation in high level science positions, he sparked an intense public controversy that mirrors a continuing debate in the scientific community. Despite the striking reversal of the gender gap in educational attainment (Buchmann and DiPrete 2006) and the near gender parity in math performance (Hyde et al. 2008), women still pursue science, technology, engineering, and mathematics (STEM) degrees at much lower rates than their male peers do. Figure 3.1 illustrates these trends. It shows, on the one hand, how women have made impressive gains in college attainment compared to men and now clearly outnumber men among college graduates in recent decades. On the other hand, women continue to lag behind in terms of bachelor degrees awarded in the physical sciences, mathematics,



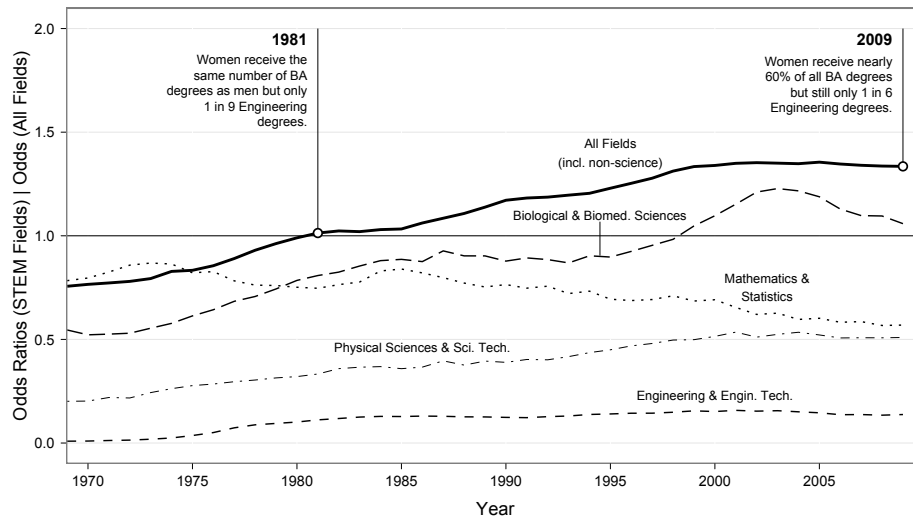
and engineering (illustrated in the graph for different STEM sub-fields).<sup>1</sup> The gender gap in STEM degrees has negative implications for the supply of qualified labor in science and engineering and for the closing of the gender gap in earnings. From the perspective of gender theory, the gender gap in STEM degrees is a prominent example of how gender stereotypes shape educational decisions and behavior. As such, the persistent pattern of gender differences in college science majors and the implications for later career choices and labor market earnings has been a major concern for scholars and policy makers alike.

In this paper, we argue that the high school years play a particularly important role in solidifying gender differences in orientations towards STEM fields that begin to emerge in early childhood. During this period, students develop a more realistic and cognitively grounded understanding about the world of work and refine conceptions of where they fit in this world by relying on gendered expectations about appropriate jobs, considerations about work-family balance, and self-assessment of career-relevant abilities. The formation of these expectations and understandings during adolescence is necessarily shaped by the environment, but the literature has under-appreciated the implications of the fact that the global character of societal gender beliefs is filtered by the local environment, including the influence of peers and teachers, the level of exposure to information about STEM fields and occupations, and local variation in the ways in which such information is gendered. In this paper, we document and explain the critical importance of experiences during the high school years for the persisting gender gap in STEM degrees, and we demonstrate how gender differences during this period are shaped by variations in high school environments across the United States. We show that these two dimensions of life course timing and local environment open concrete avenues for policy inter-

---

<sup>1</sup>Exceptions to this trend are the biological, biomedical and life sciences, in which women today outnumber men.

Figure 3.1: Gender Gap in Bachelor Degrees Awarded by Field of Study, 1969-2007



Data Source: Digest of Educational Statistics 2009, Table 268, 299, 303, 305, 312 and 313

Note: The trend line for all fields shows the odds that a BA degree is awarded to a woman, and the lines for the different subfields show the female/male odds ratio for the respective STEM field.

vention as well as being of central theoretical importance to our understanding of the gender gap in orientations to STEM fields and in graduation with STEM degrees from college.

Our theoretical argument yields a number of concrete expectations that we evaluate with data from the National Education Longitudinal Study (NELS). In particular, gender differences in the orientation towards STEM fields are already pronounced in middle school but these early orientations are unstable. Because early orientations begin to solidify after middle school, the high school environment plays a decisive role for the persisting gender gap in STEM degrees. During the high school years, we expect substantial variation in the gender gap across schools, and we hypothesize that the gender gap is smaller in schools with a strong STEM high school curriculum. In order to evaluate our argument, we first decompose the gender gap in STEM BA degrees into different

pathways, and thereby show how orientations towards STEM fields emerge and change from eighth grade through college. We find that the substantial gender gap in eighth grade orientation is relatively inconsequential for the persisting gender gap in STEM degrees at the completion of college. Instead, the high school years play a major role in shaping gendered orientations towards science and engineering. Second, we use multilevel models and special NELS subsamples to document that the gender gap in STEM orientation in twelfth grade varies across high schools, and – using an approach that resembles 11 models in research on school and teacher effectiveness – we show large differences between high schools in the ability to attract students to STEM fields. We also estimate the causal effect of the high school curriculum on the gender gap in STEM orientation as a first step to pinpoint the concrete characteristics of high schools that attract students to STEM fields. Far from being a fixed attribute of adolescent development, we find that the size of the gender gap in STEM orientation is sensitive to local high school influences; going to school in a high school that is supportive of a positive orientation by females towards math and science can reduce the gender gap in STEM bachelor degrees by 25% or more. The significant effect of high school curriculum on the gender gap in STEM orientation also provides the beginning of an understanding about the source of the high school effect.

### **3.2 Explanations for the Persisting Gender Gap in STEM Degrees**

The most prominent explanations of the persisting gender gap in STEM degrees either focus on gender differences in math performance or a number of social psychological explanations connected to gender stereotypes, self-assessment of abilities, and work-family balance. Recent research on differences in math abil-

ity has shown that the gender gap in math performance (Hyde et al. 2008) and course taking (Xie and Shauman 2005, Ch. 2) has largely closed; female performance on math tests is very similar to that of males. Girls take at least as many math classes in high school as do boys, and the classes are at a similar level of rigor (Lee et al. 2007). These facts notwithstanding, gender differences in math ability continue to play an important role in the debate about the persisting gender gap in STEM degrees. The debate largely revolves around findings that males excel at spatial orientation and visualization (Kimura 2002, 142f), and are more likely to fall on the extremes of the performance distribution in standardized mathematics tests (Ellison and Swanson 2010; Lohman and Lakin 2009; Hedges and Nowell 1995). Biological theories suggest that these differences are at least partly the result of innate genetic, hormonal, and brain structure differences between males and females, and that they largely emerged through evolutionary processes driven by the different reproductive roles of men and women (Lippa 2005; Halpern 2000, Cha. 4). Recent investigations, however (Ceci et al. 2009; Penner 2008; Guiso et al. 2008; Andreescu et al. 2008), downplay the relative importance of biological factors and point at substantial cross-national variations in the size of the gender gap as evidence for the importance of cultural factors. Meanwhile, Ceci et al. (2009) casts persuasive doubt on the power of the spatial ability theory to account for observed gender differences in STEM degrees, and a recent study by Hoffman et al (2011) finds strong nurture effects based on a comparison of a patrilineal to an adjacent matrilineal society.

The second major debate focuses on cultural factors. Sociological and social-psychological gender theories view gender as socially constructed – i.e., as a product of gender stereotypes about femininity and masculinity. Gender stereotypes set up expectations about appropriate preferences and behavior, and thereby influence how boys and girls perceive themselves, how they perform their gender to construct their own identity, and how others perceive and

react to them as boys and girls.<sup>2</sup> Status expectation theory further argues that gender stereotypes typically include status beliefs that attach greater competence in valued skills to the advantaged status (Ridgeway 2001).

Gender stereotypes are relevant for the persisting gender gap in STEM degrees because they encompass beliefs about the implications of gender for housework and child-rearing, math and science ability and other academic skills, and occupational selection and career trajectories (Charles and Bradley 2002). In this line, surveys consistently show that women are less interested in STEM fields from early adolescence (e.g. Lapan et al. 2000; Lubinski and Benbow 1992). A number of studies also suggest that females are more interested in jobs involving people and social interactions, and emphasize intrinsic, altruistic, and social rewards associated with an occupation. Males, in contrast, are more interested in jobs involving physical objects and abstract concepts, and place a higher value on extrinsic rewards such as money, prestige, and power (Eccles 2007; Beutel and Mooney Marini 1995; Johnson 2001; Davies and Guppy 1996; Konrad et al. 2000). Gender stereotypes are also associated with the division of labor in the family. Since the construction of masculinity commonly places work at the center of adult life, boys tend not to experience conflict between their work and family roles (Arnold 1995; Eccles and Hoffman 1984). But because the construction of femininity emphasizes the primacy of the domestic sphere, conflict between work and family is a prominent feature of women's lives (Duxbury and Higgins 1991; Williams 2000). Young women anticipate this career-family conflict long before they experience it firsthand (Shauman 2008). Even career-oriented women may take a contingency approach to planning their future by choosing career paths that they perceive to be compatible with future family roles (Almquist et al. 1980; Angrist and Almquist 1993;

---

<sup>2</sup>This view is similar to the "doing gender" perspective West and Zimmerman (1987), according to which – using a recent interpretation from England (2005, p. 269) – "each of us is held accountable to make sense to others in terms of gender norms, even if none of us actually prefer or believe in the rightness of the norms."

Felmlee 1993; Okamoto and England 1999; Gerson 1985; Seymour and Hewitt 1997). Accordingly, stereotypes about appropriate occupations and social roles in the family might explain, in part, why adolescents' expectations about college major and occupation remain gender-typed (Lueptow et al. 2001; Wilson and Boldizar 1990).

Based on status expectation theory, Correll (2001) also argues for a gender bias in the self-assessment of career relevant tasks such as math skills and shows how bias affects career relevant decisions. According to the Expectancy-Value model (Eccles 1994; Eccles 2007) an individual's expectations for success and the value that he or she attaches to the task are directly related to individuals' educational and occupational choices. These status beliefs, like other aspects of STEM-related gender stereotypes, arise from the socio-cultural environment (Correll 2004; Hill et al. 2010). Along with gender differences in job values and expected adult roles, gender differences in perceived skills appear to attenuate women's interest in STEM fields (Correll 2001; Pajares 2005).

### **3.3 Timing and Local Variations of the Gender Gap in Science and Engineering Orientation**

While existing explanations of the persisting gender gap in STEM degrees have implications for the timing and the sensitivity of the gender gap to the local and global environment, existing research does not adequately appreciate the importance of these dimensions. In this paper, we extend existing theories to overcome this limitation. In particular, we argue that while gender differentiation in orientation towards science, math, and STEM fields unfolds from early childhood throughout adolescence, the high school years play a central role in solidifying these gender differences as students develop a more realistic and cognitively grounded understanding of their future work lives. Dur-

ing this period, gendered expectations about appropriate jobs, considerations about work-family balance, and the self-assessment of career relevant abilities are not only shaped by widely shared and hegemonic gender beliefs but also by the local environment in schools. This local influence on the gender gap in STEM fields and occupations is based on three processes that are related to the influence of peers and teachers, the salience of gender in the local context, and the level of exposure to information about STEM fields and occupations. We elucidate these processes below. Together, these two dimensions not only shed new light on the existing explanations of gender differences in occupational plans but also point to important directions for future policy interventions.

### 3.3.1 Life Course Timing in the Emergence of the Gender Gap

Psychological research consistently shows that gender stereotypes develop in early childhood (Maccoby 1998), but developmental psychologists also recognize that these gender stereotypes elaborate throughout childhood and adolescence. Based on a U.S. study of 247 children between 6 and 10 years, Cvencek et al. (2011) shows that math-gender stereotypes and self-concepts already exist in second grade. Martin et al. (1990) found that children by the age of 10 have attached gender stereotypes to predominantly male occupations such as plumber or construction worker and to predominantly female occupations such as nurse or hairdresser.<sup>3</sup> Importantly, while meta-analyses suggest that gender stereotypes in general become less rigid after ages five or six (Signorella et al. 1993), the gender stereotyping of occupations became *more* pronounced as chil-

---

<sup>3</sup>A number of studies support these findings. Liben et al. (2001), for example, found that both 6-8 year old and 11-12 year old children are aware of the gendered character of occupations (see also McGee and Stockard 1991), and that children generally see male occupations (both familiar and fictitious) as higher status than female. But boys saw a higher status difference between male and female dominated occupations than did girls, and boys showed much lower interest in female dominated occupations than girls did in male-dominated occupations, regardless of whether the occupation in question was real or fictitious. Other studies show that sex-typed behavior already exists as early as 2.5 years of age before children enter school or kindergarten (Golombok et al. 2008).

dren aged into the 11-12 range, according to Liben et al. (2001).

Children's occupational aspirations and self-evaluation of math skills are influenced by the gender stereotypes they acquire early in childhood, but a number of studies emphasize that early childhood processes become more realistic in high school (Hossler and Stage 1992; Horn and Nuñez 2000; Ginzberg et al. 1951; Howell et al. 1977). In their often cited theory of occupational choice, Ginzberg et al (1951) argue that the pre-high school "fantasy" aspirations are "diverted into more realistic anticipations mainly through awareness of the work world and work roles and perceived reality factors (perceived goal-blocks) that function to inhibit the maintenance of such "fantasy" desires" (Howell et al. 1977, 332). The theory develops three phases of decision making; early childhood up to about eleven years of age is dominated by "fantasy" choices, middle school is the time for tentative (but unstable) interests, capacities and values, and high school is the time for the beginning of realistic interests in specific occupations. In this final phase, adolescents consider their own abilities, life goals, and interests and the compatibility between their developing occupational interests and their educational goals.

While criticized by some researchers (e.g. Super 1953), this perspective on the importance of the high school years and particularly the argument about increased realism in career aspirations has been adopted by other theories (Coulson et al. 1967; Super 1980) and has found support in empirical research (Csikszentmihalyi and Schneider 2001; McNulty and Borgen 1988). Studies have shown that the occupational aspirations of high school seniors have a considerable association with the socioeconomic status of the jobs they work in as adults (Sewell, Hauser, and Wolf 1980). More recently, Tracey et al (2005, p. 1) examined changes in interests for particular fields as well as the congruence of these interests with occupational choice and emphasize "the importance of grade 12 as a time of focus for research and intervention, and for continuing investi-



gation of gender differences in adolescent career and academic development.” Part of this changing focus comes from emotional and cognitive maturation as the content of academics (including mathematics and science) becomes more complex and as adolescents gain a more sophisticated understanding of the world of work. The sharpening focus is also a response to environmental influence, including the socializing influence of family, peers, and teachers (Sewell et al. 1969c; Sewell et al. 1980). Tracey et al. (2005, p. 19) express the widespread argument in this literature that “the environment is imposing the need to make choices about where to go to school and what to study. This could result in some more realistic assessment of what the student likes and chooses.”

The solidification of interests, moreover, is heavily influenced by gender. Expectations about success in various educational or career routes, values attached to these alternative routes, and concerns about their implications for other life interests (particularly work-family balance) all vary by gender (Eccles 1994). The gender bias in self-assessment of math skills, for example, has immediate implications for expectations about success, and, as Correll (2001) has documented, affects occupational plans at the end of high school. Boys and girls also differ in terms of their relative advantage in various skill areas. While gender differences in math performance have faded over the last decades, girls outperform boys in reading, art, and language skills. As a consequence boys might choose STEM fields because of their relative disadvantage in this area, while girls might gravitate to the area in which they have a relative advantage. Overall, these factors jointly influence the considerations of high school boys and girls when they begin to seriously and realistically contemplate their career options.

### 3.3.2 The Local and Global Environment, and the Formation of Educational and Occupational Plans

Both biological theories about gender and cultural theories that conceptualize culture as a coherent system of global norms and expectations imply that gender differences in the orientation towards STEM fields are relatively insensitive to variation in the local environment. In a similar vein, Xie and Shauman (1997) argued that occupational aspirations are formulated via cognitive processes that involve “the whole social environment at the societal level as the ultimate source of sex-typing.” According to this model, environmental knowledge about gender is global in character (e.g., knowledge about sex typing of occupations in the global labor market) so that the global and not the local environment matters. They argued that “specific actors are viewed only as socializing agents of the larger environment [...] this cognitive process involves too many individuals for any single actor to play a dominant role” (pp. 238-239).

Xie and Shauman’s (1997) argument implies that gender differences are based on society-wide gender knowledge with little room for significant influence from the local environment. Other theories and most of the existing explanations of the persisting gender gap in STEM degrees similarly do not explicate whether and how the local context in addition to global gender beliefs shape the gender gap. We acknowledge that gender stereotypes and gender status beliefs are institutionalized at the level of the global environment as part of the broader gender system in modern societies (Ridgeway 2001; Ridgeway and Correll 2004). However, we argue that variation in the local transmission and salience of these stereotypes and beliefs produces substantial variations in gendered outcomes across local contexts during the decisive high schools years.

We emphasize three processes as being of primary importance. First, peers and teachers are important actors in the local environment who support and

encourage certain career paths for boys and girls while disparaging others.<sup>4</sup> Peer influence is partly based on the pressure to conform with expectations and partly on the tendency to model own behavior on the behavior of other students. Frank et al (2008), for example, argue that “girls are highly responsive to the social norms in their local positions” and show how math course-taking is influenced by the social context. Building on ethnographic studies that document the multiple strategies used by boys and girls to construct their own gender identities (e.g. Francis 2000; an Ghail 1994), Legewie and DiPrete (2012) argue that local environmental variation in the support and sanctions for certain ways of “doing gender” affects the size of the classroom- and school-specific gender gap in academic performance. Other studies have found that teachers can influence gender differences in orientation towards STEM for elementary school and college students (Habashi et al. 2009; Carrell et al. 2010). These considerations challenge Xie and Shauman’s model of global environmental influence and argue that heterogeneity in math and science interests emerge not just from individual differences but from differences in the local environment that shape how boys and girls perceive themselves and how they construct their own gender identities.

