Stockholm School of Economics
Department of Economics
659 Degree project in economics
Spring 2017

# An Intersectional Analysis of Gender Peer Effects in Swedish Classrooms

Olivia Elf (23428) and Artur Obminski (23411)

---

**Abstract:** In this study, we examine the impact of the proportion of girls among classroom peers on fourth-grade mathematics achievement in Sweden. To allow for the possibility that different groups of students are affected in distinct manners, we separate the effects based on students' gender and whether they are new immigrants. Furthermore, we distinguish the effect unique to students who are at the intersection of these two identities—new immigrant girls. This intersectional approach provides a novel and valuable contribution to the study of peer effects. By using microdata from the international TIMSS study, we are able to compare classrooms within each school, and hence control for systematic between-school variation in students' backgrounds. We find that the positive average effects of a higher share of girls previously found in other countries are not present among Swedish fourth-grade students. New immigrant girls, however, are substantially positively affected, in contrast to new immigrant boys. This suggests that an intersectional approach is essential in order to target the most disadvantaged groups and thereby effectively enhance primary school achievement in mathematics.

---

Keywords: Peer effects, education, intersectionality, gender, TIMSS

JEL: I21, I24, J15, J16

| | |
|---|---|
| Supervisor: | Örjan Sjöberg |
| Date submitted: | May 14, 2017 |
| Date examined: | June 1, 2017 |
| Discussants: | Paula Dahlman & Gustav Tillman |
| Examiner: | Mark Sanctuary |

# Acknowledgements

We would like to express our heartfelt gratitude to our supervisor Örjan Sjöberg for his indispensable support, helpful critique, and genuine enthusiasm throughout the process of writing this thesis. In addition, we would like to thank Per-Olov Edlund for his valuable comments and advice. Last, we extend our appreciation to all those who have contributed comments and suggestions during our undertaking of this project, as well as to the International Association for the Evaluation of Educational Achievement (IEA) for granting us the permission to use their TIMSS data.

# Contents

# 1 Introduction

The effect of peers on individual behavior in group settings is of relevance in many areas of study, but the classroom setting in most schools implies that the impact of peers' behavior and background on students' achievement is of particular interest. In primary school especially, students often learn and spend large amounts of time interacting with a relatively stable composition of peers. This implies that any influences of class composition may have substantial effects on achievement, which previous research indeed suggests is the case (Coleman, 1966; Sacerdote, 2011).

Understanding peer effects and their magnitudes is central to enhancing academic achievement. The existence of such effects may be exploited by schools or policy makers to allocate students to schools and classes in such a way as to maximize outcomes. Furthermore, negative peer effects causing lower achievement can be predicted and compensated for.

The purpose of this paper is to examine how the gender composition among classroom peers affects student achievement and how these effects differ depending on the characteristics of the affected group. While the effects of classmates' gender composition on individual achievement have been extensively evaluated—also allowing for different effects on girls than on boys, or on immigrants than on non-immigrants—this paper adds an intersectional perspective not present in previous literature. We use the notion in intersectional theory of multiple interacting identities to assess if and how new immigrant girls are affected not only by being new immigrants and girls, but also by being at the intersection of these identities. This approach, we argue, sheds new light on the sources of variation in student achievement and where policy efforts may help improve outcomes especially for the most disadvantaged.

The hypotheses of this paper are examined in a Swedish context, as we find no achievement differences between girls and boys among Swedish fourth-grade students in mathematics. Hence, we are able to distinguish gender peer effects from otherwise potentially correlated ability effects. In addition, few studies have examined gender peer effects in Swedish schools.

The structure of the paper is as follows. Section 2 provides an overview of previous literature on peer effects in education, the impacts of classroom gender composition, and the implications of intersectional theory, before outlining statistical challenges with and approaches to identifying peer effects. Section 3 then outlines the research design, including contributions and hypotheses. The Swedish school system, and its allocation of control, is then considered in Section 4.

In Section 5, the dataset applied and the associated sampling procedure are featured, before the model and variables are discussed in Section 6. After a presentation of results and sensitivity analyses in Section 7, the results and implications are discussed in Section 8. We conclude in Section 9 with a summary and potential areas of future work.

# 2 Literature Review

The following literature review begins in Section 2.1 with an overview of different types of peer effects in educational settings and empirical evidence for their direction and magnitude in a variety of contexts. In particular, evaluations of peer effects of gender composition and of heterogeneous effects that may differ between subgroups will be emphasized. An account then follows of potential explanatory mechanisms behind observed effects of gender composition in classrooms in Section 2.2. Section 2.3 then provides a brief discussion of intersectional theory and its general predictions. Last, in Section 2.4, the literature review ends by considering challenges in identifying peer effects and potential measures to address these challenges.

## 2.1 Peer effects in education

The importance of students' peer environment on educational outcomes has been a prominent subject in economics and other social sciences since the publication of the so called *Coleman Report*, a comprehensive inquiry into educational equality and effects on student achievement in the United States, in 1966. Among the main findings in the report was that "the social composition of the student body is more highly rated to achievement, independently of the student's own social background, than is any school factor" (Coleman, 1966, p. 325).