Second, following Ridgeway and Correll (2004), we argue that widely shared and hence global gender beliefs such as stereotypes about appropriate occupations or status expectations manifest themselves differently in everyday interactions. According to Ridgeway and Correll (2004), the core aspects of gender or the gender system are “widely shared, hegemonic cultural beliefs about gender” (2004, 510) and local interactions (or what the authors call *social relational contexts*) in which these gender beliefs are evoked, enacted, and ultimately reproduced in a self-fulfilling manner. While this perspective highlights

---

<sup>4</sup>Parents are, of course, an important source of individual variation, and they will indirectly affect the broader local environment through peer processes.

the global nature of many gender beliefs, it simultaneously points at the importance of the local context. The situation shapes the salience of gender as well as the ways in which the widely shared gender beliefs are evoked. "We might expect, for instance, that the way the sex composition of a student-teacher interaction implicitly evokes gender beliefs will shape not only the way the individuals enact their roles but also how they evaluate each other's performance in that situation" (Ridgeway and Correll 2004, 512). In this line, Park et al. (2011) recently found evidence that all-boys high schools increase the level of male interest in STEM fields in South Korea, but that all-girls schools do not have a corresponding effect on the proportion of females who major in STEM fields while in college. Correll (2004) illustrates the importance of the specific situation based on an experimental study. Her work shows that boys assess their own abilities higher compared to girls when students are exposed to the belief that men perform better at this particular task, whereas boys and girls evaluate their own performance in a similar way when they are told that men and women do not differ in their abilities. Accordingly, the salience of gender in the local environment can affect the extent to which widely shared gender beliefs such as status expectations are relevant in certain situations. While gender always plays an important role for the life of young adolescents, we argue that some environments foreground gender and magnify its influence, while other contexts put gender more in the background and diminish its influence. This process can occur through the actions of peers or teachers, the use of certain instruction methods, or through organizational characteristics of the school such as the gender segregation of extracurricular activities. As a consequence, widely shared and hegemonic cultural beliefs about gender are more salient in some schools than in others, and these differences produce variations in the gender gap in STEM orientations across high schools.

A third important aspect of the local environment is the local availability

of academic knowledge about STEM fields and of information about STEM careers. Information about occupations in adolescence is highly incomplete and therefore subject to local environmental variation. These differences are of potentially great importance for females, given that STEM fields – most notably in engineering and the physical sciences – are typically viewed as pathways to male careers. It is also important given the evidence from psychology that females tend to react more strongly to gender stereotypes involving STEM fields than males.<sup>5</sup> The greater level of sensitivity of women in this context opens the way for greater importance of the local environment in the extent of gender stereotyping of STEM occupations and fields of study, and its impact on the development of educational plans and occupational aspirations. As a consequence, we expect that knowledge about the actual character of science and mathematics in the local environment will weaken and counteract gender beliefs such as stereotypes and status expectations.

Together, these three processes imply that gender differences in occupational aspirations are shaped not only by global factors as suggested by Xie and Shauman (1997) but also by the local environment in school. Accordingly, we argue that high schools as an important part of the local environment shape orientations towards STEM fields and the gender gap in these orientations. We cannot as a matter of theoretical deduction make concrete predictions about the extent of local variation in peer culture and teacher orientations or in the availability of knowledge about STEM fields and STEM careers. However, we expect the variation to be considerable, and we of course know that organizational and curricular characteristics of high schools are far from uniform. We expect that local environment heterogeneity and the three theoretical processes

---

<sup>5</sup>Steele (2010) reported that men who watched videos that advertised a math, science, and engineering leadership conference were unaffected by the experimentally manipulated gender ratio of people in the video, but that women who were exposed to videos where 75% of the people were men had elevated heart rates, blood pressure, and sweating, and remembered more incidental features of the video and the experimental room.

identified above combine to produce meaningful variation in the size of the gender gap in orientation towards STEM fields across high schools. While this argument alone has implications for our understanding of the persisting gender gap in field of study and points at important directions for policy interventions, the three described processes also point at particular characteristics of the high school environment that might affect the gender gap. In addition to measuring the overall importance of the local high school environment, our empirical analysis addresses one specific aspect of this environment, namely the elaboration of the high school curriculum in math and science. Studies of high school curricula show wide variation in the extent and depth of course offerings in science and mathematics (Adelman 2006; Owings 1998). High schools with the strongest science and mathematics curricular offerings arguably also offer a more effective antidote to gender stereotyping and the discouragement of female interest in STEM fields that it stimulates. Accordingly, we expect that the gender gap in orientation towards STEM fields is smaller in high schools with a strong STEM curricular. We discuss in the conclusion some other specific high school characteristics that may have similar effects and that are promising directions for future research.

### **3.3.3 Summary**

The process of forming gender differences in orientation towards STEM fields unfolds from early childhood throughout adolescence but the high school years play a particular role for solidifying gender differences. During this period, students develop a more realistic and cognitively grounded understanding of their future work lives. Also during this period, gendered expectations about appropriate jobs, considerations about work-family balance, and self-assessment of career relevant abilities play an important role. These processes are shaped not only by widely shared and hegemonic gender beliefs but also by the local

environment. This local influence occurs through the influence of peers and teachers, the salience of gender in the local context, and the level of exposure to information about STEM fields and occupations.

Based on this argument, we formulated four concrete hypothesis. First, gender differences in the orientation towards STEM fields emerge early in childhood and are already pronounced in middle school (our pre-high school measure of orientation towards STEM). Second, these early orientations are unstable and begin to solidify after middle school so that the high schools years play a decisive role for the persisting gender gap in STEM degrees. Third, during the high school years, the local environment shapes the gender orientation towards STEM fields, and thereby produces substantial variation in the gender gap across schools. Fourth, the strength of the science and mathematics curriculum in high school influences gender orientations towards STEM fields differently for male and female students, and as a consequence, we expect a smaller gender gap in schools with a strong STEM high school curriculum. In the following sections, we examine the timing in the emergence of gender differences in the orientation towards STEM fields to evaluate our first two hypotheses, and then study variation in the gender gap at the high school level to evaluate hypotheses three and four.

### **3.4 Pathways to a STEM Bachelor Degree**

To evaluate our argument about the life-course timing of the gender gap in STEM orientations, we decompose the persisting gender gap in STEM degrees using data from the National Education Longitudinal Study 1988-2000 (NELS). NELS provides a large sample of eighth grade students, who were followed over time as they graduated from high school and entered the labor force or pursued post-secondary degrees. The panel structure of the data together with

the availability of detailed information on educational careers allows us to examine the educational paths that lead to a bachelor degree in STEM fields. Appendix A provides details about the sample restrictions and the variables used in the analysis.

Extending previous work down by Xie and Shauman (2005, Ch. 4), we decompose the probability that an individual graduates from college with a STEM bachelor degree into different possible pathways as defined by transition rates between STEM orientations at three stages of the educational career.<sup>6</sup> We use the orientation towards science and engineering in eighth grade as the origin state, which captures pre-high school gender differences.<sup>7</sup> As the second stage towards a STEM BA, we use the expressed intention to study a STEM field in college at the end of high school (twelfth grade). As the third and final outcome stage, we use graduating from a four-year college with a STEM bachelor degree by 2000 (8 years after the expected high school graduation date). Figures 3.2, 3.3, and 3.4 all show the same distribution of these three states for boys and girls but highlight different components of the transitions that produce the gender gap in STEM bachelor degrees. In particular, Figure 3.2 shows the pathway of persistence, which is defined as a continuing science and engineering orientation from one state to the next. The leakage pathway shown in Figure 3.3, in contrast, is defined as having a science and engineering orientation in one state but not the next. Finally, Figure 3.4 shows the gender gap in late entry, which is defined by transfer from a non-science orientation in one state to a science orientation in the next state. We calculate these transition rates within high

---

<sup>6</sup>Compared to Xie and Shauman's results (2005, Cha. 4), we find that the high school orientation is relatively more important and late entry transitions are relatively less important to the gender gap in STEM bachelor degrees. Our results are based on more recent data (NELS and not HSB) but our different findings are not an historical change; we find the same pattern in our reanalysis of High School and Beyond data.

<sup>7</sup>The orientation during eighth grade is measured using the occupational expectation by age 30. Respondents were asked 'What kind of work do you expect to be doing when you are 30 years old?' and one of the response categories was "Science/Engineering" (5.8%), which we used to define a science and engineering orientation in eighth grade.



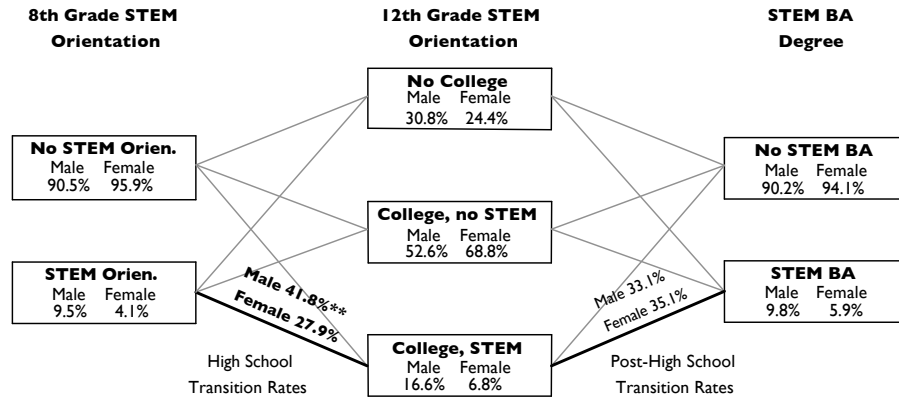


Figure 3.2: Persistence in the Pathway to a STEM BA Degree

Source: *National Education Longitudinal Study, 1988-2000*

Note: Asterisks (\*  $p < 0.05$ , \*\*  $p < 0.01$ ) and bold font indicate whether the gender difference in a specific transition rate is significant. The overall sample size is 7,060 with 3,700 female and 3,360 male students so that the smallest cell-size is 150 for girls with a 8th grade STEM orientation.

school (*high school transition rates*) as the probability of each twelfth grade state conditional on the eighth grade orientation, and we calculate *post-high school transition rates* as the probability of obtaining a STEM BA conditional on the orientation at the end of high school. This decomposition allows us to examine our argument about the life-course timing of the gender differences. It shows at what stage of the life course gender differences in the orientation towards science and engineering emerge, and at what point these differences become stable and have lasting implications.<sup>8</sup>

The results presented in all three figures show a substantial gender gap in eighth grade orientation towards science and engineering. Boys are more than twice as likely as girls to expect to work in science or engineering in middle school (9.5% compared to 4.1%). This finding confirms our first expectations and is in line with earlier studies reporting gender differences in the orienta-

<sup>8</sup>For simplicity, we do not distinguish between respondents who did not graduate from college and those who graduate with a non-STEM major. We thereby simplify the decomposition to exclude gender differences in rates of STEM bachelor degrees that arise from gender differences in the probability of getting a BA.

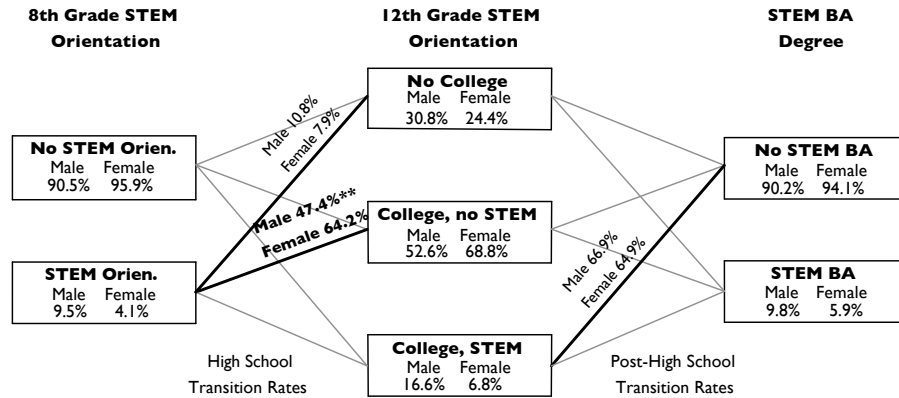


Figure 3.3: Leakage from the Pathway to a STEM BA Degree  
(see Figure 3.2 for data source and details)

tion towards and perception of math and science from early childhood to adolescence (Jacobs et al. 2002). Comparing Figures 3.2 and 3.4, we see that eighth grade STEM orientation predicts twelfth grade STEM orientation for both boys and girls. Thus, 41.8% of males with an eighth grade STEM orientation have a twelfth grade STEM orientation (persistence), as compared with only 13.9% of males who lacked an eighth grade STEM orientation (high school entry). Similarly, 27.9% of females with an eighth grade STEM orientation have a twelfth grade STEM orientation, as compared with only 5.8% of females who lacked a STEM orientation in eighth grade. Accordingly, boys are more likely to persist in and enter a science orientation during the high school years than are girls (41.8% compared to 27.9% for persistence and 13.9% compared to 5.8% for late entry; see Figure 3.2 and 3.4). The gender gap in persistence rates, however, disappears after high school. In other words, once high school seniors have developed an orientation towards science and engineering, boys and girls are equally likely to pursue this orientation after high school and actually graduate from college with a STEM BA (33.1% compared to 35.1%, the difference is not statistically significant). The same pattern is shown in Figure 3.3 in terms

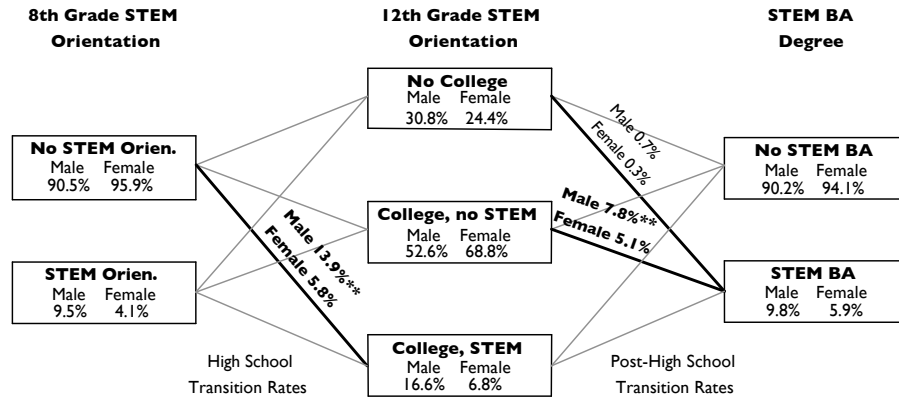


Figure 3.4: Late Entry as a Pathway to a STEM BA Degree  
(see Figure 3.2 for data source and details)

of leakage rates. It shows that girls are more likely to change their orientation from a STEM to a non-STEM orientation from eighth to twelfth grade (64.2% for females; 47.4% for males), whereas these leakage rates are the same for the post-high school period (64.9% for females, 66.9% for males). The gender gap in entry, however, remains substantial even in the post-high school period so that boys are more likely to be recruited for STEM field both during the high school years as well as after graduating from high school. Thus, the results in Figure 3.4 show pronounced (and highly significant) gender differences in the rate of entry into the science track for both the high school and the post-high school transitions rates.

Overall, these gender differences in initial eighth grade distributions and in the three transition rates of persistence, leakage, and late entry lead to a substantial gender gap in STEM degrees by the end of college: only 5.9% of female college graduates obtained a STEM bachelor degree, as compared with 9.8% of males. To determine the contribution of the different components of the decomposition to the overall gender gap, we successively assigned women the male orientation distribution as of eighth grade and the various male transition rates,

and calculated how the gender gap would change under these hypothetical scenarios. Table 3.1 presents the results from these simulations, which confirm our expectation that the high school year play an important role for the gender gap in STEM degrees. They show that the science orientation during eighth grade only plays a marginal role even though we observed a substantial gender gap in the orientation towards science during eighth grade. In particular, the gender gap would be reduced by 10.5% if women had the same eighth grade science orientation as men. The reason for this small importance of pre-high school orientation is the role of the high school years in shaping the science orientation of boys and girls. If males and females had the same transition rates within high school, the gender gap in STEM BAs would be reduced by a substantial 55.3%. The combination of eighth grade orientation and within-high school transition rates mathematically identify the twelfth grade science and engineering orientation, which accounts for 68.5% of the gap (i.e., 68.5% of the gap would disappear if women had the same twelfth grade orientation as men). Post-high school transition rates, and in particular gender differences in the rate of late entry into the science track, play the second most important role for the gender gap and account for 47.1% of the gap.

In Appendix B, we also present the same results for the subset of academically talented students, and for different STEM subfields, which seems particularly important considering the substantial differences in the trends for STEM subfields shown in Figure 3.1. As it turns out, the results from these additional analyses closely resemble the findings for the overall gender gap reported above (for further details see Appendix B).

In sum, the results suggest that the gender differences in transition rates during the high school years play a decisive role in shaping personal orientations toward science. During these years, girls are much more likely to abandon a science career even when they expressed interest in eighth grade, and boys

Table 3.1: Decomposition of Gender Gap in STEM BAs, NELS 1988–2000

Percent with STEM BA Degree	Male	Female	Gender Gap	Percent Reduced (Personal)
Observed	9.83	5.98	3.84	
<i>Changes if Females are Assigned Male Values</i>				
Same 8th Grade Orientation		6.38	3.44	10.45
Same 12th Grade Orientation		8.61	1.21	68.51
Same HS (8th →12th) Transition Rates		8.11	1.72	55.31
Same HS Entry		8.7	1.12	70.8
Same HS Persistence		6.18	3.65	5.17
Same Post-HS (12th → BA) Transition Rates		7.79	2.03	47.1
Same Post-HS Entry		7.93	1.9	50.66
Same Post-HS Persistence		5.84	3.98	-3.56
Same Late Entry Rates		10.65	-0.82	121.46
Same Persistence Rates		6.03	3.79	1.31

*Note: Late Entry - transition rate from no science orientation to science orientation either from 8th to 12th or from 12th to STEM BA. Persistence rate - science orientation in 8th grade to science orientation in 12th grade and from there to a STEM BA.*

are much more likely than girls either to persist or to enter a science and engineering oriented educational path. The importance of the high school years is also reflected in the high proportion of the gap accounted for by gender differences in the transition rates. These findings confirm our expectations and show that although a large gender gap in STEM orientation already exists in 8th grade, this gender gap is relative inconsequential. Instead, the high school years play a decisive role for the gender gap as a period when students begin to realistically consider different career options and develop a more cognitively grounded understanding of their future life, which renders many of the processes that produce the gender gap highly relevant during this period (e.g. self-evaluation of own skills). The second most important component for the gender gap in STEM BAs is gender differences in the entry into the science track after high school. Once graduated from high school, girls are much less

likely to change from a non-STEM orientation to a STEM major.

### **3.5 The Role of High School for the Gender Gap in STEM Orientation**

The last section demonstrated our first central contention that the high school years play a crucial role in shaping the orientation towards science and engineering among boys and girls. We now examine the role of the high school context for shaping orientations towards science and engineering during this decisive period. Our argument suggests that the local context in high school plays an important role for the gender gap in orientation towards STEM fields. In particular, we expect substantial variations in the gender gap across high schools even after adjusting for selection processes and a positive influence of the high school STEM curriculum as a concrete characteristic of schools on the STEM orientation of girls. In order to evaluate these hypotheses, we first use multilevel models that document how the gender gap in STEM orientation at the end of high school varies across schools and then estimate the causal effect of the high school curriculum on the gender gap in STEM orientation.

For this purpose, we use two special samples from the National Education Longitudinal Study. Compared to the 1988 to 2000 panel study (NELS 88-2000), these two special samples only follow the students until their senior year in high school, but they offer important advantages for our analytic goals. NELS 88-92 includes the full eighth grade sample of NELS (about 25,000), which is a much larger sample than NELS 88-2000<sup>9</sup>, as well as important pre-high school variables about the early science and engineering orientation. The NELS 88-92 sample does not, however, generally include a large number of students per high school because eighth-grade students in the same school typically tran-

---

<sup>9</sup>Only a randomly selected subset of students were followed after high school.

sitioned to more than one high school. The NELS High School Effectiveness Study (HSES), which is the second dataset we use in the following analysis, addresses this problem. As an independent component of NELS, the HSES extended the sample of students in a subset of 250 high schools in the first follow-up 1990 so that these schools had a sufficiently large number of students per school to support our analytic strategy. In contrast to NELS 88-92, however, HSES does not include pre-high school information. The sample restrictions for both datasets, the multiple imputation procedure used to recover missing data, and the variables are described in Appendix A.

In the following analysis, we use both NELS 88-92 and HSES to study the variation of the gender gap across schools conditional on regional and urban variations and a large set of pre-high school control variables. For this purpose, we specify a logistic multilevel model that captures variation in the gender gap in twelfth grade STEM orientation across schools. The outcome variable is the STEM orientation in twelfth grade at the end of high school. A value of 0 indicates that a student does not intend to study a STEM field in college after graduating from high school, whereas a value of 1 indicates that a student intends to study a STEM field after high school. The results from this analysis show substantial variation in the gender gap across high schools even after controlling for a comprehensive set of pre-high school covariates. This finding, described in more detail over the next paragraphs, indicates that the high school context plays an important role in shaping the gender gap in STEM orientation and therefore confirms our argument about the importance of the local context.

We begin with a simple model that only includes female as an independent variable along with a random intercept and a random slope at the school level, which allows the effect of gender on high school STEM orientation to vary across schools. The results from this multilevel model fitted with the HSES as well as the NELS 88-92 data are presented in Table 3.2. The estimated coef-

Table 3.2: Gender Effect in STEM Orientation across Schools

	NELS HSES		NELS HSES		NELS 88-92		NELS 88-92		NELS 88-92	
	Coef.	(se)	Coef.	(se)	Coef.	(se)	Coef.	(se)	Coef.	(se)
<i>Fixed Effects</i>										
Intercept	-2.093***	(0.06)	-1.991***	(0.14)	-1.599***	(0.03)	-1.719***	(0.09)	-2.127***	(0.10)
Female	-0.906***	(0.08)	-0.985***	(0.09)	-1.109***	(0.05)	-1.118***	(0.06)	-0.897***	(0.06)
<i>Control Variables</i>										
Std. Demographic Variables			✓				✓		✓	
Region-Urban Variables			✓				✓		✓	
Pre-High School Variables									✓	
<i>Random Effects</i>										
SD for Intercept	0.548		0.393		0.247		0.243		0.275	
SD for Female	0.355		0.424		0.423		0.361		0.303	
<i>Log-Likelihood Ratio</i>										
Chi-square	104.2***		109.2***		350.6***		279.42***		270.62***	
P-value	0.000		0.000		0.000		0.000		0.000	
Schools	250		250		1,280		1,280		1,280	
Avg. Students per School	37.1		37.1		10.2		10.2		10.2	
Students	9,120		9,120		13,640		13,640		13,640	

Note: A detailed description of the control variables is in Appendix Table A3.4. The likelihood ratio test compares the model reported here with a model that omits the random effect for gender and therefore tests whether the effect of female varies across high schools. \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$



ficients show a substantial gender effect; the odds of reporting an intention to study a STEM field in college at the end of high school for female are about 60% (HSES) or 70% (NELS 88-92) lower than the odds for males (the female/male odds ratio are 0.4 and 0.3 as calculated from the coefficients on the log-odds scale reported in the table). The results in the last section showed that this substantial gender gap at the end of high school is decisive for the later gender gap in STEM bachelor degrees, accounting for nearly 70% of the gap. The gender gap in personal STEM orientation, however, varies substantially across high schools. Specifically, the estimated standard deviation of the random effect on the school level implies that the gender gap ranges from 0.20 to 0.82 female/male odds ratios for 95% of the schools (these are the more conservative estimates from a HSES dataset). In other words, the odds for girls having a STEM interest are only 18% lower than the odds for boys in schools at one end of this spectrum, whereas in schools at the other end the difference is 80%. This variation is illustrated in Figure 3.5, which shows the distribution of the empirical Bayes estimates for the 250 high schools in HSES and the 1,280 high school in NELS 88-92. The graph also illustrates the analytical limitations of the NELS 88-92 dataset: Even though the estimated random slope for the variation of the gender effect across schools is bigger in NELS 88-92 than in HSES, the empirical Bayes estimates do not vary as strongly. The reason for this difference is the smaller average number of students per school in NELS 88-92. Empirical Bayes estimates are so-called “shrinkage” estimates; they are a weighted sum of the estimates from a single school and the estimates predicted for that school by data for the larger population. Empirical Bayes estimates for schools with a large number of students put more weight on the school-specific estimate, while empirical Bayes estimates for schools with a small number of students put more weight on the overall gender gap so that their estimates are pulled more strongly towards the overall mean (for a discussion of this see Gelman

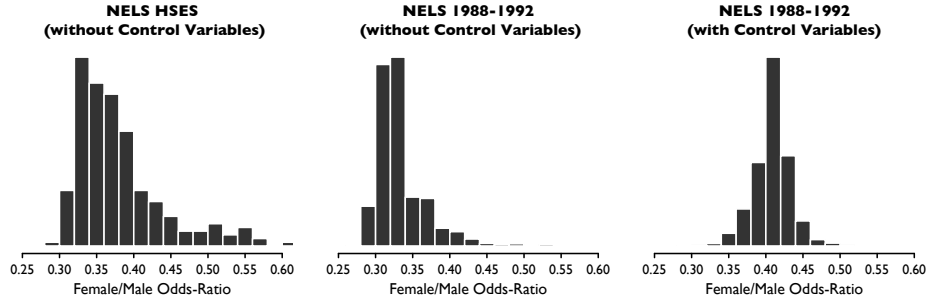


Figure 3.5: Variation of Gender Gap in 12th Grade STEM Orientation across Schools, NELS 1988-92

*Note: The graph reports female/male odds-ratios so that a value of 1 indicates gender equality and values closer to one –i.e. higher values in this graph– a smaller gender gap.*

and Hill 2007). Despite considerable shrinkage towards the overall mean, the NELS 88-92 data do contain enough students per school to reveal substantial variation (from 0.3 to 0.45 for the female/male odds ratio) in the gender slope across schools.

The revealed variation in the gender gap across schools might reflect the importance of the local school context, but it is also possible that they arise from a non-random sorting of students into different high schools such that girls with a strong science orientation are more likely to go to one school rather than another. In order to address this problem, we first add a number of standard demographic measures and a categorical region-urban variable to both the HSES and the NELS 88-92 and then use the NELS 88-92 sample to also condition on a large number of eighth grade orientation and performance measures (the variables are described in Appendix Table A3.4). We thereby obtain an estimate of the high school effect on science and engineering orientation that is conditional on the pre-high school science and math orientation as well as performance of students. The eighth grade orientation measures include not only the expressed occupational plans of eighth grade students used in the pathway analysis, but also four measures that assess whether middle school students like math and

science and whether they think that math and science is important for their future. The performance measures are comprehensive and include three eighth grade test scores (math, science, and English), and four GPA measures (math, science, English, and social studies). Because of this comprehensive set of control variables for family background, region-urban, pre-high school science and engineering orientation, and academic performance, these models can be understood as “value-added” models for STEM orientation.<sup>10</sup> Similar to value-added models in educational research on the effect of schools and teachers on performance (e.g., Kane and Staiger 2008), the empirical Bayes estimates from these models show the extent to which schools vary in supporting a science and engineering orientation among high school students, conditional on their previous orientation. They also show the extent to which schools are particularly supportive or unsupportive of a science orientation for girls net of the school’s support for a science orientation for boys.

The results from these models are presented in Table 3.2 as well as Figure 3.5. They show that although the estimated standard deviation for the school variation is smaller after pre-high school variables are controlled, the remaining variation in the effect of the local environment is still substantial and statistically significant. In particular, the estimated random slope from the multilevel model suggests that the gender gap ranges from 0.22 to 0.75 female/male odds ratio in 95% of the schools. Figure 3.5 shows the distribution of the empirical Bayes estimates, which indicate substantial variations across schools (the female/male odds ratio ranges from 0.34 to 0.50) despite the considerable pooling towards the overall mean in the NELS 88-92 sample.

---

<sup>10</sup>In the literature on teacher effectiveness, “value-added” is usually defined as the average gain in test-scores of the students a particular teachers taught conditional on student characteristics such as previous scores. Based on this definition, our model can be understood as “value-added” not in terms of test scores but rather in terms of the STEM orientation of the students in a particular high school conditional on a large set of student characteristics such as the 8th STEM orientation, performance and other measures. From this perspective, our model estimates the extent to which a school contributes to the development of an orientation towards STEM fields among it’s students.

Overall, the results so far confirm our first hypothesis about the role of high schools and show substantial variation in the gender gap in science and engineering orientation across schools. Net of science and math orientation in eighth grade, high schools appear to play an important role in shaping these orientations during the decisive high school years among boys and girls. This finding provides support for our argument that the local as well as the global environment shapes the gender gap in orientation towards STEM fields. It remains unclear, however, which particular characteristics of the high school explain these variations. In the remaining part of this section, we begin to explore this question and also address two important follow-up questions.

### **3.5.1 The Effect of High School Curriculum on the Gender Gap in STEM Orientation**

The analyses so far have demonstrated the importance of the high school years for the persisting gender gap in STEM degrees, and have shown substantial variation in the gender gap in STEM orientations across high schools, net of pre-high school orientation. In this section, we estimate the causal effect of the math and science curriculum in high school on the STEM orientation in twelfth grade for boys and girls. Based on the theoretical argument developed above, we would argue that a math and science orientation in high school as reflected in the course offerings in math and science has a positive effect on the STEM orientation for both boys and girls, and that the effect should be especially large for girls. We expect the strength of gender stereotypes about occupations to vary inversely with the level of information about these careers provided by the local environment, and about the relevance of gender to success in these careers. Given that girls currently perform at the same level as boys in advanced math and science high school courses, we expect that experiential knowledge of

this fact and of the actual character of science and mathematics – as revealed in advanced coursework – will weaken gender stereotypes and lead to a reduced gender gap in STEM orientation during high school.

In order to estimate the causal effect of the math and science curriculum in high school, we use the fact that the original NELS sample was first interviewed in eighth grade before students attended high school. This feature of the data allows us to condition on the same comprehensive set of pre-treatment variables used before. These variables are directly related to the selection of students into high schools with a strong math and science curriculum. A number of recent studies that compare experimental with observational estimates have shown that such a comprehensive set of pre-treatment variables is essential to reduce the bias in estimates that are based on regression or matching methods (Cook et al. 2009; Shadish et al. 2008b). These studies also suggest that the actual method used to estimate the effect – regressions based on the raw data or based on a matched sample – plays a negligible role relative to the importance of the right pre-treatment controls and despite the theoretical advantages of matching (for corroborating arguments, see Angrist and Pischke 2008). Accordingly, the estimates presented below are based on logistic regressions using a comprehensive set of pre-treatment control variables from eighth grade, including not only standard demographic measures but also the eighth grade orientation towards math and science, the extent to which a student reports that s/he likes math and science, and a set of seven GPA and test score performance measures for reading, math and science (for a detailed description of the variables see Appendix Table A3.4). The focal treatment variable is the intensity of the high school course offerings in math and science. We measure this variable based on a set of questions asked in the tenth grade school questionnaire about the courses offered at a school. In particular, we create an index based on the AP or college or university level courses offered at a school. The

concrete courses are selected based on the eight currently defined STEM AP classes in the US (the definition is based on standards set by the CollegeBoard, which is the sponsor of AP classes in the US).<sup>11</sup> Our focal treatment indicator is the standardized sum index with a mean of zero and a standard deviation of one from the questions that most closely match these eight STEM AP classes offered by the CollegeBoard. Appendix C contains a detailed description of the estimation strategy, the sample, the variables, an assessment of the balance between different levels of the treatment indicator, and additional sensitivity analysis.

A STEM-supportive high school environment can affect students by inhibiting conversion to a non-STEM orientation or by stimulating conversion to a STEM orientation during the high school years. Table 3.3 presents estimates of the effect of the curriculum index on the probability of having a STEM orientation at the end of high school for students who did not indicate a science orientation in eighth grade, conditional on a large set of pre-treatment covariates (hereafter, *the high school entry analysis*; the corresponding results for the persistence analysis are in Appendix C). Our prior analysis revealed a substantial gender gap in both the late entry and the persistence rate. It also indicated that the difference in male and female rates of late entry by gender (which can be understood as the ability of a high school to recruit boys and girls into the science track) plays a decisive role for the gender gap in STEM degrees. Accordingly, the high school entry analysis is of particular interest in revealing the potential impact of the high school environment.

The results in Table 3.3 show a substantial gender gap in the late entry rate as previously observed. In particular, boys have 2.3 times the odds of girls to transfer into a science track during high school. Consistent with our hypoth-

---

<sup>11</sup>The CollegeBoard AP classes in STEM are biology, calculus (AB & BC), chemistry, computer science, environmental science, different physics classes, and statistics.

Table 3.3: Logistic Regression Estimates for the HS Entry Pathway of STEM Orientation in 12th Grade on High School Math and Science Curriculum

	Model I		Model II		Model III		Model IV		Sensitivity Analysis	
	Coef	(se)	Coef	(se)	Coef	(se)	Coef	(se)	Coef	(se)
Intercept	-2.792***	(0.06)	-2.960***	(0.11)	-3.152***	(0.20)	-3.721***	(0.81)	-3.894***	(0.25)
Male	1.058***	(0.07)	1.111***	(0.07)	1.037***	(0.08)	1.059***	(0.08)	0.779***	(0.09)
Curriculum Index (CI)	0.262***	(0.06)	0.194**	(0.06)	0.142*	(0.06)	0.131*	(0.06)	0.071	(0.07)
Curriculum Index (CI) x Male	-0.247***	(0.07)	-0.256***	(0.07)	-0.217***	(0.07)	-0.224***	(0.07)	-0.064	(0.08)
<i>Pre-High School Control Variables</i>										
Standard Demographic Variables			✓		✓		✓		✓	
Urban/Region Variables			✓		✓		✓		✓	
8th Grade Variables					✓		✓		✓	
<i>High School Control Variables</i>										
Students	10,478		10,478		10,478		10,478		11,270	

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ ; Robust standard errors in parentheses.