Peer effects take a variety of forms, as Sacerdote (2011) explains:

> if a student's classmates have higher incoming ability and the student learns directly from her classmates, that is a peer effect. If the classmates have higher incoming ability and this enables the teacher to teach at a higher level or a more demanding pace, that is also a peer effect. If the student is disruptive and

> consumes more of the teacher's attention, thereby reducing her classmates' test scores, that too is a peer effect. (pp. 250–251)

The substantial variety in the field is evidenced in the range of research setups used in the literature on peer effects in education. A distinction is often made between effects of peer *ability*, typically proxied by test scores, and effects of peers' *characteristics*, such as gender or race (Cooley, 2010). These will be discussed further in the following sections, 2.1.1, 2.1.2, and 2.1.3. Although most studies use an achievement outcome as their dependent variable, research exists also on topics such as desire to pursue higher education and substance abuse (Sacerdote, 2011).

Furthermore, some studies allow for nonlinear effects: when students are affected differently based on their characteristics, such as gender, race, or initial ability (a phenomenon also called *heterogeneous effects*), or when the effects of these students on their peers depends on the students' own characteristics. Evidence of heterogeneous effects is relatively common within race, gender, and ability dimensions. With regards to race and gender, for instance, intra-group peer effects appear particularly strong (Cooley, 2010).

Two distinct methods of isolating peer effects are described by Cooley (2010). These are event studies and achievement production studies. Event studies use natural events such as student reallocation programs to compare students affected in different ways, sometimes with unaffected control groups. The introduction of mixed-gender classrooms in a previously gender-segregated setting is one example. Achievement production studies, in contrast, use existing variation in peer group composition to measure the effects of changes in peers' ability, behavior, or characteristics while controlling for other factors that may also affect these outcomes.

In a large share of the peer effects literature, the focus is on students and schools in the United States. In the U.S., a number of event studies have been conducted on student reassignment programs (including Hoxby and Weingarth, 2005; Vigdor and

Nechyba, 2007; Whitmore, 2005), and also among achievement production studies, many use United States data (Hoxby, 2000; Burke and Sass, 2013). In addition, countries including Israel (Lavy and Schlosser, 2011; Gould et al., 2009) and the Netherlands (Booij et al., 2016; Ohinata and van Ours, 2016) are subject to several studies. Where other, and sometimes multiple, countries are discussed, data from the OECD PISA study in mathematics and science (Entorf and Lauk, 2008; Schneeweis and Winter-Ebmer, 2007; Brunello and Rocco, 2013) is often used, while studies based on the IEA studies TIMSS (Kang, 2007; Raitano and Vona, 2013), in mathematics and science, and PIRLS (Ammermueller and Pischke, 2009), in reading, are less prevalent.

To a large extent, the peer effects literature focuses on primary and secondary school children (Lavy and Schlosser, 2011; McEwan, 2003; Sund, 2009), and particularly in the United States often in public schools (Hoxby and Weingarth, 2005; Vigdor and Nechyba, 2007; Fruehwirth, 2013). When university students are studied, the data is typically limited to one college or university, dormmates are sometimes studied rather than classmates, and the outcome of interest is not always grades or test scores, but sometimes choice of major or participation in social activities on campus (Sacerdote, 2001; Zimmerman, 2003; Foster, 2006).

### 2.1.1 Ability peer effects

Previous research on ability peer effects shows a wide range of estimates (Sacerdote, 2011). Hoxby and Weingarth (2005) argue that the sometimes contradictory results may be a consequence of the frequent use of the linear-in-means model, which assumes that every student affects every other student identically (i.e. has a homogeneous effect). This assumption masks the different effects experienced by different students and implies that any rearrangement of peers would generate the same average level of outcomes—a zero sum game. As a result, more recent literature generally presumes some type of nonlinear or heterogeneous effects.

Hoxby and Weingarth (2005) and Fruehwirth (2013) use data from a student reassignment program in public elementary schools of North Carolina to identify the effect of peers on student achievement. Hoxby and Weingarth find that students benefit from peer homogeneity with regards to ability, independently of their own ability level. Moreover, they find students to be more affected by peers within their own gender and racial subgroups. Also, on average, female peers have a more positive effect on achievement than male peers, even when accounting for ability differences.

These findings are in line with those of Fruehwirth (2013), who provides evidence that peer effects work solely within, and not between, racial subgroups. Moreover, she finds that lower-achieving students are relatively more affected by average (intraracial) peer achievement.

Such heterogeneous effects can also be found examining effects on girls and boys separately. Lavy et al. (2012) find that girls benefit from high-achieving peers while boys are unaffected. Girls are also negatively affected by low-achieving peers to a larger extent than boys are.