Note: Control variables are described in Appendix Table A3.4. Model I-IV are only for the high school entry pathway - i.e. the students who had no STEM orientation in 8th grade. The sensitivity analysis is based on the full sample and replaces the dependent variables with a pre-treatment measure of the outcome variable (STEM orientation in 8th grade).

esis, the results also show that the curriculum index has a substantial positive effect on the late entry rate for girls but not for boys. The estimated effect (in units of odds ratios) is 1.12 (Model III), which implies that a one standard deviation change on the scale of the high school's math and science curriculum index produces a 12% increase in the odds that a girl will transition from a non-science to a science orientation than in the corresponding odds for a boy. This result was obtained after controlling for a comprehensive set of pre-treatment control variables for STEM orientation, academic performance, and other variables, and is stable across different model specifications. Because schools with strong science curriculums plausibly have greater resources and are of higher quality, one might ask whether the effect we measure is properly interpreted as a curriculum effect or whether our curriculum index is functioning as a proxy for high school quality. Evidence for our interpretation is the fact that the positive effect of the curriculum index persists after controlling for additional high school variables that measure school quality and resources such as the socioeconomic and racial composition of the student body, the drop-out rate, the student-teacher ratio, teacher salaries, and teacher education (Model IV).<sup>12</sup>

While the large and statistically significant positive effect of the high school curriculum for girls confirms our hypothesis, we did not expect to find a negative point estimate of the curriculum on the behavior of boys (main effect plus interaction) after controlling for the large set of pre-treatment control variables. Although the male effect is only marginally significant in a regression just for boys and generally smaller for other definitions of the treatment indicator (also see Appendix C), it might still be substantively meaningful. One possible interpretation would be that boys who over-evaluated their own performance

---

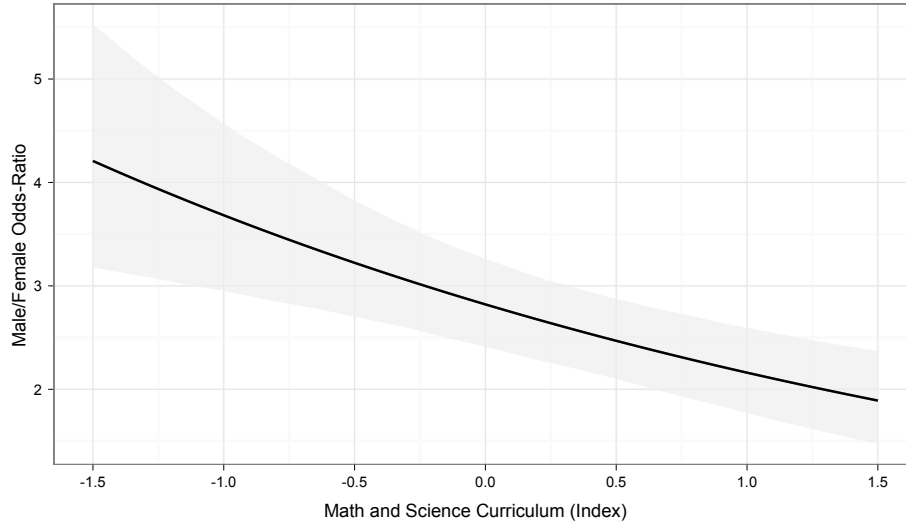
<sup>12</sup> These additional variables are not pre-treatment measures and as such might actually remove part of the HS curriculum effect from our estimate.



in middle school are suddenly confronted with the higher performing peers in STEM oriented schools, which in the end inhibits them from switching to a STEM orientation. This interpretation is plausible because boys are known to over-evaluate their skills compared to girls, but more research is needed to determine whether the negative effect is reproduced in future studies with data that allow the same quality analysis as the NELS data.

As a consequence of the interaction between curriculum and gender, the gender gap in STEM orientation narrows in high schools with a strong math and science curriculum, net of pre-treatment controls. Figure 3.6 illustrates this finding graphically and shows how the gender gap in terms of the male-female odds ratio in STEM orientation at the end of high school is smaller in schools with a strong math and science curriculum (the graph covers the range between the 10th and the 90th percentile of the math and science curriculum index).

Although we control for a large set of highly relevant pre-treatment control variables, our results might still be affected by unobservable variables that are related to the treatment and the outcome conditional on these variables. We perform two sensitivity analysis in order to evaluate this problem. First, we estimate the effect of our treatment indicator (the HS STEM curriculum) on a pre-treatment measure of the outcome variable, the STEM orientation in 8th grade (e.g. Imbens 2004). If the conditional independence assumption is satisfied, this effect should be close to zero and statistically insignificant simply because a pre-treatment measure of the outcome variable *cannot* be causally affected by the treatment. A positive effect, on the other hand, indicates that a selection process is at work that invalidates the conditional independence assumption. The results of this regression are presented in the right column of Table 3.3 and show that the effect is small and statistically insignificant. Accordingly, this sensitivity analysis suggest that no selection process is at work. Second, we examine how robust our female estimate is to additional unobserved cofounders.



**Figure 3.6: Effect of High School Curriculum on STEM Orientation**

*Note: The graph shows the gender gap in the entry into a science track during high school in terms of the male-female odds ratio as a function of the intensity of the math and science curriculum (estimates from third late entry model in Table 3.3). The graph covers the range between the 10th and the 90th percentile of the math and science curriculum index. The grey lines visualize the uncertainty in the estimated interaction effect based on 25 simulations (Gelman and Hill 2007, 140).*

Building on Ichino et al. (2008), we simulate an additional cofounder that is designed to eliminate the measured treatment effect. Our findings show that this additional, unobserved cofounder has to be relatively large compared to any of the observed covariates (including such key variables as eighth grade STEM orientation or eighth grade math test score) in order to invalidate our findings. This result supports the conclusion that our results are robust to certain violations of the conditional independence assumption. Appendix C contains a detailed description of this sensitivity analysis.

The logistic regression results for persistence are similar in size without the negative point estimate for boys (for details see Appendix C). The results are not statistical significant though, which might be related to the much smaller sample size for the persistence pathway compared to the high school entry

pathway (792 compared to 10,478). Accordingly, we can only conclude that the high school math and science curriculum helps to recruit girls into science and engineering, whereas it remains unclear whether it also helps to retain girls in science who have previously reported a STEM orientation. This pattern of results is consistent with the findings from our pathways model. The gender gap in STEM BAs is not primarily a consequence of girls losing their personal STEM orientation at a greater rate than boys, but rather from the lower rate of recruitment of girls into a STEM orientation between eighth grade and the senior year of college.

### **3.5.2 Is the High School Effect Lasting and How Big is the Effect?**

A common argument in the debate on the the effect of teachers on the learning of students is that potential gains in performance abate over the following years (Rothstein 2010; Jacob et al. 2010). A similar concern should apply to the effect of high school on the science and engineering orientation of boys and girls. If girls who were enrolled in high schools that were especially good recruiters of girls into a personal STEM orientation were to leak from the science pipeline at higher rates, the school effect would not be an important determinant of the gender gap in STEM bachelors degrees. In a recent review of interventions to increase female interest in science and technology (e.g., Turner and Lapan 2005; Plant et al. 2009), Hill et al. (2010) noted the uncertainty about the long term effects of these interventions that arises simply from the lack of long-term followup data. In this respect, the NELS data are attractive because they allow a direct assessment of the durability of high school effects on STEM orientations.

In order to conduct this assessment, we group high schools by the size of the gender gap in science and engineering orientation and examine the post-

high school transition rates used in the pathway analysis above. The results, reported in Appendix D, show that the post-high school transition rates are remarkably constant across the three samples. Neither of the transition rates differs significantly between the three samples. Accordingly, students from high schools that encourage a science and engineering orientation among women do not have higher leakage rates from the science pipeline than their peers from schools with a big gender gap. This finding suggests that the effect of high schools on the science and engineering orientation of women is not temporary, but instead endures after high school and ultimately reduces the gender gap in the attainment of STEM BAs. Accordingly, high schools seem to be an effective agent for policy initiatives to reduce the gender gap in STEM degrees.

Building on this findings, we ask how much the gender gap in STEM BAs would be reduced if all schools would encourage women to study science and engineering at the same rate as schools in the bottom tercile of the gender gap. As reported in Appendix D, the gender gap in STEM BAs would be reduced by about 25% if the environment in all schools would encourage girls to study science and engineering at the same rates as the top third of schools (from 1.7 male/female odds ratio in the entire sample to 1.3 odds ratio in the sub-sample of students who attend high schools with a small gender gap). The reduction would presumably be even larger if all schools could achieve the same results as the most gender-egalitarian schools in our sample.

### **3.6 Conclusion**

Despite the striking reversal of the gender gap in educational attainment and the near gender parity in math performance, women still pursue science, technology, engineering, and math degrees at much lower rates than their male peers do. Existing explanations of this persisting pattern of gender differences

focus on mathematical abilities or different gender beliefs that are related to gendered expectations about appropriate jobs, considerations about work-family balance, and the self-assessment of career relevant tasks. In this paper, we have extended these theories and explored two important and related dimensions of the persisting gender gap in STEM degrees: First, the timing in the emergence of the gender gap in orientation towards STEM fields, and second, variations in the gender gap across high schools. In particular, we have argued that the process of forming gender differences in orientation towards STEM fields unfolds from early childhood throughout adolescence, but that the high school years play a particular role for solidifying these gender differences as students develop a more realistic and cognitively grounded understanding of their future work lives. During this period, gendered expectations about appropriate jobs, considerations about work-family balance, and the self-assessment of career relevant abilities play an important role and are not only shaped by widely shared and hegemonic gender beliefs but also by the local environment. This local influence on the gender gap in STEM fields and occupations is based on three processes: the influence of peers and teachers, the salience of gender in the local context, and the level of exposure to information about STEM fields and occupations.

We then turned to an empirical examination of the two dimensions. First, we decomposed the gender gap in STEM bachelor degrees into various pathways to examine the emergence and solidification of gender differences in the orientation towards science and engineering in the adolescent life course. In particular, we used the National Education Longitudinal Study to follow the 1988 cohort of eighth grade students through adolescence and young adulthood, and we observed how orientations towards STEM fields emerge and change during these years. Our findings show that the substantial gender gap in eighth grade orientation is relatively inconsequential for the persisting gen-

der gap in STEM degrees at the completion of college. Instead, the high school years play a major role in shaping gendered orientations towards science and engineering. Second, we used multilevel models to examine how the gender gap in STEM orientation at the end of high school varies across schools, net of pre-treatment controls. The results show substantial variation in the gender gap in STEM orientation across schools, and support our argument that the local environment plays a major role in shaping and gendering orientations towards education and career among boys and girls. Our additional analyses show that this high school effect seems to be related to the math and science orientation of the school.

While our focus has been on STEM fields, our results potentially have broader implications for the distribution of majors for both males and females, for gender occupational segregation, and even for the gendered character of household work. Simply put, our results suggest that the local environment in which adolescents spend their high school years plays an important role in the strengthening or weakening of gender stereotypes. Our focus in this paper has been on the gender gap in STEM fields, but similar processes could be at work with respect to gender stereotypes concerning elementary or secondary school teaching, or interest in the humanities and the performing arts. Thus, just as some local environments pull adolescent girls away from an orientation consistent with gender stereotypes, and toward an interest in STEM fields, the same or other local environments might pull adolescent boys toward an interest in humanities, performing arts, or elementary school teaching. However, gender integration of occupations has occurred more through women moving into formerly male dominated occupations than through men moving into female dominated occupations, and the trend with respect to college majors has the same qualitative profile. This pattern reinforces other research to suggest that males are more concerned about violating gender stereotypes than are females, and it

may therefore be the case that males are more resistant to local environments that challenge gender stereotypes than are girls. Nonetheless, similar research to what we have reported here could be applied to a broader set of life course outcomes, and, we predict, the results would be highly informative about how variation in the coding of gender in local environments affects the distribution of gender roles and identities in adulthood.

From a policy perspective, our findings point to important directions for research about concrete interventions. The pathway analysis shows that the high school years are the decisive life period during which the gender gap emerges, and the examination of variations across contexts shows that the local context in high school plays an important role for the gender gap in orientations towards STEM fields. As such, our findings not only point at the life course period that should be targeted by policy interventions, but also provide evidence that high school interventions might be effective. Not all local environmental effects are necessarily durable, though. In light of recent research asserting only a temporary effect from exposure to Head Start programs or to individual above-average teachers (Jacob et al. 2010), it is of considerable importance that the effects of the high school environment on the formation of STEM orientations appear to be durable. Some existing interventions have indeed targeted high school students and shown success in promoting a STEM orientation among girls. Eisenhart (2008), for example, discusses a seemingly effective outreach project that educates high-achieving, minority girls in high school about science and engineering jobs. While such policy interventions have to withstand the serious scrutiny of experimental field trials, the evidence presented in this paper encourages researchers and policy makers alike to take seriously the potential impact of high school interventions on the STEM orientations of female students. Our finding that the intensity of the math and science curriculum reduces the gender gap in science orientation strongly supports this conclusion.

Our results also have implications for the future trend of gender segregation in STEM fields. At a minimum, they suggest that the propagation of more supportive local environments would increase the proportion of women interested in STEM fields. We noted in Figure 3.1 the increase between the early 1980s and 2005 in biological and biomedical sciences bachelors degrees obtained by women as well as their less-dramatic but still notable progress in the physical sciences and science technology bachelor degrees. This was also a period when the mathematics and science curriculums of high schools were strengthening, as measured by the fraction of students who took precalculus or calculus, or by the percentage of high school graduates who completed chemistry, physics, or advanced biology (Dalton et al. 2007). The fact that these trends both move in the same direction suggests that the expansion of the science curriculum in high schools may have been one factor increasing the fraction of STEM degrees awarded to women over these years.

The present study obviously falls short in adequately addressing all the characteristics of high schools that influence the gender gap. Similar to the state of knowledge about teacher quality, our findings suggest that high schools have the potential to shape the orientation towards STEM fields and suggest that the math and science orientation of the school might play an important role, but we still know relatively little about other high school characteristics or programs that achieve this goal. Our own theoretical argument suggests that the ways in which gender identities are constructed plays an important role. Our argument also suggests that commonly held stereotypes are strengthened by the lack of adequate information about science and engineering careers in the local environment, and conversely that the power of these stereotypes over behavior can be reduced through greater exposure to knowledge about science and engineering through the academic curriculum. A third argument was presented recently by Frank et al. (2008), who argue that social dynamics play an im-



portant role for the propensity of girls and boys to take math courses. Greater efforts to measure directly the strength of gender stereotypes concerning science and other careers might provide particularly valuable information about how the high school environment shapes gender identities and the career orientation of male and female students. Future research should investigate these issues in greater depth.

## Appendix A Samples, Variables, and Missing Data

The analyses presented in this paper are based on three samples from the National Education Longitudinal Study of 1988 (NELS). NELS is a nationally representative sample of about 25,000 eighth grade students who were first surveyed in the spring of 1988. Subsamples of these students were resurveyed in 1990, 1992, 1994, and 2000 so that the students were followed over time as they graduated from high school and entered the labor force or pursued post-secondary degrees. The panel structure of the data combined with the fact that it includes detailed information on educational careers allows us to examine the different questions addressed in this paper. We have briefly described each of the three samples in the main text and provide further details in this appendix. Note that all sample sizes are rounded to the nearest 10 (National Center of Educational Statistics requirement).

**NELS 1988-2000 Sample (Pathway Analysis)** We restrict the NELS 88-2000 sample to students who participated in the eighth and 12 grade survey (base year and second follow up), and the 2000 follow up (fourth follow up).<sup>13</sup> The size of this restricted sample is 8,320. From this sample, 1,260 (15.2%) cases are dropped because of missing data on the relevant variables, which brings the analysis sample down to 7,060.<sup>14</sup> All of the analysis use the appropriate weights provided by NELS.

**NELS 1988-1992 Sample (High School Context Analysis)** We restrict our analysis of the NELS 88-1992 to students who participated in the base year as well as the first and second follow up and to those for which the school

<sup>13</sup>This sample restriction excludes high school drop-outs. Although NELS followed students who dropped out of high school, information on intentions to go to college are meaningless so that it makes sense to exclude this group from our analysis.

<sup>14</sup>We obtained similar results using multiple imputations to recover the missing values (available from the authors). The pathway results presented here, however, are based on case-wise deletion because the literature on multiple imputations is not conclusive about using multiple imputations in our situation.

Table A3.4: Description of Variables

Variables	Description
<i>STEM Orientation in School and STEM Bachelor Degree Attainment</i>	
8th Grade STEM Orientation	Binary indicator based on occupational aspiration in eighth grade (“What kind of work do you expect to be doing when you are 30 years old?”) 0=not science or engineering (12 categories such as craftsperson, housewife, business owner and other professions) 1=“science or engineering professional, such as engineer or scientist”
12th Grade STEM Orientation	Our coding first uses the filter question “Do you plan to continue your education past high school at some time in the future?” to determine the people who plan to go to college. We then use the intended field of study question to distinguish between STEM fields (science, technology, engineering, or mathematics) or non-STEM fields. Pathway analysis: 1=no college, 2=College, no STEM field, 3=College, STEM field HS Effect analysis: 0=no college or college but no STEM, 1=College, STEM field
STEM bachelor	Attainment of bachelor degree from a four year college in a STEM field within eight years after the normal high school graduation. STEM field was determined based on the Classification of Instructional Programs (CIP) and is defined as a degree in science, technology, engineering, or mathematics.
<i>Demographic Control Variables (NELS 88-1992 and NELS HSES)</i>	
Gender	0 - Female; 1 - Male
Race	Categorical (reference category is White): Asian, Hispanic, Black (not Hispanic), White (not Hispanic), Native American
Age	continuous (standardized)
Family Status	Standardized socio-economic status composite constructed from father’s and mother’s educational degree and occupation, and family income (from eighth grade data for NELS 88-2000 and 88-1992, and from 10th and 12th grade data for NELS HSES)
<i>Pre-High School Control Variables (NELS 88-1992)</i>	
Test Scores	eighth grade reading, math, and science test scores (separate, continuous variables)
Performance (GPA)	Self reported English, math, science, and social studies grades from 6th to 8th grade (separate, continuous variables)
8th Grade Science Orientation	see above
Math/Science Interest	“I usually look forward to mathematics class” “I usually look forward to science class” (four point Likert scale, 8th grade)
Math/Science Usefulness	“Math will be useful in my future” “Science will be useful in my future” (four point Likert scale, 8th grade)
Math/Science Extra Curricular Activities	Three dichotomous indicators (8th grade) for participation in math club, science club, and science fair.
Middle School Variables	SES composition, average STEM orientation, presence of gifted program, math and science, student-teacher ratio, and school type
<i>Additional Control Variables (NELS 88-1992 and NELS HSES)</i>	
Region-Urban	Categorical variable with twelve groups defined by all the possible combinations of four large US regions (Northeast, North Central, South, West) and the urbanicity of the area (urban, suburban, and rural).

Note: All continuous variables have been standardized for the analysis.

filled out a school questionnaire during the first follow up. The size of this sample is 11,270.<sup>15</sup> 5,310 or 38.9% of these cases have missing data on at least one of the large number of pre-treatment control or other variables. To address this problem, we use multiple imputations based on the multivariate normal model approach with a large number of auxiliary variables such as 10th and 12th grade test scores in reading, math and science.

**NELS HSES Sample (High School Context Analysis)** The HSES sample used in our analysis contains all 9,740 students who participated in the twelfth grade survey. 1,720 or 17.63% of the students have missing data on at least one of the variables. Similar to our analysis based on the NELS 1988-1992, we use multiple imputations to recover cases with missing data.

Table A3.4 contains a detailed description of all the variables used in the different analyses. The main variables are the three measures of STEM orientation in school and the attainment of a STEM Bachelor degree. The 8th grade measure is based on occupational aspirations at age 30, and the 12th grade measure is based on the intended field of study for those who plan to go to college (for details about the question wording and answer categories, see Table A3.4). While the two measures are not based on the same question, they both capture the present and expected future orientation towards science and engineering, specifically, the intention to work and to study in STEM fields. The reliability of the eighth grade measure is supported by the high proportion of students that do not change their orientation. Over 40% of men persist in their science and engineering orientation from 8th grade to 12th grade, which is higher than the

<sup>15</sup>For the estimation of the curriculum effect we further restricted the sample. First, we also exclude students who did not change school between eighth and tenth grade and those who changed school between tenth and twelfth grade. Accordingly, the analysis of the curriculum effect is based on students who changed from a middle to a high school and proceeded to 12th grade without dropping out of school or changing the school. This restriction excludes 450 cases. Second, we drop about 1,500 students from schools that had no information at all about the course offerings.

persistence rate from 12th grade to a STEM BA degree and also high compared to other results reported in the literature (e.g. Jacobs 1989).

## **Appendix B Pathways for Academically Talented Students and Different STEM Subfields**

In this appendix, we present the findings from the pathway analysis for academically talented students, and for different STEM subfields. Academically talented students are defined in terms of their math performance in eighth grade. Restricting the sample to the top 33% students in terms of math performance leads to similar results but the patterns reported above are even more pronounced (see Table A3.5). The transition rates during high school play an ever bigger role (64.1% compared to 55.3%) and the different rates of late entry after high school a slightly smaller role (44.9% compared to 50.7%).

The trends presented in Figure 3.1 suggest that analyzing gender differences in STEM degrees should take into account the substantial differences between sub-fields. Accordingly, we analyze engineering, math and physics as well as bio/life-science respectively and look at how the pathways for these sub-fields differ from STEM fields in general. Table A3.5 presents the results of these decompositions. For both engineering as well as math and physics, the results resemble the findings from STEM fields in general. In all three cases, the transition rates in high school play the most important role in explaining gender differences in an engineering BA, while the lower rate of late entry for females after high school plays the second most important role. For biology and life science, the gender gap in bachelor degrees is small and not statistically significant, and therefore a decomposition analysis is not informative.

Table A3.5: Decomposition of Gender Gap in STEM BAs, Academically Talented Students and STEM Sub-fields

	Personal Orientation	Academically Talented Students	Engin- eering	Math and Physics	Bio/Life Sci- ences
Observed Gender Gap in %	3.844	6.034	4.025	0.424	-0.605
...in M/F Odds Ratio	1.713	1.619	3.421	1.382	0.799
<i>Percent Reduction in Gender Gap if Females are Assigned Male Values</i>					
Same 8th Grade Orientation	10.45	14.47	4.89	22.67	-23.78
Same 12th Grade Orientation	68.51	78.78	71.81	92.18	-70.99
Same HS Transition Rates	55.31	64.08	62.24	66.45	-48.37
Same HS Late Entry	70.8	83.96	61.2	-8.88	-11.05
Same HS Persistence	5.17	4.95	6.36	-0.94	7.01
Same Post-HS Transition Rates	47.1	38.16	32.26	48.43	28.56
Same Post-HS Late Entry	50.66	44.85	31.96	37.91	-34.36
Same Post-HS Persistence	-3.56	-6.69	0.30	10.52	62.92
Same Late Entry Rates	121.46	128.8	93.16	29.04	-45.41
Same Persistence Rates	1.31	-2.12	6.79	9.48	65.45

*Note: For the STEM sub-fields, "Late Entry" refers to entry into the STEM sub-field both from a prior non-STEM as well as from an orientation towards another STEM field.*

## Appendix C Estimating the Effect of High School Curriculum on the Gender Gap in STEM Orientation

The literature on the estimation of causal effects using conditioning on observed covariates as an identification strategy has recently focused on matching procedures. Theoretically, estimates based on matched samples provide some advantages over regression estimates based on the raw data. Most notably, matching estimates are quasi-nonparametric, and reduce the need to extrapolate by discarding or down-weighting observations without common support in the treatment and control group. A number of recent studies that compare experimental with regression and matching estimates, however, have shown that these theoretical advantages are typically negligible in practice (Cook et al.

2009; Shadish et al. 2008b; Angrist and Pischke 2008). These studies instead point at the importance of the actual set of observed covariates used in the analysis, and the measurement reliability of these constructs. They show that only a comprehensive set of pre-treatment variables that are measured reliably and that are directly related to the selection process (i.e., that go beyond a set of standard demographic measures) can effectively reduce bias from non-random selection into treatment.

In order to estimate the causal effect of the math and science curriculum, we use the fact that the original NELS sample was first interviewed in eighth grade, i.e., before these students attended high school. This feature of the data allows us to use a comprehensive set of pre-treatment variables that are directly related to the selection of students into high schools with a strong math and science curriculum. These variables include a comprehensive set of family background variables and other standard demographic measures, plus a set of variables related to the science and engineering orientation in eighth grade such as the occupational aspiration in eighth grade, the extent to which a student looks forwards to their math and science classes, and whether the student thinks that math and science is useful for their future. In addition, we control for eighth grade GPA and test scores in math, science, and reading (for a full list of variables see Table A3.4). These measures were selected because that they are directly related to the selection of students into high schools with a strong math and science curriculum. The measure of occupational aspirations in eighth grade can also be understood as a “proxy pretest” variable, which is particularly important for the reduction of bias (Steiner et al. 2010b). In addition, the high number of measures connected to the same underlying construct reduces the potential bias introduced by unreliable measurement of the key pre-treatment covariates. For our analysis, we refrain from using matching procedures because of the now-common finding that regression and matching

based estimates are equally good and because our treatment measure is continuous.<sup>16</sup> We do, however, analyze the balance of the data across different values of the treatment indicator below, and explore alternative model specifications to maximize balance.

**The Treatment Indicator** Our focal treatment variable is the intensity of the high school course offerings in math and science. We measure this variable based on a set of questions asked in the tenth grade school questionnaire about the courses offered at a school. In particular, we create an index based on the STEM AP or university level courses as defined by the CollegeBoard – the sponsor of standardized AP classes in the US. The CollegeBoard currently lists 32 AP classes, of which 8 fall into the the STEM area (biology, calculus AB & calculus BC, chemistry, computer science, environmental science, different physics classes, and statistics). Our focal treatment indicator is the standardized sum index with a mean of zero and a standard deviation of one from the classes that most closely match these STEM AP classes offered by the CollegeBoard.

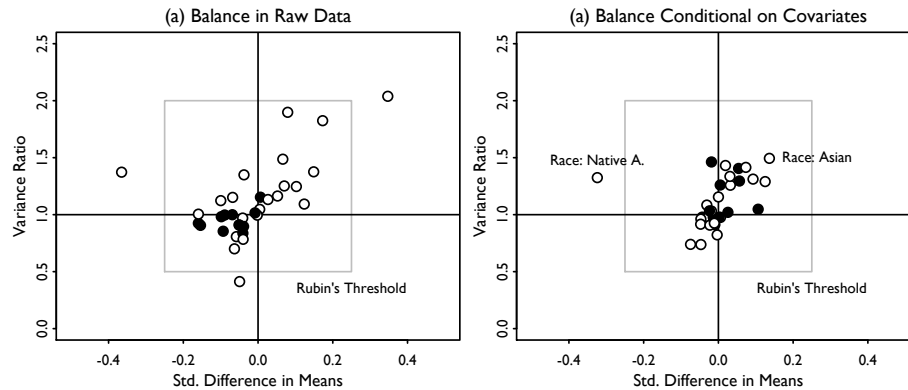
**Selection of Courses and Sensitivity of Findings to Definition of Treatment Indicator** The results reported in this paper are based on the count of AP or university-level courses offered in schools, where we selected specific courses based on the CollegeBoard definition of AP classes. These AP classes are nationally standardized with a clear curriculum and exam-based evaluation. The NELS school questionnaire in the first follow up, however, allows school administrators to designate classes in the broader math and science area as “Advanced Placement (AP) Courses” that most probably are not CollegeBoard-certified AP courses or university-level math and science courses.<sup>17</sup> Our results

<sup>16</sup>While the treatment indicator could be dichotomized, we believe that any gains from matching methods are more than offset by the precision lost from dichotomizing the treatment indicator.

<sup>17</sup>The NELS questionnaire asks the school administrators to ‘circle all that apply on each line’ whereby each line refers to one of the 34 classes and the options are “Course Not Offered”, “Regular



Figure A3.7: Balance between Top and Bottom Quartile of Continuous Treatment Variable



are somewhat sensitive to the actual selection of classes for the definition of the treatment indicator. Estimates from separate regressions where each of the 34 courses was used as a dichotomous treatment indicator range from -0.171 to 0.608 (with an outlier at 1.77) for girls (the estimate for the AP course-based index presented in this paper is 0.14). Most of these estimates are positive. The single-course dichotomous effects are particularly strong for math classes (including many that are not part of our AP course-based index) and for most but not all of the CollegeBoard AP classes.

**Balance between Levels of the Treatment Indicator** To examine the balance between different levels of the treatment indicator, we compare the first and fourth quartile of the treatment indicator with respect to forty pre-treatment covariates in the raw data as well as on the residualized covariates after conditioning on the control variables. These covariates include the control variables used in the final analysis but also a number of interaction terms between these

Course", "Advanced Placement (AP) Course", and "College or University Level Course". As a consequence of this question design, some of the cells in this 34x4 matrix do not make sense. The list includes, for example, an AP remedial math class, which supposedly is offered at one school. Other examples of classes that are dubious are a separate 9th grade AP class in general mathematics that is listed in addition to a 10-12 grade class or the many classes for which no "Advanced Placement" course is defined by the CollegeBoard.

variables that are not included in the final analysis. Figure A3.7 shows balance in terms of the standardized difference in means (x-axis) and the variance ratio (y-axis) for the raw data as well as conditional on the covariates in the final analysis. Solid circles indicate key variables such as eighth grade orientation, and math and science performance. These variables were selected based on theoretical considerations as well as the size of their effect in the selection and outcome model. The grey rectangle indicates Rubin's (2001) rule of the thumb, which suggests that the absolute standardized differences in means should not be greater than 0.25 and the variance ratio should be between 0.5 and 2. In general, however, imbalance should be reduced without limit. The results show that the covariates clearly increase the balance between the treatment and control group and that the highly relevant variables cluster around the point of no imbalance.<sup>18</sup> Overall, the results indicate that our estimation strategy is effective in reducing the observable imbalance between different levels of the treatment indicator. Alternative specifications did not reduce the imbalance further.

**Sensitivity Analysis - Robustness to Violations of Conditional Independence Assumption** Although we control for a large set of pre-treatment control variables that are directly related to the selection process, unobserved confounding variables might nonetheless bias our estimated effects. To estimate their potential impact, we conduct a simulation based sensitivity analysis. Other sensitivity analysis have been proposed for estimating the effect of confounding variables in propensity score matching analyses (Rosenbaum 2002), linear regression models (Frank 2000), or instrumental variable regression (DiPrete and Gangl 2004). Here we apply a simulation-based sensitivity analysis proposed by Ichino et al (2008; also see Nannicini 2007) for matching methods to the case

---

<sup>18</sup>The only variable that falls outside the rectangle is the dummy variable for Native Americans. This group, however, is extremely small with only 30 students in the control group and 20 in the treatment group.

of regression analysis with a binary dependent variable. The starting point of this and similar sensitivity analysis is to posit an unobserved variable,  $U$  (here assumed to be binary), that violates the conditional independence assumption. The binary covariate  $U$  can be simulated based on different assumptions and added to the regression model as an additional covariate in order to get an understanding about the robustness of the results to specific failures of the independence assumption.

For simplicity, consider a set of observed pre-treatment covariates  $X$  and three binary variables: the treatment variable  $T$ , the outcome  $Y$ , and an unobserved confounding variable  $U$ . In order to qualify as a confounding variable,  $U$  has to be associated with both the treatment and the outcome variable after controlling for  $X$ . If we make the further simplifying (and conservative) assumption that  $U$  and  $X$  are independent, conditional on  $T$  and  $Y$ <sup>19</sup>, the distribution of  $U$  can be characterized with a set of four probabilities  $p_{ij}$  that define  $U$  depending on the treatment and outcome status (see Ichino et al. 2008, 317)

$$\begin{aligned} p_{ij} &\equiv P(U = 1 \mid T = i, Y = j, X) \\ &= P(U = 1 \mid T = i, Y = j) \\ \text{with } & i, j \in \{0, 1\} \end{aligned}$$

Hence,  $p_{ij}$  defines the probability that  $P(U = 1)$  when  $T = i$  and  $Y = j$ . Following Ichino et al (2008), we focus our sensitivity analysis on two statistics based on these four parameters  $p_{ij}$  that reflect different assumptions about the unobserved confounding variable. In particular, “the real threat to the baseline estimate is coming from a potential confounder that has both a positive effect on the untreated outcome ( $p_{01} - p_{00} > 0$ ), [hereafter, “ $d$ ”] and on the selection into treatment ( $p_{1.} - p_{0.} > 0$ ) [hereafter, “ $s$ ”]” (Ichino et al. 2008, p. 318).

<sup>19</sup>The marginal association of  $U$  and  $X$  will be nonzero in the sample because of the association between  $U$ ,  $T$ , and  $Y$  along with the association between  $X$ ,  $T$ , and  $Y$ .

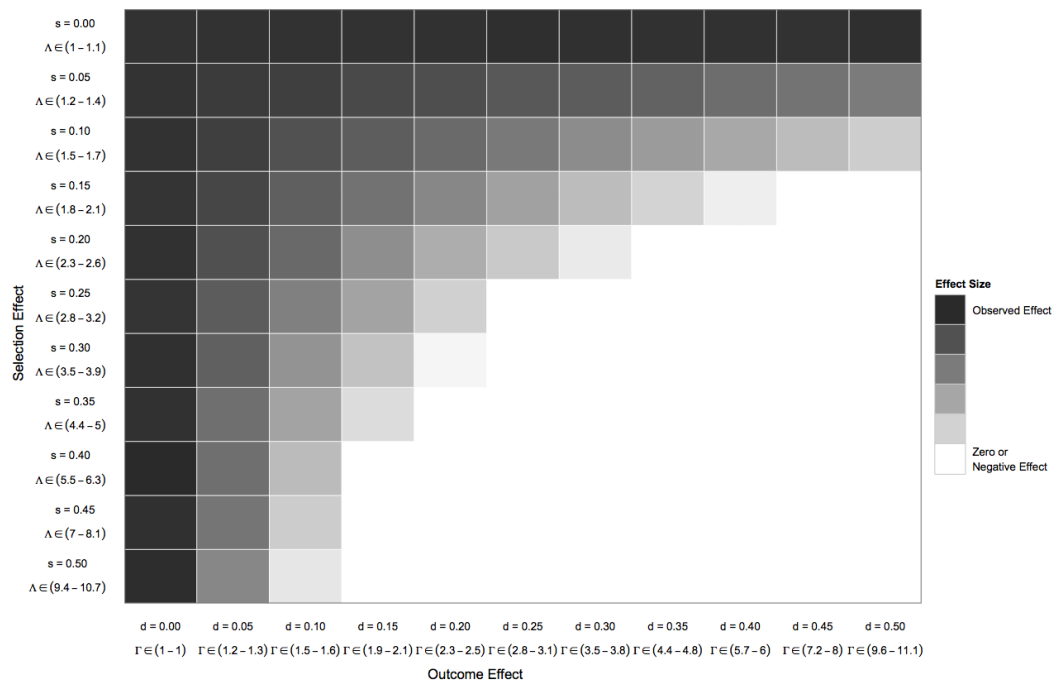
The two statistics,  $d$  and  $s$ , together with the marginal probability  $P(U = 1)$  and the difference  $p_{11} - p_{10}$  determine the four values of  $p_{ij}$ . Accordingly, fixing two secondary statistics  $P(U = 1) = 0.4$  and setting  $p_{11} - p_{10} = 0$  allows us to simulate  $U$  for each observation in our dataset using random draws from a binomial distribution with  $p_{ij}$  as the probability parameter so that  $U \sim \text{Binominal}(p_{ij})$ .<sup>20</sup> After simulating  $U$ , we reestimate Model III presented in Table 3.3 with the additional control variable  $U$  and confined to the female respondents. We then compare the observed effect from Model X with the one obtained with the additional simulated confounder  $U$ . Changing the parameter  $d$  and  $s$  in this simulation and comparing the obtained effects helps us to understand how robust the estimated effect is to additional unobserved covariates. The parameter  $d$  is associated with, but not the same as the effect of  $U$  on the untreated  $Y$ , and the parameter  $s$  is associated with, but not the same as the effect of  $U$  on  $s$ . For each  $d$  and  $s$ , we can compute the average odds ratio in the data of the effect of  $U$  on  $Y$ , conditional on  $X$  (hereafter,  $\Gamma$ ), and also the average odds ratio of  $U$  on  $T$ , conditional on  $X$  (hereafter,  $\Lambda$ ). We can thereby produce unobserved confounding variables that have effects similar to those of observed covariates. For further details about this sensitivity analysis, we refer the reader to Ichino et al. (2008) and Nannicini (2007).