Hanushek et al. (2003) and Burke and Sass (2013) examine ability peer effects in U.S. elementary schools, allowing for heterogeneous results by separating the effects on different ability levels. Hanushek et al. (2003) find that average peer achievement has a significant effect on student achievement across the test score distribution, with a smaller effect on the highest-achieving students. However, because of data limitations, these results are based on school-by-grade fixed effects while school-by-class fixed effects are assertedly preferable since it allows for a more accurate definition of peer groups. Burke and Sass, in contrast, find that low-achieving students are negatively affected by an increase of students at either end of the ability distribution while middle-achieving students benefit from a larger fraction of high-achievers and, lastly, high-achievers benefit from an increase in either low-ability or high-ability peers. Moreover, they compare classroom-level and grade-level peer effects and find that only effects on the classroom level are significant. Therefore, they argue, peer effect estimates are highly dependent on the correct identification of a student's relevant peers.

The identification of peer effects in post-secondary education often relies on roommate configurations since these are easily identified (Sacerdote, 2011). Results from a number of these studies are similar, as there appears to be an agreement that roommate academic achievement has a positive and significant impact on own academic outcomes (Sacerdote, 2001; Zimmerman, 2003; Griffith and Rask, 2014; Booij et al., 2016). Furthermore, Griffith and Rask (2014) find that male, low-ability, minority, or financially aided students benefit the most from being paired with a high-ability roommate, while the high-ability student is not hurt by this pairing. Booij et al. (2016) also present evidence that high-ability students affect students in the lower two thirds of the GPA distribution positively while high-ability students themselves are unaffected.

Previous research on peer effects in Swedish classrooms is scarce. Sund (2009) uses ability differences between different subjects for each student to examine peer effects among high school students in the municipality of Stockholm. He finds that an increase in peer achievement has a positive impact on student outcomes. Furthermore, students at the lower end of the achievement distribution benefit more from a heterogeneous class composition with regards to ability than high-achieving students, although the overall effect is positive.

Sweden also appears in some international comparisons on ability peer effects (Kang, 2007; Entorf and Lauk, 2008; Raitano and Vona, 2013). The former uses TIMSS data while the two latter use OECD PISA data. Kang compares the magnitude of peer effects across countries by examining within-student differences between mathematics and science test scores. He finds a positive correlation between own and peer achievement for most TIMSS countries and the magnitude of these correlations to be notably close across countries considering the substantial institutional differences in middle school education.

Entorf and Lauk (2008) find stronger peer effects in countries with school systems that to a greater extent sort students based on ability. Comprehensive Scandinavian school systems encompass the smallest peer effects, with only native-to-native peer effects being significant. However, results based on PISA data may be less reliable than those based on registry data for two reasons. First, the selection of available control variables is limited, resulting in a greater risk of incorrectly specified models. Second, cohorts of fifteen-year-olds in a school are sampled, rather than classes, capturing incomplete school-level peer effects rather than classroom-level effects. These school-level effects, compared to class-level effects, are less likely to capture and control for unobserved factors that may vary systematically between schools, which would have at least partially compensated for the lack of control variables.

In short, although previous literature provides a wide range of conclusions on ability peer effects, a few results appear to be recurring. First, the majority of the studies that test for heterogeneous effects find such to be present. In particular, low-ability and other disadvantaged groups are affected to a larger extent. Second, ability peer effects appear to work to a larger extent within subgroups such as race or gender than between them. Last, higher-achieving peers seem to almost always be positive for other students' achievements, except for particularly low-achieving students who are occasionally negatively affected by this ability gap.

### 2.1.2 Characteristics peer effects

An alternative to focusing on ability peer effects is to estimate the influence of peer characteristics, typically demographic characteristics such as gender or race. Other characteristics studied, and for which positive effects have been observed, include books at home among fourth-graders in six European countries (Ammermueller and

Pischke, 2009) and class size in German eighth grade (Fertig, 2003). These characteristics may both be proxies for other, unobserved factors correlated with the characteristics in question, such as the level of disrupting behavior, and represent direct effects of the characteristics themselves. However, the distinction is not necessarily important. In a study of the effects of racial integration, for instance, the aggregate effects may be the primary target of interest rather than the separate direct and indirect effects (Cooley, 2010).

#### Immigrant peer effects

A common focus among characteristics studies is on immigrants, although the definition of 'immigrant' varies, as well as the age and nationality of the population studied. For instance, defining immigrants as foreign born and arriving in Israel after 1989, Gould et al. (2009) find a small yet significant negative effect on native Israeli fifth-grade children's high school matriculation and dropout rates, and that the effect on socioeconomically disadvantaged natives is particularly large.

The literature on immigrant educational peer effects is diverse, but heterogeneous analyses often indicate the existence of such effects. Although Ohinata and van Ours (2016) do not find any significant effects of classroom immigrant share on Dutch eleven-to-twelve-year-old students, their definition of immigrant is wide, as it requires only one foreign-born parent. Diette and Oyelere (2014) use data from North Carolina, U.S., and instead apply the state categorization of some students, often with Latin American background, as 'limited English.' Using data on grade levels three to eight they find large negative effects of a higher proportion of immigrants on reading achievement for black males, and small but still negative effects on white males' mathematics scores.