Figure A3.8 shows the results of our simulation-based sensitivity analysis for values of  $d$  and  $s$  that range from 0 to 0.5 and reflect different relations between the simulated cofounder  $U$  and the treatment and outcome variable. To restate, our treatment variable is the index value of the number of AP math and science courses in the high school, and our outcome variable is the STEM orientation in 12th grade. As indicated by the  $\Gamma$  and  $\Lambda$  values in the figure, these values of  $d$  and  $s$  correspond to an odds-ratio effect of  $U$  on  $Y$  of between 1.1 and 11.1, and of between 1 and 10.7 on the treatment indicator after con-

---

<sup>20</sup>The results are consistent across different values for  $P(U = 1)$  and  $p_{11} - p_{10}$ .

Figure A3.8: Sensitivity of Estimate to Additional Confounding Variable



*Note: For each combination of  $d$  and  $s$ , we conducted 100 simulation runs so that the whole graph is based on 12,100 simulation. A Stata implementation of the simulation-based sensitivity analysis for matching procedures is available from Nannicini (2007). Our R implementation for both matching procedures and regression methods together with the graphical presentation of the results is available from the first author of the paper.*

ditioning on all the covariates used in the main regression. The shading of each square indicates how the estimated effect size changes depending on the two parameters  $d$  and  $s$  with black indicating the observed effect size of  $T$  in the sample, and white indicating a zero or negative effect. For most of the observed pre-treatment covariates, the odds ratio for the outcome effect (conditional on other covariates) is between 0.8 and 1.2, while some have slightly higher values. The estimated effect of a specific covariate  $x$  on  $T$  conditional on the other observed covariates (the selection effect of  $x$ ), is generally smaller than is the estimated outcome effect. Accordingly, most estimated covariates conditional on observable variables are equivalent to a confounding variable that would

be located in the four squares in the top-left corner of Figure A3.8, where  $d$  and  $s$  are both in the range  $[0, 0.05]$ . Confounding variables that had a similar strength relationship with  $Y$  and  $T$  as do nearly all of the observed covariates would fall in the slightly wider range of the upper left nine squares where both  $d$  and  $s$  are in the range  $[0, 0.10]$ . For these values of  $d$  and  $s$ , the estimated effect with the simulated confounder  $U$  is still substantial. For example, 8th grade math performance, which is one of the most important control variables, has an odds-ratio outcome effect of 1.06 and a selection effect of 1.212 and therefore lies in the square region defined by  $d \in [0, 0.05]$  and  $s = 0.05$ . As the diagram shows, confounding variables as powerful as 8th grade math performance still leave a substantial portion of the positive curriculum effect on STEM orientations intact. Accordingly, our estimates are relatively robust to an additional cofounder that is similar to the currently used control variables and unrelated to any of the covariates in the current model.<sup>21</sup>

### **The Effect of High School Curriculum on the Gender Gap in STEM Orientation for the Persistence Pathway**

In Table 3.3, we reported the effect of the high school curriculum on the gender gap in STEM orientation for the late entry pathway. Table A3.6 extends these results and shows the corresponding estimates for the persistence pathway, i.e. students who reported an orientation towards STEM fields in 8th grade. The results show a highly significant difference in the persistence rate between boys and girls, which confirms the pattern documented in the pathway analysis (Figure 3.2). Similar to the estimates for the late entry pathway (see Table 3.3), the point estimate for the female curriculum effect is positive

---

<sup>21</sup>Note that the assumption that the confounding variable is unrelated to the pre-treatment control variables in the current model produces a conservative sensitivity analysis considering that we use a large set of variables that are directly related to the selection process and highly relevant for the outcome.

and substantial whereas the effect for boys is clearly smaller as indicated by the negative interaction effect (note that the point estimate for the model without additional control variables is smaller for the persistence analysis but slightly larger in Model IV with the full set of control variables). The estimated effects for the persistence pathway, however, are far from statistically significant so that we are unable to determine whether the high school curriculum also influences the persistence rates. At the same time, these results also do not support the conclusion that the high school curriculum has no effect on the persistence rate (only point estimates that are close to zero with small confidence intervals would support such a conclusion). The estimated effects are similar in size to the effects for late entry but the uncertainty around these estimates is simply too large to draw a meaningful conclusion. This difference between the two analyses is mainly related to the much smaller sample size for the persistence pathway in comparison to the high school entry pathway (792 compared to 10,478).

## **Appendix D Analysis for the Section 'Is the High School Effect Lasting and How Big is the Effect?'**

In the section 'Is the High School Effect Lasting and How Big is the Effect?', we reported results about the post-high school transitions rates for students from different high schools as well as for the reduction in the gender gap in STEM BAs if all high schools would encourage women to study science and engineering. In order to conduct this assessment, we group high schools by the size of the gender gap in science and engineering orientation. In particular, we use the empirical Bayes estimates of the gender gap from the "value-added" multilevel

Table A3.6: Logistic Regression Estimates for the Persistence Pathway of STEM Orientation in 12th Grade on High School Math and Science Curriculum

	Model I		Model III		Model IV	
	Coef	(se)	Coef	(se)	Coef	(se)
Intercept	-1.008***	(0.16)	-2.319***	(0.64)	0.815	(2.14)
Male	0.487**	(0.18)	0.482*	(0.21)	0.436*	(0.22)
Curriculum Index (CI)	0.156	(0.16)	0.170	(0.20)	0.179	(0.21)
Curriculum Index (CI) x Male	-0.093	(0.18)	-0.200	(0.21)	-0.145	(0.22)
<i>Pre-High School Control Variables</i>						
Std. Demographic Variables			✓		✓	
Urban/Region Variables			✓		✓	
8th Grade Variables			✓		✓	
<i>High School Control Variables</i>					✓	
Students	792		792		792	

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ ; Robust standard errors in parentheses.

Note: Control variables are described in Appendix Table A3.4. The estimates presented in this table are only for the persistence pathway - i.e. the students who indicated a STEM orientation in 8th grade. Note that we omitted Model II with only standard demographic and urban/region control variables from the table.

model (NELS 88-92) above to group schools into those with a small gender gap (bottom terciles) and those with a big gender gap (top terciles). We then match this newly created school-level variable to the students in NELS 88-2000. Table A3.7 reports the post-high school transition rates used in the pathway analysis for the full school sample (already shown in the graphs above) as well as the high schools with a small and big gender gap. The results show that the post-high school transition rates are remarkably constant across the three samples indicating that high schools have a lasting effect on gender differences.

For the second part of the analysis, we again group high schools into terciles according to the size of their gender gap in STEM orientation. We then calculate the gender gap in STEM BA degrees assuming the same eighth grade orientation and post-high school transition rates across all three samples. In other words, we assume that differences in the gender gap across the three samples



Table A3.7: Post-High School Transition Rates for Full Sample, and School with Small/Big Gender Gap

Post-HS Transition Rates	Gender	Full Sample	Schools with Small Gender Gap	Schools with Big Gender Gap
Leakage Rate	male	0.669	0.615	0.692
	female	0.649	0.686	0.637
Late Entry Rate	male	0.078	0.084	0.082
	female	0.051	0.06	0.039
Persistence Rate	male	0.331	0.385	0.308
	female	0.351	0.314	0.363

*Note: Late Entry refers to late entry from a college but non-STEM orientation at the end of high school.*

Table A3.8: Gender Gap in STEM BAs for Full Sample, Schools with Small Gender Gap, and Schools with Big Gender Gap

Proportion of Students with STEM bachelor degree	Male	Female	Gender Gap	
			in %	Odds Ratio
Full Sample	0.098	0.060	0.038	1.713
Schools with Small Gender Gap	0.124	0.095	0.029	1.349
Schools with Big Gender Gap	0.077	0.039	0.039	2.089

only emerge because of differences in the transition rates within high school, and not from group differences in eighth grade orientation and transition rates after high school. As shown in Table A3.8, boys are 1.7 times as likely as girls to graduate from college with a STEM BA degree in the entire sample. However, this substantial male advantage is reduced to 1.3 (male/female odds ratio) in the sub-sample of students who attend high schools with a small gender gap. Accordingly, the gender gap would be reduced by about 25% if the environment in all schools would encourage girls to study science and engineering at the same rates as the top third of schools.



# Conclusion

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”. At the same time, women still pursue science, technology, engineering, and math degrees at much lower rates than their male peers do. Some point at deeply entrenched possibly innate gender differences as an explanation for the gender gap in educational outcomes but others document different ways in which gender is constructed and performed in school. This dissertation examines the role of the school context and the connection between school resources and the gender differences in educational outcomes in three papers that each highlight a different aspect of this broader question. The first article “School Context and the Gender Gap in Educational Achievement” builds on theories about gender identity and reports from prior ethnographic classroom observations to argue that schools and peers in school play an important role for gender differences in education. 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. To evaluate this argument, the paper uses a quasi-experimental research design that estimates the gender difference in the causal effect on test scores focusing on peer SES as an important school resource. The findings from this quasi-experimental research design show that boys benefit more strongly from peers with higher socioeconomic background. Additional analysis support the argument that the gender difference in the causal effect can be explained by differences in the social support for academic work in the male and female peer culture. A second quasi-experimental case study based on a different setting supports the main results and thereby reaffirms the conclusions. The findings, however, seem to be at odds with recent research on the effect of policy interventions that aim to place families in low-poverty neighborhoods or move students to better schools, which report larger benefits for girls than boys. To address this seeming contradiction and develop a better understanding of gender differences in the effect of peers, the second paper "Disruptive Change: Peer Effects and the Social Adjustment Process of Mobile Students" integrates the literature on student mobility and peer effects. I argue that changing school is connected with a social adjustment process that has important implications for the influence of peers and helps us to understand the contradicting findings in the literature on gender differences in context and policy effects. Using a large-scale administrative dataset and a quasi-experimental research design based on a difference-in-difference, matching approach, I find that the effect of peers is substantially smaller for mobile students with a clear temporal adjustment process that is more pronounced for boys. As a consequence, boys who transfer to a school with higher achieving peers initially learn at the same rate as their peers in the previous school despite the improved learning environment. Only after several years, they begin to experience the benefits associated with higher performing peers. The final and third paper "High School Envi-

ronments, STEM Orientations, and the Gender Gap in Science and Engineering Degrees” shifts the focus from educational performance to persisting gender differences in field of study. Using the National Education Longitudinal Study (NELS), I first evaluate whether the high school years play an important role in forming orientations towards science, technology, engineering, and math fields that are consequential for the later gender gap in STEM BA degrees and then determine that the local environment in school shapes the gender-specific formation of career aspirations.

Together, these three articles make important contributions to our understanding of gender differences in educational outcomes, and suggest concrete policy implications about the educational shortcomings of boys, and the persisting gender gap in STEM degrees. They show that peer effects are larger for boys than girls and that this gender difference can be explained by differences in the social support for academic work in the male and female peer culture. These findings shift the focus from masculinity as inherently based on resistance to school towards the importance of the local school environment for the construction of gender identities as well as school-related attitudes, behavior, and the performance of boys and girls. My findings also point to the high school years as the life course period that should be targeted to increase the number of women with STEM BAs, and provide evidence that high school interventions might be effective to achieve that goal. Overall, they contribute to several areas of research with important implications for future directions of research on policies.

First, the three papers make a critical contribution to the debate about the well-publicized under-performance of boys and the persisting gender difference in field of study. The outlined mechanisms highlight the significance of schools and peers in school, which challenges the focus on deeply entrenched, possibly innate gender differences. Instead, they emphasize the pertinent role

of the local cultural environment. This argument suggests that boys' resistance to school and girls' relative disinterest in science fields is not purely a consequence of entrenched possibly innate gender differences or universal gender stereotypes but instead depends on the local cultural environment in school. For educational performance, the findings 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. In terms of interest in science, technology, engineering, and math fields, however, the third article presents evidence that *girls* are more sensitive to certain school resources. While these findings in terms of educational performance and field of study might initially contradict each other, a broader theoretical argument reconciles the different results: supportive peers or more generally a supportive school environment are particularly beneficial for the disadvantaged group - boys in the case of work habits and educational performance, and girls in the case of STEM interests. This understanding of gender differences in peer effects is in line with previous psychological studies, which document that the influence of peers is stronger for certain outcomes on boys and for other outcomes on girls (Brechwald and Prinstein 2011: 172). Finally, our findings about the pertinent interaction between exposure and mobility effects in the second article speak to the recent debate on gender differences in context and policy effects (Kling et al. 2005; Clampet-Lundquist et al. 2011; Legewie and DiPrete 2012). A number of studies indicate that girls but not boys benefit from moving to a higher-resource environment (Hastings et al. 2006; Clampet-Lundquist et al. 2011; Kling et al. 2005) while other research suggests that boys are more sensitive to peers in their context (Legewie and DiPrete 2012; Carrell and Hoekstra 2010). Our findings contribute to this debate by showing that the temporal adjustment process is particularly pronounced for boys. Accordingly, boys might

well be more sensitive to peer effects but after changing school they have more problems adjusting to the new environment, which temporarily reduces the effect of peers and explains the contradictory findings in previous research.

Second, our results point to useful directions for new research on policies to raise boys' achievement levels and reduce the persisting gender differences in field of study. It is obviously important to know that school resources affect the educational performance of boys and girls and how this influence depends on exposure dynamics connected to student mobility. This knowledge has direct implications for existing policy interventions that aim to place families in low-poverty neighborhoods or move students to better schools. Context effects do play an important role for all students and particularly for boys but at the same time moving students across schools moderates the benefits and as such might still not provide a tangible solution. Accordingly, an important goal and a potential direction for future research should be to understand the conditions that alleviate the consequences of moving and aid the adjustment process for students that are given the chance to change to higher quality schools. Haynie and South's (2005) work on the effect of residential mobility on adolescent violence is the only paper to my knowledge that addresses this question focusing on the parent-child relationships, psychological distress, experiences of victimization, and peer networks as possible mediators. An alternative would be to focus resources on improving neighborhood schools that provide benefits for all students by increasing the quality of schools. But the theoretical focus on peer processes among boys and girls also points at another possible direction for future research that informs concrete policy interventions. For this direction, the key unanswered question raised by our research is 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. My research suggests, for example, that teaching methods that emphasize academic competition are particularly beneficial for boys. More precisely, the peer processes at the core of my theoretical argument hint at a group-based reward structure that has the potential to channel the value and reward system among students (and particularly boys) towards academics and therefore foster an learning oriented peer culture (for an early formulation of such an argument see Spilerman 1971).

In terms of gender difference in field of study, the pathway analysis shows that high school is the decisive life period during which the gender gap emerges, and the examination of variations across contexts shows that the local context in high school plays a critical role for the gender gap in orientations towards STEM fields. As such, my findings not only point at the life course period that should be targeted by policy interventions, but also provide evidence that high school interventions might be effective. Not all local environmental effects are necessarily durable, though. In light of recent research asserting only a temporary effect from exposure to Head Start programs or to individual above-average teachers (Jacob et al. 2010), it is of considerable importance that the effects of the high school environment on the formation of STEM orientations appear to be durable. Some existing interventions have indeed targeted high school students and shown success in promoting a STEM orientation among girls. Eisenhart (2008), for example, discusses a seemingly effective outreach project that educates high-achieving, minority girls in high school about science and engineering jobs. While such policy interventions have to withstand the serious scrutiny of experimental field trials, the evidence presented in the



third paper encourages researchers and policy makers alike to take seriously the potential impact of high school interventions on the STEM orientations of female students. Our finding that the intensity of the math and science curriculum reduces the gender gap in science orientation strongly supports this conclusion.

Finally, this dissertation makes a methodological contribution to the literature on the estimation of causal effects. The approach to causality in the two quasi-experimental case studies differs from the predominant focus on matching as a technique to condition on a set of observable covariates. Instead, my case studies identify some source of variation in the treatment indicator that is as good as random and then provide strong institutional, and empirical evidence based on statistical simulations, and qualitative interviews to support this claim. Accordingly, article 1 illustrates how a detailed study of the relevant selection process – in the Berlin case, the 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 the researcher to evaluate the extent of bias but also enables the design of targeted sensitivity analysis (in the Berlin case based on instrumental variables and sample restrictions). Overall, this knowledge about the selection process can help researchers improve the accuracy of causal effect estimates such as in this case for compositional peer effects in school. Considering these benefits, my dissertation invites sociologists to rethink their predominant focus on matching and to take selection processes seriously as an independent object of study – an argument previously made by Sampson (2008) who conceptualizes “selection bias as a fundamental social process worthy of study in its own right rather than a statistical nuisance”.



# Bibliography

- Adelman, Clifford. 2006. "The Toolbox Revisited. Paths to Degree Completion from High School Through College." Technical report, U.S. Department of Education.
- Alexander, Karl L., Doris R. Entwisle, and Susan L. Dauber. 1996. "Children in Motion: School Transfers and Elementary School Performance." *The Journal of Educational Research* 90:3–12.
- Almquist, E.M., S.S. Angrist, and R. Mickelsen. 1980. "Women's Career Aspirations and Achievements: College and Seven Years After." *Work and Occupations* 7:367–384.
- 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* 27:315–348.
- an Ghail, Mairtin Mac. 1994. *The Making of Men: Masculinities, Sexualities and Schooling*. Buckingham: Open University Press.
- Andreescu, T., J. A Gallian, J. M Kane, and J. E Mertz. 2008. "Cross-cultural analysis of students with exceptional talent in mathematical problem solving." *Notices of the AMS* 55:1248–1260.
- Angrist, Joshua D. and Kevin Lang. 2004. "Does School Integration Generate Peer Effects? Evidence from Boston's Metco Program." *The American Economic Review* 94:1613–1634.
- Angrist, Joshua D. and Jörn-Steffen Pischke. 2008. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton: Princeton University Press.
- Angrist, S.S. and E.M. Almquist. 1993. "The Carnegie Mellon Class of 1968: Families, careers, and contingencies." In *Women's Lives through Time: Educated American Women of the Twentieth Century*, edited by K.D. Hulbert and D.T. Schuster. San Francisco: Jossey-Bass.

- Arnold, K.D. 1995. *Lives of Promise: What becomes of High School Valedictorians*. San Francisco: Jossey-Bass.
- Barry, Carolyn Mcnamara and Kathryn R. Wentzel. 2006. "Friend Influence on Prosocial Behavior: The Role of Motivational Factors and Friendship Characteristics." *Developmental Psychology* 42:153–163.
- Beutel, A.M. and M. Mooney Marini. 1995. "Gender and values." *American sociological review* 60:436–448.
- 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* 6:141–199.
- Brechwald, Whitney A. and Mitchell J. Prinstein. 2011. "Beyond Homophily: A Decade of Advances in Understanding Peer Influence Processes." *Journal of Research on Adolescence* 21:166–179.
- 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* 71:515–541.
- 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.
- Carrell, Scott E., Richard L. Fullerton, and James E. West. 2009. "Does your cohort matter? Measuring peer effects in college achievement." *Journal of Labor Economics* 27:439–464.
- Carrell, Scott E. and Mark L. Hoekstra. 2010. "Externalities in the Classroom: How Children Exposed to Domestic Violence Affect Everyone's Kids." *American Economic Journal: Applied Economics* 2:211–228.
- Carrell, S. E, M. E Page, and J. E West. 2010. "Sex and Science: How Professor Gender Perpetuates the Gender Gap." *Quarterly Journal of Economics* 125:1101–1144.
- Carter, Prudence L. 2003. "'Black' Cultural Capital, Status Positioning, and Schooling Conflicts for Low-Income African American Youth." *Social Problems* 50:136–155.

- Ceci, S. J., W. M. Williams, and S. M. Barnett. 2009. "Women's underrepresentation in science: Sociocultural and biological considerations." *Psychological Bulletin* 135:218–261.
- Charles, M. and K. Bradley. 2002. "Equal but separate? A cross-national study of sex segregation in higher education." *American Sociological Review* 67:573–599.
- Chudgar, Amita and Thomas F. Lushei. 2009. "National Income, Income Inequality, and the Importance of Schools: A Hierarchical Cross-National Comparison." *American Educational Research Journal* 46:626–658.
- Clampet-Lundquist, Susan, Kathryn Edin, Jeffrey R. Kling, and Greg J. Duncan. 2011. "Moving Teenagers Out of High-Risk Neighborhoods: How Girls Fare Better than Boys." *American Journal of Sociology* 116:1154–89.
- Cohen, M. 1998. "A habit of healthy idleness: Boys' underachievement in historical perspective." In *Failing Boys?*, edited by A. D. Epstein, J. Elwood, V. Hey, and J. Maw, pp. 19–34. Buckingham, UK: Open University Press.
- Coleman, James. 1966. *Equality of Educational Opportunity*. U.S. Dept. of Health, Education, and Welfare, Office of Education.
- Coleman, James S. 1960. "The Adolescent Subculture and Academic Achievement." *American Journal of Sociology* 65:337–347.
- 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. 1970. "Reply to Cain and Watts." *American Sociological Review* 35:242–249.
- Cook, Thomas D, Peter M. Steiner, and Steffi Pohl. 2009. "How Bias Reduction Is Affected by Covariate Choice, Unreliability, and Mode of Data Analysis: Results From Two Types of Within-Study Comparisons." *Multivariate Behavioral Research* 44:828–847.
- Correll, Shelley J. 2001. "Gender and the Career Choice Process: The Role of Biased Self-Assessments." *American Journal of Sociology* 106:1691–1730.
- Correll, Shelley J. 2004. "Constraints into Preferences: Gender, Status, and Emerging Career Aspirations." *American Sociological Review* 69:93–113.
- Coulson, Margaret A, E. Teresa Keil, David S Riddell, and John S Struthers. 1967. "Towards a Sociological Theory of Occupational Choice — a Critique." *The Sociological Review* 15:301–309.

- Crane, Jonathan. 1991. "The epidemic theory of ghettos and neighborhood effects on dropping out and teenage childbearing." *American Journal of Sociology* 96:1226–1259.
- Crosnoe, Robert. 2009. "Low-Income Students and the Socioeconomic Composition of Public High Schools." *American Sociological Review* 74:709–730.
- Crowder, Kyle and Scott J. South. 2011. "Spatial and temporal dimensions of neighborhood effects on high school graduation." *Social Science Research* 40:87–106.
- Csikszentmihalyi, Mihaly and Barbara Schneider. 2001. *Becoming Adult: How Teenagers Prepare For The World Of Work*. Basic Books.
- Cullen, Julie Berry, Brian A. Jacob, and Steven Levitt. 2006. "The Effect of School Choice on Participants: Evidence from Randomized Lotteries." *Econometrica* 74:1191–1230.
- Cullen, Julie Berry, Brian A. Jacob, and Steven D. Levitt. 2005. "The impact of school choice on student outcomes: an analysis of the Chicago Public Schools." *Journal of Public Economics* 89:729–760.
- Cvencek, Dario, Andrew N Meltzoff, and Anthony G Greenwald. 2011. "Math–Gender Stereotypes in Elementary School Children." *Child Development* 82:766–779.
- 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.
- Dalton, Ben, Steven J Ingels, Jane Downing, and Robert Bozick. 2007. *Advanced Mathematics and Science Course-taking in the Spring High School Senior Classes of 1982, 1992 and 2004*. Washington: U.S. Department of Education.
- Davies, Bronwyn. 2003. *Frogs and Snails and Feminist Tales: Preschool Children and Gender*. Hampton Pr.
- Davies, S. and N. Guppy. 1996. "Fields of Study, College Selectivity, and Student Inequalities in Higher Education." *Social Forces* 75:1439–1462.
- Deluca, Stefanie and Elizabeth Dayton. 2009. "Switching Social Contexts: The Effects of Housing Mobility and School Choice Programs on Youth Outcomes." *Annual Review of Sociology* 35:457–491.
- DiPrete, T.A. and M. Gangl. 2004. "Assessing bias in the estimation of causal effects: Rosenbaum bounds on matching estimators and instrumental variables estimation with imperfect instruments." *Sociological Methodology* 34:271–310.

- 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* 22:415–440.
- Dresel, Markus, Heidrun Stöger, and Albert Ziegler. 2006. "Klassen- und Schuleffekte bei Geschlechtsunterschieden im schulischen Leistungsbereich." *Psychologie in Erziehung und Unterricht* 53:44–61.
- Duflo, Esther, Pascaline Dupas, and Michael Kremer. 2011. "Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya." *American Economic Review* 101:1739–1774.
- Duxbury, L.E. and C.A. Higgins. 1991. "Gender differences in work-family conflict." *Journal of Applied Psychology* 76:60–74.
- Eccles, J.S. 2007. "Where are all the women? Gender differences in participation in physical science and engineering." In *Why Aren't More Women in Science? Top Researchers Debate the Evidence*, edited by S. J. Ceci and W. M. Williams. Washington, D.C.: American Psychological Association.
- Eccles, JS and LW Hoffman. 1984. "Sex roles, socialization, and occupational behavior." *Child development research and social policy* 1:367–420.
- Eccles, J. S. 1994. "Understanding women's educational and occupational choices." *Psychology of Women Quarterly* 18:585–609.
- Eisenhart, M. 2008. "We can't get there from here: High school girls consider engineering."
- Eitzen, D. Stanley. 1975. "Athletics in the Status System of Male Adolescents: A Replication of Coleman's 'The Adolescent Society'." *Adolescence* 10:267–276.
- Ellison, G. and A. Swanson. 2010. "The Gender Gap in Secondary School Mathematics at High Achievement Levels: Evidence from the American Mathematics Competitions." *The Journal of Economic Perspectives* 24:109–128.
- England, P. 2005. "Gender inequality in labor markets: The role of motherhood and segregation." *Social Politics: International Studies in Gender, State & Society* 12:264–288.
- Epple, Dennis and Richard E. Romano. 2011. "Peer Effects in Education: A Survey of the Theory and Evidence." In *Handbook of Social Economics*, edited by Jess Benhabib, Alberto Bisin, and Matthew O. Jackson, pp. 1053–1163. Amsterdam: North-Holland.

- Epstein, Debbie. 1998. "Real Boys Don't Work: Underachievement, Masculinity and the Harassment of Sissies." In *Failing boys? Issues in gender and achievement*, edited by Debbie Epstein, Jannette Elwood, Valerie Hey, and Janet Maw, pp. 96–108. Buckingham: Open University Press.
- Felmlee, D.H. 1993. "The Dynamic Interdependence of Women's Employment and Fertility." *Social Science Research* 22:333–360.
- Francis, Becky. 2000. *Boys, Girls and Achievement: Addressing the Classroom Issues*. Routledge Falmer.
- Frank, Kenneth. 2000. "Impact of a Confounding Variable on the Inference of a Regression Coefficient." *Sociological Methods & Research* 27:147–194.
- Frank, Kenneth A., Chandra Muller, Kathryn S. Schiller, Catherine Riegler-Crumb, Anna Strassmann Mueller, Robert Crosnoe, and Jennifer Pearson. 2008. "The Social Dynamics of Mathematics Coursetaking in High School." *American Journal of Sociology* 113:1645–1696.
- 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.
- Gelman, Andrew and Jennifer Hill. 2007. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge: Cambridge University Press.
- Gerson, K. 1985. *Hard Choices: How Women Decide about Work, Career, and Motherhood*. University of California Press.
- Ginzberg, Eli, Sol W. Ginsburg, Sidney Axelrad, and Herma John L. 1951. *Occupational Choice*. New York: Columbia University Press.
- Goldsmith, Pat Rubio. 2011. "Coleman Revisited: School Segregation, Peers, and Frog Ponds." *American Educational Research Journal* 48:508–535.
- Golombok, Susan, John Rust, Karyofyllis Zervoulis, Tim Croudace, Jean Golding, and Melissa Hines. 2008. "Developmental Trajectories of Sex-Typed Behavior in Boys and Girls: A Longitudinal General Population Study of Children Aged 2.5–8 Years." *Child Development* 79:1583–1593.
- 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* 119:1243–1269.
- Greenwald, Rob, Larry V Hedges, and Richard D Laine. 1996. "The Effect of School Resources on Student Achievement." *Review of Educational Research* 66:361–396.



- Grigg, Jeffrey. 2012. "School Enrollment Changes and Student Achievement Growth: A Case Study in Educational Disruption and Continuity." *Sociology of Education* 85:388–404.
- Guiso, Luigi, Ferdinando Monte, Paola Sapienza, and Luigi Zingales. 2008. "Culture, Gender, and Math." *Science* 320:1164–1165.
- Habashi, M. M, W. G Graziano, D. Evangelou, and I. Ngambeki. 2009. "Teacher Influences on Child Interest in STEM Careers." In *Proceedings of the Research in Engineering Education Symposium*, Palm Cove, QLD.
- Hagan, John, Ross Macmillan, and Blair Wheaton. 1996. "New Kid in Town: Social Capital and the Life Course Effects of Family Migration on Children." *American Sociological Review* 61:368–385.
- Hallinan, M. T. 1988. "Equality of Educational Opportunity." *Annual Review of Sociology* 14:249–268.
- Halpern, D. F. 2000. *Sex Differences in Cognitive Abilities*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Hanushek, Eric A. 1989. "The impact of differential expenditures on school performance." *Educational researcher* 18:45–62.
- Hanushek, Eric A., John F. Kain, and Steven G. Rivkin. 2004. "Disruption versus Tiebout improvement: the costs and benefits of switching schools." *Journal of Public Economics* 88:1721–1746.
- Hanushek, Eric A. and Steven G. Rivkin. 2009. "Harming the best: How schools affect the black-white achievement gap." *Journal of Policy Analysis and Management* 28:366–393.
- Hastings, J. S., T. J. Kane, and D. O. Staiger. 2006. "Gender and performance: Evidence from school assignment by randomized lottery." *American Economic Review* 96:232–236.
- Haynie, Dana L. 2001. "Delinquent Peers Revisited: Does Network Structure Matter?" *American Journal of Sociology* 106:1013–1057.
- Haynie, Dana L. and Scott J. South. 2005. "Residential Mobility and Adolescent Violence." *Social Forces* 84:361–374.
- Hedges, LV and A. Nowell. 1995. "Sex differences in mental test scores, variability, and numbers of high-scoring individuals." *Science* 269:41–45.
- Heilbron, Nicole and Mitchell J. Prinstein. 2008. "Peer influence and adolescent nonsuicidal self-injury: A theoretical review of mechanisms and moderators." *Applied and Preventive Psychology* 12:169–177.

- Hill, Catherine, Christianne Corbett, and Andresse St. Rose. 2010. *Why So Few? Women in Science, Technology, Engineering, and Mathematics*. Washington, DC: American Association of University Women.
- Hoffman, Moshe, Uri Gneezy, and John A. List. 2011. "Nurture affects gender differences in spatial abilities." *Proceedings of the National Academy of Sciences* 36:14786–14788.
- Hoffmann, John P. and Robert A. Johnson. 1998. "A National Portrait of Family Structure and Adolescent Drug Use." *Journal of Marriage and Family* 60:633–645.
- Horn, Laura and Anne-Marie Nuñez. 2000. "Mapping the Road to College: First-Generation Students' Math Track, Planning Strategies, and Context of Support." NCES rep. no. 2000-153, National Center for Educational Statistics, Washington, DC.
- Hossler, Don and Frances K. Stage. 1992. "Family and High School Experience Influences on the Postsecondary Educational Plans of Ninth-Grade Students." *American Educational Research Journal* 29:425–451.
- Howe, Christine. 1997. *Gender and Classroom Interaction: A Research Review*. The SCRE Centre.
- Howell, F. M, W. Frese, and C. R. Sollie. 1977. "Ginzberg's theory of occupational choice: A reanalysis of increasing realism." *Journal of Vocational Behavior* 11:332–346.
- Hoxby, Caroline. 2000. "Peer Effects in the Classroom: Learning from Gender and Race Variation." NBER Working Paper 7867.
- Hoxby, Caroline M. and Gretchen Weingarth. 2005. "Taking Race Out of the Equation. School Reassignment and the Structure of Peer Effects."
- Hyde, J.S., S.M. Lindberg, Linn M.C., Ellis A.B., and Williams C.C. 2008. "Gender Similarities Characterize Math Performance." *Science* 321:494–495.
- Ichino, Andrea, Fabrizia Mealli, and Tommaso Nannicini. 2008. "From temporary help jobs to permanent employment: what can we learn from matching estimators and their sensitivity?" *Journal of Applied Econometrics* 23:305–327.
- Imbens, Guido W. 2004. "Nonparametric Estimation of Average Treatment Effects under Exogeneity: A Review." *The Review of Economics and Statistics* 86:4–29.
- Imberman, S., A. Kugler, and B. Sacerdote. 2012a. "Katrina's Children: A Natural Experiment in Peer Effects from Hurricane Evacuees." *American Economic Review* 102:2048–2082.

- Imberman, Scott, Adriana Kugler, and Bruce Sacerdote. 2012b. "Katrina's Children: A Natural Experiment in Peer Effects from Hurricane Evacuees." *American Economic Review* 102:2048–2082.
- Jackson, Margot I. and Robert D. Mare. 2007. "Cross-sectional and longitudinal measurements of neighborhood experience and their effects on children." *Social Science Research* 36:590–610.
- Jacob, B. A, L. Lefgren, and D. P Sims. 2010. "The persistence of teacher-induced learning." *Journal of Human Resources* 45:915–943.
- Jacobs, J. A. 1989. *Revolving doors: Sex segregation and women's careers*. Stanford Univ Pr.
- Jacobs, J. E, S. Lanza, D. W Osgood, J. S Eccles, and A. Wigfield. 2002. "Changes in children's self-competence and values: Gender and domain differences across grades one through twelve." *Child Development* 73:509–527.
- Jencks, C. and S. E. Mayer. 1990a. "The social consequences of growing up in a poor neighborhood." In *Inner-city poverty in the United States*, pp. 111–186. Washington, DC: National Academy Press.
- Jencks, Christopher and Susan E. Mayer. 1990b. "The social consequences of growing up in a poor neighborhood." In *Inner-city poverty in the United States*, edited by Laurence E. Lynn and Michael G. H. McGeary, p. 111–186. Washington, DC: National Academy Press.
- Jennings, J. L and T. A DiPrete. 2010. "Teacher Effects on Social and Behavioral Skills in Early Elementary School." *Sociology of Education* 83:135–159.
- Johnson, M.K. 2001. "Social Origins, Adolescent Experiences, and Work Value Trajectories during the Transition to Adulthood." *Social Forces* 80:1307–1340.
- 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* 35:477–496.
- Kahlenberg, Richard D. 2001. *All Together Now: Creating Middle-Class Schools through Public School Choice*. Washington, D.C.: Brookings Institution Press.
- Kane, T. J and D. O Staiger. 2008. *Estimating teacher impacts on student achievement: An experimental evaluation*. National Bureau of Economic Research Cambridge, Mass., USA.
- 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.