Negative peer effects are more notable in studies using OECD PISA data. Both Brunello and Rocco (2013), studying nineteen different countries, and Jensen and

Rasmussen (2011), focusing on Denmark, find negative effects of a high share of immigrants in school on test results in mathematics, and Brunello and Rocco also in reading. These studies both define foreign-born students with two foreign parents as immigrants, while Jensen and Rasmussen also allow for Danish-born students with immigrant parents. However, the limitations of PISA data discussed in Section 2.1.1 apply also to these studies.

In brief, estimates of immigrant peer effects vary with relatively large differences in the use of the 'immigrant' term, although there is some indication that negative effects exist and are stronger on socioeconomically disadvantaged peers.

### 2.1.3 Gender peer effects

The perhaps most studied characteristic in the educational peer effects literature is gender, where studies on the whole find significant positive effects of increasing the proportion of girls in the classroom, as discussed by Sacerdote (2011). These effects seem to operate not only through gender differences in average ability, but also contain nonlinear dimensions. The effects increase in magnitude at higher percentages of girls and vary with students' gender, immigrant status, and socioeconomic background.

Clear positive effects are found on mathematics and reading test scores in Hoxby (2000) and Hoxby and Weingarth (2005), in primary school grades three to six in Texas and three to eight in North Carolina, respectively. Hoxby (2000) finds these effects, on both genders, to be nonlinear and increasing at higher percentages of girls in the classroom.

Both Whitmore (2005) and Gottfried and Graves (2014) provide similar findings, also based on US data, although for students in kindergarten to third grade. While Whitmore analyzes an average of mathematics and reading test scores, and finds similar results for both genders, Gottfried and Graves identify gendered subject differences: although boys are affected markedly in reading, girls are affected primarily in mathematics. They argue that these differences can be attributed to boys' higher sensitivity to peers' disrupting behavior and girls' sensitivity to gender norms and identity issues appearing primarily in the later part of the early elementary school years.

Contrasting evidence, with little indication of significant peer effects, is found in Oosterbeek and van Ewijk (2014) and Vigdor and Nechyba (2007). Oosterbeek and van Ewijk find only small declines in absenteeism and delays in male dropouts in a study conducted on Dutch university students—suggesting that gender peer effects may be less likely among students in tertiary education than in primary school classrooms. Vigdor and Nechyba instead focus on fifth-grade students in North Carolina public schools. Although they originally find a positive effect of higher shares of girls in the classroom on end-of-grade scores in mathematics, this effect is shown to be correlated also with lagged scores, suggesting inaccurate estimates due to missing variables. Based on an additional test using North Carolina experimental data, the authors conclude that the gender peer effects observed are not causal, but due to sorting.

Nevertheless, Vigdor and Nechyba (2007) do not allow for heterogeneous effects in their study of gender peer effects. As evidenced by Lavy and Schlosser (2011) in their study of Israeli elementary, middle, and high schools, such differences in effects between different affected groups may be substantial. Although the authors, in contrast to Vigdor and Nechyba, find evidence of positive peer effects of a high percentage of girls in the classroom on average test scores in mathematics and science, these effects are particularly noticeable for students with low parental education and new immigrants, defined as having lived in Israel for five years or less. These conclusions indicate that informational value is added in heterogeneous analyses of educational peer effects.

## 2.2 Theoretical background to gender peer effects

The findings in Section 2.1.3 indicate that a higher percentage of girls in the classroom has positive effects on student achievement, and particularly on the achievement of girls and disadvantaged groups. With a focus on mathematics test scores, Hoxby (2000) presents three theories as to why this might be the case:

> First, since learning math requires reading and reading scores are higher in more female classes, females may affect subjects like math *through* their (quite plausible) peer effect on reading. Second, more female classes may simply have fewer disruptive students or a more learning-oriented culture. Third, classroom observers argue that pressure to be feminine makes girls unenthusiastic about math. Perhaps in female-dominated classrooms, females do not experience much pressure and therefore remain enthusiastic about math—allowing the teacher to teach it better to all students. (p. 24)

In this section, this classification will be used to further examine how such impact mechanisms of gender composition in the classroom may work and affect gender and immigrant subgroups differently.

First, although boys tend to outperform girls in mathematics and science across the OECD countries, according to OECD (2015), no such gender gap exists in Sweden in these subjects. However, girls consistently perform better than boys in reading, with Sweden experiencing a larger reading gender gap than the OECD average. Hoxby (2000) suggests that, through such a gender difference in reading skills, a higher percentage of girls in the classroom can indirectly affect mathematics achievement positively for both genders and, hence, increase average class ability.

Furthermore, as previous literature indicates that a higher classroom ability level benefits minorities and students with low socioeconomic status to a greater extent (e.g. Sund, 2009; Griffith and Rask, 2014; Booij et al., 2016), it is plausible that immigrants are among the groups that can be more positively affected by such an increase in average class ability.

Second, as evidenced in previous literature, gender composition in the classroom is likely to have an impact on the peer environment. For instance, Lavy and Schlosser (2011) find that a higher percentage of girls lowers the level of disruption and violence, favors cooperation between peers, and improves teacher-student relationships. Moreover, they find that the improved study environment that results from a higher proportion of girls in the classroom has a larger effect on new immigrants' achievement than on that of non-immigrants. It is argued that, just like reducing class size, a higher proportion of female students can be regarded as a strategy to reduce classroom violence and disruption. Reducing these factors of disturbance, hence improving the study environment, will to a greater extent favor students who are in need of additional instructional time and who are easily affected by disruptive peers. In Lavy and Schlosser (2011), such students are found to be new immigrants and students from low socioeconomic backgrounds.

Third, existing gender stereotypes could affect girls' performance in mathematics. For instance, Wolter and Hannover (2016) find that "even in primary school children associate high capabilities and motivation in mathematics with men and boys, but high capabilities and motivation in reading with girls and women" (p. 682). Muzzatti and Agnoli (2007) also find negative gender stereotypes for girls in mathematics. More specifically, they find that fourth grade students of both genders regard boys as better in mathematics than girls, although no actual differences in ability between the genders were detected.

Evidence shows, however, that a higher share of girls in the classroom could weaken gender stereotypes such as the above and hence mitigate their negative impact on girls' performance in mathematics. Inzlicht and Ben-Zeev (2000), for instance, find that females who participated in mathematics tests performed worse when surrounded by male participants while the male participants performed equally well independently of the gender composition. Moreover, Schneeweis and Zweimüller (2012) find that girls who had been exposed to a higher share of girls during grades five to eight were more likely to choose a traditionally male dominated school type, which is considered an indication of mitigated gender stereotypes.

In brief, a higher share of girls in the classroom can affect the ability level, classroom environment, or the magnitude of existing gender norms. Although such changes can affect all students, they appear to impact girls and disadvantaged groups, such as immigrants, to a higher extent.

## 2.3 Intersectionality

The term intersectionality was first coined by American civil rights advocate Kimberlé Williams Crenshaw in 1989 and can be used to describe how multiple identities interact and reinforce each other, which creates an effect that is greater than the aggregate of the separate identities (Crenshaw, 1989). Such identities can include race, gender, social class, physical or mental disability, sexual orientation, or other factors. As a result, social inequalities occur on a multidimensional level which any analysis that addresses only one form of marginalized identity fails to acknowledge. As Crenshaw explains:

> Because the intersectional experience is greater than the sum of racism and sexism, any analysis that does not take intersectionality into account cannot sufficiently address the particular manner in which Black women are subordinated. (p. 140)

These effects have since been found in a variety of contexts, such as on the labor market and within organizations (Acker, 2006; Browne and Misra, 2003).

## 2.4 Statistical identification of peer effects

Identifying peer effects of interest is complex, as different types of effects are easily confounded and data on important explanatory variables is often missing. A distinction is commonly made between three different peer effect mechanisms, each with distinct policy implications. These will be discussed further in Section 2.4.1. The consequences of limited data availability will be expanded upon in Section 2.4.2.

### 2.4.1 Peer effects mechanisms

Indications of similarities in the behavior or results of peers, such as test scores, can be explained through three different mechanisms. Manski (1993, 2000) makes a distinction between endogenous effects, exogenous (or contextual) effects, and correlated effects. Endogenous effects imply that individual behavior is to some extent dependent on peers' behavior, such as student achievement on peers' achievement. Exogenous effects instead indicate that behavior depends on peers' exogenous characteristics, such as gender or race. Both of these mechanisms explain how individuals are affected by their peers. The third mechanism, correlated effects, instead suggests that individuals act in a similar manner due to common background characteristics or a common external environment. Manski (1993) illustrates the differences with an example:

> Consider the high school achievement of a teenage youth. There is an endogenous effect if, all else equal, individual achievement tends to vary with the average achievement of the students in the youth's school, ethnic group, or other reference group. There is an exogenous effect if achievement tends to vary with, say, the socio-economic composition of the reference group. There are correlated effects if youths in the same school tend to achieve similarly because they have similar family backgrounds or because they are taught by the same teachers. (p. 533)

This difference is of importance to policy. Manski (2000) explains how endogenous effects allow for so-called social multipliers. If they are present, an increase in the performance of one student will raise the performance of classmates, whose improved performance will, in turn, affect the student herself positively—a feedback loop. Exogenous effects do not allow for such multiplication, and correlated effects are entirely non-social and hence typically not of interest in peer effects studies.

### 2.4.2 Further challenges in identifying peer effects

The risk of confounding the three types of mechanisms as well as data limitations result in challenges in identifying causality correctly. First, the difficulty in isolating direct effects of peers on individuals from indirect effects from an individual through her peers on the individual herself constitutes the *reflection problem.* Second, selection into classrooms and schools is unlikely to be completely random, and this nonrandomness may result in biased estimates—a *selection problem.* Finally, *omitted variables* generate additional challenges. These will be discussed in turn, followed by an overview of *potential solutions.*

#### The reflection problem

The reflection problem, also referred to as 'simultaneity,' is well discussed by Hoxby and Weingarth (2005). It is based on that an individual not only is influenced by her peers, but also influences those same peers. However, these peers have already been affected by the individual. Hence, to some extent, peer effects may capture characteristics or behavior of the individual herself in the peer measures. This problem is especially large in estimations of endogenous peer effects, as the social multipliers discussed in Section 2.4.1 increase the difference between the original direct effect and the observed total effect. The reflection problem may result in measurement biases in peer effects studies.