- Kerbow, David. 1996. "Patterns of urban student mobility and local school reform." *Journal of Education for Students Placed at Risk* 1:147–169.
- Kerbow, David, Carlos Azcoitia, and Barbara Buell. 2003a. "Student Mobility and Local School Improvement in Chicago." *The Journal of Negro Education* 72:158–164.
- Kerbow, David, Carlos Azcoitia, and Barbara Buell. 2003b. "Student Mobility and Local School Improvement in Chicago." *The Journal of Negro Education* 72:158–164.
- Kimura, D. 2002. "Sex hormones influence human cognitive pattern." *Neuroendocrinology Letters* 23:67–77.
- Kling, J. R., J. B. Liebman, and L. F. Katz. 2007. "Experimental analysis of neighborhood effects." *Econometrica* 75:83–119.
- Kling, Jeffrey R., Jens Ludwig, and Lawrence F. Katz. 2005. "Neighborhood Effects on Crime for Female and Male Youth: Evidence from a Randomized Housing Voucher Experiment." *The Quarterly Journal of Economics* 120:87–130.
- Konrad, A.M., J. E. Ritchie, P. Lieb, and E. Corrigall. 2000. "Sex differences and similarities in job attribute preferences: A meta-analysis." *Psychological bulletin* 126:593–641.
- Lapan, Richard T., Angela Adams, Sherri Turner, and Jeanne M. Hinkelman. 2000. "Seventh graders' vocational interest and efficacy expectation patterns." *Journal of Career Development* 26:215–229.
- Lauen, Douglas Lee and S. Michael Gaddis. 2012. "Exposure to Classroom Poverty and Test Score Achievement: Contextual Effects or Selection?" *American Journal of Sociology* (forthcoming).
- Lee, J., W. Grigg, and G. Dion. 2007. *The nation's report card: Mathematics 2007*. Washington, DC: National Center for Education Statistics, Institute of Education Sciences, US Department of Education.
- Lee, Valerie E., Robert F. Dedrick, and Julia B. Smith. 1991. "The Effect of the Social Organization of Schools on Teachers' Efficacy and Satisfaction." *Sociology of Education* 64:190–208.
- 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.
- Legewie, Joscha and Thomas A. DiPrete. 2012. "School Context and the Gender Gap in Educational Achievement." *American Sociological Review* 77:463–485.

- Lehmann, Rainer and Jenny Lenkeit. 2008. "ELEMENT. Abschlussbericht über die Untersuchungen 2003, 2004 und 2005 an Berliner Grundschulen und grundständigen Gymnasien." Technical report, Humboldt Universität zu Berlin, Berlin.
- Liben, L. S, R. S Bigler, and H. R Krogh. 2001. "Pink and Blue Collar Jobs: Children's Judgments of Job Status and Job Aspirations in Relation to Sex of Worker\* 1." *Journal of Experimental Child Psychology* 79:346–363.
- Lippa, Richard A. 2005. *Gender, Nature, and Nurture*. Psychology Press.
- Lohman, David F. and Joni M. Lakin. 2009. "Consistencies in sex differences on the cognitive abilities test across countries, grades, test forms, and cohorts." *British Journal of Educational Psychology* 79:389–407.
- Lubbers, Miranda J., Tom A. B. Snijders, and Margaretha P. C. Van Der Werf. 2011. "Dynamics of Peer Relationships Across the First Two Years of Junior High as a Function of Gender and Changes in Classroom Composition." *Journal of Research on Adolescence* 21:488–504.
- Lubinski, D. and C. P Benbow. 1992. "Gender Differences in Abilities and Preferences Among the Gifted: Implications for the Math-Science Pipeline." *Current Directions in Psychological Science* 1:61–66.
- Ludwig, Jens, Greg J. Duncan, Lisa A. Gennetian, Lawrence F. Katz, Ronald C. Kessler, Jeffrey R. Kling, and Lisa Sanbonmatsu. 2012. "Neighborhood Effects on the Long-Term Well-Being of Low-Income Adults." *Science* 337:1505–1510.
- Lueptow, L.B., L. Garovich-Szabo, and M.B. Lueptow. 2001. "Social Change and the Persistence of Sex Typing: 1974-1997." *Social Forces* 80:1–35.
- Mac an Ghail, Mairtin. 1994. *The Making of Men: Masculinities, Sexualities and Schooling*. Buckingham: Open University Press.
- Maccoby, E.E. 1998. *The two sexes: Growing up apart, coming together*. Cambridge, MA: Harvard University Press.
- Machin, Stephen and Sandra McNally. 2005. "Gender and student achievement in English schools." *Oxford Review of Economic Policy* 21:357–372.
- MacLeod, Jay. 2008. *Ain't No Makin' It: Aspirations and Attainment in a Low-Income Neighborhood*. Oxford: Westview Press.
- Martin, C. L, C. H Wood, and J. K Little. 1990. "The development of gender stereotype components." *Child Development* 61:1891–1904.

- McGee, J. and J. Stockard. 1991. "From a child's view: Children's occupational knowledge and perceptions of occupational characteristics." *Sociological studies of child development* 4:113–136.
- McNulty, William B and William A Borgen. 1988. "Career expectations and aspirations of adolescents." *Journal of Vocational Behavior* 33:217–224.
- Morris, E.W. 2008. "Rednecks, Rutters, and Rithmetic: Social Class, Masculinity, and Schooling in a Rural Context." *Gender & Society* 22:728–751.
- 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. 1983. "Quantitative Studies of Effective Schools: What Have We Learned?" In *School Finance and School Improvement: Linkages for the 1980*, edited by Allan Odden and L. Dean Webb, pp. 193–209. Cambridge: Ballinger.
- Nannicini, Tommaso. 2007. "Simulation-based sensitivity analysis for matching estimators." *Stata Journal* 7:334–350.
- Nechyba, Thomas J. 2000. "Mobility, Targeting, and Private-School Vouchers." *The American Economic Review* 90:130–146.
- OECD. 2007. *Education at a Glance 2007*. Paris: OECD.
- Okamoto, D. and P. England. 1999. "Is There a Supply Side to Occupational Sex Segregation?" *Sociological Perspectives* 42:557–582.
- Owings, Jeffrey. 1998. "High School Curriculum Structure: Effects on Course-taking and Achievement in Mathematics for High School Graduates." Technical report, US Dept. of Education, Office of Educational Research and Improvement.
- Pajares, F. 2005. "Gender Differences in Mathematics Self-Efficacy Beliefs." In *Gender differences in mathematics: An integrative psychological approach*, edited by A. M Gallagher and J. C Kaufman, pp. 294–315. Cambridge: Cambridge University Press.
- Park, Hyunjoon, Jere R Behrman, and Jaesung Choi. 2011. "Causal Effects of Single-Sex Schools on Students' STEM Outcomes by Gender and Parental SES." Unpublished working paper.
- Park, Robert E. 1928. "Human Migration and the Marginal Man." *American Journal of Sociology* 33:881–893.

- Penner, A.M. 2008. "Gender Differences in Extreme Mathematical Achievement: An International Perspective on Biological and Social Factors." *American Journal of Sociology* 114:138–170.
- 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.
- Plant, E. Ashby, A. L. Baylor, C. E. Doerr, and R. B. Rosenberg-Kima. 2009. "Changing middle-school students' attitudes and performance regarding engineering with computer-based social models." *Computers & Education* 53:209–215.
- Power, S., G. Whitty, T. Edwards, and V. Wigfall. 1998. "Schoolboys and schoolwork: gender identification and academic achievement." *International Journal of Inclusive Education* 2:135–153.
- Prinstein, Mitchell J. 2007. "Moderators of Peer Contagion: A Longitudinal Examination of Depression Socialization Between Adolescents and Their Best Friends." *Journal of Clinical Child & Adolescent Psychology* 36:159–170.
- 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* 22:728–751.
- Ravitch, Diane. 2010. *The death and life of the great American school system : how testing and choice are undermining education*. New York, NY: Basic Books.
- Ridgeway, C. L. 2001. "Gender, status, and leadership." *Journal of Social Issues* 57:637–655.
- Ridgeway, Cecilia L and Shelley J Correll. 2004. "Unpacking the Gender System: A Theoretical Perspective on Gender Beliefs and Social Relations." *Gender and Society* 18:510–531.
- Rockoff, Jonah E. 2004. "The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data." *The American Economic Review* 94:247–252.
- Rosenbaum, Paul. 2002. *Observational Studies*. New York: Springer Verlag.
- Rothstein, J. 2010. "Teacher Quality in Educational Production: Tracking, Decay, and Student Achievement." *Quarterly Journal of Economics* 125:175–214.
- Rubin, Donald B. 2001. "Using Propensity Scores to Help Design Observational Studies: Application to the Tobacco Litigation." *Health Services and Outcomes Research Methodology* 2:169–188.

- Rumberger, Russell W. 2003. "The Causes and Consequences of Student Mobility." *The Journal of Negro Education* 72:6–21.
- Rumberger, Russell W. and Katherine A. Larson. 1998. "Student Mobility and the Increased Risk of High School Dropout." *American Journal of Education* 107:1–35.
- Rumberger, Russell W and Gregory J Palardy. 2005a. "Does Segregation Still Matter? The Impact of Student Composition on Academic Achievement in High School." *Teachers College Record* 107:1999–2045.
- Rumberger, Russell W. and Gregory J. Palardy. 2005b. "Does Segregation Still Matter? The Impact of Student Composition on Academic Achievement in High School." *Teachers College Record* 107:1999–2045.
- Sacerdote, Bruce. 2010. "Peer Effects in Education: How Might They Work, How Big are They, and How Much Do We Know Thus Far?" In *Handbook of Economics of Education*, edited by Eric A. Hanushek, Stephen Machin, and Ludger Woessmann. Amsterdam: North Holland.
- 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*. 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* 114:189–231.
- Sampson, Robert J., Jeffrey D. Morenoff, and Thomas Gannon-Rowley. 2002. "Assessing "Neighborhood Effects": Social Processes and New Directions in Research." *Annual Review of Sociology* 28:443–478.
- Sampson, Robert J., Patrick Sharkey, and Stephen W. Raudenbush. 2008. "Durable effects of concentrated disadvantage on verbal ability among African-American children." *Proceedings of the National Academy of Sciences* 105:845–852.
- 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. Waxmann.
- Sewell, W.H., A.O. Haller, and A. Portes. 1969a. "The educational and early occupational attainment process." *American Sociological Review* 34:82-92.
- Sewell, W.H., R.M. Hauser, and W.C. Wolf. 1980. "Sex, schooling, and occupational status." *American Journal of Sociology* 86:551-583.
- Sewell, W. H, A. O Haller, and A. Portes. 1969b. "The educational and early occupational attainment process." *American Sociological Review* 34:82-92.
- Sewell, William H., Archibald O. Haller, and Alejandro Portes. 1969c. "The Educational and Early Occupational Attainment Process." *American Sociological Review* 34:82-92.
- Seymour, E. and N.M Hewitt. 1997. *Talking About Leaving: Why Undergraduates Leave the Sciences*. Boulder, CO: Westview Press.
- Shadish, William R., M. H. Clark, and Peter M. Steiner. 2008a. "Can Nonrandomized Experiments Yield Accurate Answers? A Randomized Experiment Comparing Random and Nonrandom Assignments." *Journal of the American Statistical Association* 103:1334-1344.
- Shadish, William R., M. H. Clark, and Peter M. Steiner. 2008b. "Can Nonrandomized Experiments Yield Accurate Answers? A Randomized Experiment Comparing Random and Nonrandom Assignments." *Journal of the American Statistical Association* 103:1334-1344.
- Sharkey, Patrick and Felix Elwert. 2011. "The Legacy of Disadvantage: Multi-generational Neighborhood Effects on Cognitive Ability." *American Journal of Sociology* 116:1934-1981.
- Shauman, K.A. 2008. "Are There Sex Differences in the Utilization of Educational Capital Among College-Educated Workers?" Unpublished working paper.
- Signorella, M. L, R. S Bigler, and L. S Liben. 1993. "Developmental differences in children's gender schemata about others: A meta-analytic review." *Developmental Review* 13:147-183.
- Skelton, C. 1997a. "Primary boys and hegemonic masculinities." *British Journal of Sociology of Education* 18:349-369.

- Skelton, Christine. 1997b. "Primary Boys and Hegemonic Masculinities." *British Journal of Sociology of Education* 18:349–369.
- Sobel, Michael E. 1981. "Diagonal Mobility Models: A Substantively Motivated Class of Designs for the Analysis of Mobility Effects." *American Sociological Review* 46:893–906.
- Sobel, Michael E. 1985. "Social Mobility and Fertility Revisited: Some New Models for the Analysis of the Mobility Effects Hypothesis." *American Sociological Review* 50:699–712.
- Sorensen, Aage B. and Stephen L. Morgan. 2006. "School Effects: Theoretical and Methodological Issues." In *Handbook of the Sociology of Education*, edited by Maureen T. Hallinan, pp. 137–160. New York: Springer.
- South, Scott J. and Dana L. Haynie. 2004. "Friendship Networks of Mobile Adolescents." *Social Forces* 83:315–350.
- South, Scott J., Dana L. Haynie, and Sunita Bose. 2007. "Student mobility and school dropout." *Social Science Research* 36:68–94.
- Spender, Dale. 1982. *Invisible Women: The Schooling Scandal*. London: Writers and Readers.
- Spilerman, Seymour. 1971. "Raising Academic Motivation in Lower Class Adolescents: A Convergence of Two Research Traditions." *Sociology of Education* 44:103–118.
- Staiger, Douglas and James H. Stock. 1997. "Instrumental Variables Regression with Weak Instruments." *Econometrica* 65:557–586.
- Stanworth, Michelle. 1984. *Gender and Schooling: A Study of Sexual Divisions in the Classroom*. London: Routledge.
- Steele, C. 2010. *Whistling Vivaldi: And other clues to how stereotypes affect us*. WW Norton & Company.
- Steinberg, L.D., B.B. Brown, and S.M. Dornbusch. 1996. *Beyond the classroom: Why school reform has failed and what parents need to do*. Simon & Schuster.
- Steiner, Peter M., Thomas D. Cook, William R. Shadish, and M. H. Clark. 2010a. "The importance of covariate selection in controlling for selection bias in observational studies." *Psychological Methods* 15:250–267.
- Steiner, Peter M., Thomas D. Cook, William R. Shadish, and M. H. Clark. 2010b. "The importance of covariate selection in controlling for selection bias in observational studies." *Psychological Methods* 15:250–267.

- Super, Donald E. 1953. "A theory of vocational development." *American Psychologist* 8:185–190.
- Super, Donald E. 1980. "A life-span, life-space approach to career development." *Journal of Vocational Behavior* 16:282–298.
- 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* 37:483–504.
- Torre, Marisa de la and Julia Gwynne. 2009. "Changing Schools: A Look at Student Mobility Trends in Chicago Public Schools Since 1995." Technical report, Consortium on Chicago School Research, Chicago.
- Tracey, Terence J.G., Steven B. Robbins, and Christy D. Hofsess. 2005. "Stability and change in interests: A longitudinal study of adolescents from grades 8 through 12." *Journal of Vocational Behavior* 66:1–25.
- Turner, S. L and R. T Lapan. 2005. "Evaluation of an intervention to increase non-traditional career interests and career-related self-efficacy among middle-school adolescents." *Journal of Vocational Behavior* 66:516–531.
- Urberg, Kathryn A, Qing Luo, Colleen Pilgrim, and Serdar M Degirmencioglu. 2003. "A two-stage model of peer influence in adolescent substance use: individual and relationship-specific differences in susceptibility to influence." *Addictive Behaviors* 28:1243–1256.
- Vernberg, Eric M. 1990. "Experiences with Peers Following Relocation During Early Adolescence." *American Journal of Orthopsychiatry* 60:466–472.
- West, Candace and Don H. Zimmerman. 1987. "Doing Gender." *Gender and Society* 1:125–151.
- Whitmire, Richard. 2010. *Why Boys Fail: Saving Our Sons from an Educational System That's Leaving Them Behind*. New York: AMACOM.
- Williams, J. 2000. *Unbending Gender: Why Family and Work Conflict and What to Do About It*. Oxford: Oxford University Press.
- Willis, Paul. 1981. *Learning to Labor: How Working Class Kids Get Working Class Jobs*. New York: Columbia University Press.
- Wilson, K.L. and J.P. Boldizar. 1990. "Gender Segregation in Higher Education: Effects of Aspirations, Mathematics Achievement, and Income." *Sociology of Education* 63:62–74.

- Wilson, William J. 1987. *The Truly Disadvantaged: The Inner City, the Underclass, and Public Policy*. Chicago: University of Chicago Press.
- Wodtke, Geoffrey T., David J. Harding, and Felix Elwert. 2011. "Neighborhood Effects in Temporal Perspective: The Impact of Long-Term Exposure to Concentrated Disadvantage on High School Graduation." *American Sociological Review* 76:713–736.
- Xie, Y. and K. A. Shauman. 1997. "Modeling the sex-typing of occupational choice." *Sociological methods & research* 26:233–261.
- Xie, Yu and Kimberlee A. Shauman. 2005. *Women in Science: Career Processes and Outcomes*. Cambridge, MA: Harvard 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* 20:325–341.
- Zubizarreta, Jose R. 2012. "Using Mixed Integer Programming for Matching in an Observational Study of Kidney Failure After Surgery." *Journal of the American Statistical Association* 107:1360–1371.