The challenge of separating endogenous and exogenous effects further magnifies the reflection problem, as explained by Manski (2000). Often in analyses of ability peer effects, peers' behavior (such as achievement), peer characteristics (including gender), and common background factors (for instance teacher performance)—endogenous, exogenous, and correlated effects, respectively—all affect students' achievement. In these settings, although it may be possible to isolate correlated effects, it is often impossible to distinguish between endogenous and exogenous effects. The gender of a student, for instance, may affect peers' achievement, which in turn affects the achievement of the individual student. There is a feedback loop originating in two distinct sources of influence.

**The selection problem**

Nonrandom allocation to classes and schools, in addition, is a large source of concern. As Cooley (2010, p. 2) explains, "unobservable shared effects [...] may be correlated with peer characteristics." New immigrants, for instance, may be more likely to attend certain schools, due to segregated housing or other factors. Within schools, furthermore, the allocation of students between classes may be affected by implicit or explicit strategies to balance the genders or distribute new immigrants evenly, for example. If these sources of nonrandomness are not fully controlled for, they are likely to present a source of bias.

A similar problem occurs if there is systematic variation between peer group composition and unobserved aspects that influence outcomes (Cooley, 2010). One example, which Cooley also points to, is when teachers adapt their teaching to the ability level or demographic composition of the class. This issue is not solved by random allocation of students into classes, as the unobserved teacher response still affects outcomes.

### Omitted variables

In a more general sense, omitted and incorrectly measured variables inhibit correct estimations. If factors that influence students' outcomes are excluded from a model and are correlated with included variables, the included variables, Hoxby and Weingarth (2005) explain, "will appear to cause the student's outcomes when they are really just proxying for his own characteristics" (p. 9)—an important source of bias.

Model specifications including only contemporaneous measures, and not, for instance, the composition of previous classes, often suffer from omitted variables bias. As Todd and Wolpin (2003) explain, these specifications only generate correct estimates if (i) only contemporaneous factors affect the outcome variable, or (ii) inputs are fixed over time, so that historical values are included. Furthermore, (iii) the contemporaneous inputs have to be uncorrelated with unobserved levels of ability. Conditions (i) and (ii) in particular, if not properly addressed, limit the possibility of unbiased estimation.

### Potential solutions

To address some of the identification problems, studies often include proxy variables, lagged scores, or fixed effects. Alternatively, analyses may be based on natural experiments. However, these approaches give rise to additional problems, as discussed in the following paragraphs based on Todd and Wolpin (2003).

Proxy variables are often included to control for unobserved variables such as parental investment in children's education. Proxies that are poorly correlated with the unobserved factors, may, however, contribute to measurement biases. For instance, with both an expenditure proxy (such as books at home) and income included, holding income fixed, an increase in the expenditure proxy implies a decrease in funds for other goods, and these two effects will be confounded.

Another common estimation technique is the 'value added' approach, which uses lagged variables to control for historical inputs while studying the effects of contemporaneous factors on changes in outcomes. Nevertheless, this approach is based on the assumption that the size and direction of the effects are constant over the ages studied. Furthermore, as omitted variables are often serially correlated (family inputs are often similar between one year and the next, for example), estimations of lagged variable effects are likely to be downward biased, which makes it impossible to determine the correct model specification.

Fixed effects are used to control for unobserved factors that influence outcomes but are assumed to be fixed within a group setting, such as a classroom. Such estimations, however, require the assumption that the unobserved factors of relevance are common to all individuals within the group. Nonetheless, several of the challenges associated with the 'value added' model are omitted.

Finally, some studies (including Hoxby and Weingarth, 2005; Vigdor and Nechyba, 2007) use natural experiments, such as student reallocation programs. This approach often provides sound estimates, but these estimates are not directly comparable to those found through other approaches, as school and family inputs often change as a result of the experiments.

# 3 Design and Hypotheses

As discussed in the above literature review, positive effects on achievement of a higher percentage of girls among classroom peers have been established in a variety of contexts. These effects operate not only through differences in ability, but also appear to relate to disruption and gender norms. Furthermore, the existence and magnitude of these peer effects depend on the demographic characteristics, such as gender, of the students affected. Specifically, girls appear more susceptible to changes in peer gender composition due to the influence of gender norms, and immigrants due to their sensitivity to the class environment.

However, studies discussing peer effects in contexts with interacting characteristics, such as effects on an immigrant girl, are rare. The main contribution of this paper is to allow for these variations in peer influence depending on interacting social identities, as discussed in the intersectionality literature. Hence, the analysis presents not only the effects of gender peer effects on boys and girls, and new immigrants and non-immigrants, but also distinguishes the effects on students at the interaction of the two categories—and in particular on new immigrant girls.

Furthermore, by using the share of girls as an exogenous characteristic, rather than involving a measure of ability, we are able to mitigate the reflection problem, further discussed in Section 2.4.2, and can estimate more precisely the direct effects on peers' achievement. By using achievement in mathematics as the outcome variable, the exogenous effects of peers' gender can also be isolated from direct effects of mathematics ability that may be correlated with gender. This is possible as gender differences in mathematics achievement in our data are largely non-existent, as shown in Table 6.1.

Specifically, the impact of the share of girls in the classroom, excluding the individual student, on test scores in mathematics in the international TIMSS test will be examined. The results will be presented first on an aggregate level, and then accounting for the gender of each student and whether he or she is a new immigrant, as well as interactions between the groups.

We choose to restrict the evaluation to fourth-grade students' mathematics test scores in TIMSS in Sweden. The focus on Sweden is advantageous for at least three reasons. First, Swedish fourth-grade students are largely given instruction in the same classroom peer group in all subjects and sorting on the basis of ability is not present at this early age. (See Section 4 for a further discussion of the Swedish school system.) Second, as discussed above, we observe no substantial gender gap in mathematics achievement in the Swedish data (see Table 6.1). Third, previous research on gender peer effects in Sweden is limited.

The use of TIMSS data, further described in Section 5, provides a good basis for comparative analyses, as TIMSS tests are standardized across schools and countries.

Our hypotheses are as follows:

**Hypothesis 1:** *A larger share of girls among classroom peers on average impacts positively on students' test scores in mathematics.*

Hypothesis 1 is widely supported in previous literature (e.g. Hoxby, 2000; Sacerdote, 2011). Explanations such as that girls' higher achievement in reading may influence peers' results in mathematics (Hoxby, 2000) are also potentially applicable in the Swedish context, as these differences in reading performance exist also in Swedish schools (OECD, 2015).

**Hypothesis 2:** *Girls are on average especially positively influenced by a larger share of girls in the classroom, in addition to the effect also on boys.*

Previous literature lends support also to Hypothesis 2 (Gottfried and Graves, 2014). Furthermore, as discussed in Section 2.2, gender norms affect girls' performance in mathematics negatively, and these norms are likely to weaken with an increasing share of girls in the classroom.

**Hypothesis 3:** *New immigrants are on average affected more positively by a larger share of girls among classroom peers than students who are not new immigrants.*

In addition to Lavy and Schlosser's (2011) findings in line with Hypothesis 3, the hypothesis is supported by several findings on the particular susceptibility of minority or academically disadvantaged groups to peer influences, including those of gender share (e.g. Griffith and Rask, 2014; Diette and Oyelere, 2014).

**Hypothesis 4:** *Students who are both girls and new immigrants see greater positive effects of a higher share of girls in the classroom than do new immigrant boys.*

Finally, Hypothesis 4 follows the argument in the literature on intersectionality, discussed in Section 2.3, that multiple identities of each individual interact and create an effect that is larger than can be explained by each individual identity alone. Intersectionality also includes the notion that the effect is further magnified at this intersection, and therefore is greater than can be explained by a simple aggregation of component effects (Crenshaw, 1989). However, due to econometric constraints, we are not able to test this aspect.

# 4 The Swedish School System

The Swedish school system, as described by OECD (2015), begins with preschool, which most children aged three to five attend, although it is not mandatory. At age six, a large majority of children attend preschool classes aimed to prepare them for primary school. Primary and lower secondary education, grades one to nine, are then compulsory before most students transition into upper secondary school. Before upper secondary school, the system does not feature tracking—all students study the same curriculum.

The system includes three types of schools targeting students with special needs, as discussed in OECD (2015): Sami schools, special schools, and schools for students with learning disabilities. While Sami schools differ from other schools by providing instruction in the Sami language in addition to Swedish, special schools target students with hearing disabilities, severe vision impairments, or significant language problems. Students with special needs are, nevertheless, in most cases educated in mainstream schools, and are only segregated in particular cases.

OECD (2015) also describe the major reforms carried out in the 1990's, impacting areas such as school administration, funding, and class formation. The primary responsibility for schools, including funding, hiring, and curriculum adaptations, was transferred from the state to municipalities. In connection with this change, the school sector was also liberalized to allow for more independent schools, so called 'friskolor' organized by companies, foundations, or associations (Skolverket, 2016a). The opportunity for parents to choose a school for their children was also introduced. Since then, the share of independent schools, 16 percent in 2012, has been growing.

State control of schools has also been limited, although gradually increasing in recent periods.

Municipalities typically delegate much of their responsibility for school organization, within-school resource allocation, and staffing to principals. Freedom for principals is particularly high in the formation of classes and student groups (Skolverket, 2011). Nevertheless, the Education Act of 2014 demands of principals as well as municipalities that resources are allocated based on students' needs and abilities (OECD, 2015).

Although formal control of schools is at the municipal level, government funds exist for particular purposes, particularly advancing equity, as described in Skolverket (2015). These funds, however, represent only a small share of total school funding. Applying for them is also voluntary. Hence, the distribution of funds is uneven, with municipalities applying for, and receiving, funds more often than independent schools, for instance.

Many municipalities also allocate part of their funds for schooling based on socioeconomic factors, according to data collected by Skolverket (2015, 2013). Parents' educational level is typically an important factor in allocating these funds, accounting for at least 50 percent of total grants based on socioeconomic factors in a majority of municipalities. The gender and immigrant background of students are among other common factors.

The results of these efforts to facilitate equity in the Swedish school system nonetheless appear moderate. OECD (2015), in their PISA test results, find that the variation in performance in mathematics and reading among Swedish fifteen-year-old students is close to the OECD average. This variation is, however, to a large extent within schools. Performance variations between schools are comparatively small, yet increasing.

Differences between schools are larger in the distribution of new immigrants, as discussed in Skolverket (2016b). The number of immigrant children arriving in Sweden has been rising since 2012 and reached a peak in 2015. As the children should, according to Swedish National Agency for Education guidelines, be enrolled in school within a month after arrival, this immigration contributes to a marked increase in the total number of students enrolled in Swedish schools. However, the share of new immigrants admitted by independent schools is small, while municipal schools accept the vast majority.

Although the weak performance of the Swedish school system observed in PISA results remains after controlling for new immigrants (OECD, 2015), an important influence on Swedish schooling has been a relatively large share of new immigrants among students, and that immigration increasingly originates from non-European countries. In 2012, for instance, a large share of new immigrant children in Swedish primary and lower secondary schools were born in non-European countries such as Afghanistan, Somalia, Syria, and Iraq (SCB, 2013).

# 5 Data

The data in this paper originates from the Trends in International Mathematics and Science Study (TIMSS), administered by the International Association for the Evaluation of Educational Achievement (IEA) (IEA, 2016). The information on the TIMSS study included in Sections 5, 5.1, and 5.2 is adapted from Martin et al. (2016).

TIMSS is a study conducted every four years where students' abilities in mathematics and science are assessed through designated tests. In addition, students, parents, teachers, and principals are asked to fill in questionnaires on topics such as study habits, home environment, and school resources. Although both students in their fourth year of formal schooling (grade level four in the Swedish school system) and students in their eighth year are sampled, we focus on the data for grade level four.

We use the most recent TIMSS assessment data, for which data collection was conducted in Northern Hemisphere countries between March and June of 2015. Furthermore, although 43 countries were studied, we choose to concentrate on Swedish schools and students.

The IEA typically samples 150 schools and a total of 4,000 students within these schools in each country to reach its precision requirements. In Sweden, 149 schools and 4,505 fourth-grade students were initially sampled. Nevertheless, after exclusions and attrition, 144 schools and 4,142 students remained in the assessment. These account for 4.3 percent of all Swedish schools and 4.1 percent of all students at the time of assessment.

## 5.1 Sampling procedure

The TIMSS sampling process proceeds in two stages: first of schools, then of classes. Schools are sampled with weighted probabilities based on the number of students enrolled. They are also stratified according to country-specific variables, typically to compensate for disproportionate allocations of students across the data. In Sweden, this stratification was based explicitly on whether the school offers both grade level four and grade eight or solely the former, and within the latter category also on the average achievement of fourth-graders in the school (categorized as low, high, or missing). In addition, Swedish schools were stratified implicitly (at a second level) based on school type (public, private, or all schools).

Within schools, one or more whole classes are sampled with equal probability. In Sweden, two classrooms were sampled in schools with more than 45 students in grade four, and one classroom in remaining schools. However, in the Swedish data, one school exists where four classrooms have been sampled.

## 5.2 Exclusions and attrition

Schools, classes, and individual students may in some cases be excluded in the TIMSS data before publication. In Sweden, very small schools (with less than five students in grade four), international schools (with a curriculum different from the national), and special education schools were excluded. Among individual students, excluded were students with intellectual or functional disabilities, and non-native language speakers. In the former case, for students to be excluded, disabilities had to be such as to make it impossible to participate in the assessment. Non-native language speakers were excluded if they were "unable to overcome the language barrier in the test situation" (Martin et al., 2016, p. 45), which typically applies to students having received less than one year of instruction in Swedish, as the TIMSS test is administered only in Swedish in Sweden.

Whole schools representing 1.7 percent of the students originally sampled in Sweden were excluded, and an additional 4.0 percent of individual students within non-excluded schools. This implies a total exclusion rate of 5.7 percent, somewhat above the TIMSS country average.

In addition to deliberate exclusions, there may also be missing data points, attrition. The TIMSS data includes a distinction between 'not administered' and 'omitted or invalid' observations. The former category includes when a student was not provided with a particular test item due to deliberate rotation of assessment blocks, which is not a source of concern, but also when whole questionnaires were not filled in, such as when parents did not answer the home background questionnaire. 'Omitted or invalid' observations, furthermore, include conscious omissions by students, as well as uninterpretable answers. Both these types of values were coded missing.

While there was no attrition on the school level in Sweden after the formal exclusions discussed, the student participation rate was 95 percent after attrition. Classrooms are excluded when student participation within the class is below 50 percent, but no classroom has been excluded in the Swedish data.

# 6 Empirical Approach

In the following section, the method for approaching our research question is outlined. We begin by defining our chosen variables, followed by a specification of the contemporaneous model of student achievement, adapted from Todd and Wolpin (2003), and, lastly, model assumptions and adaptations are described.

## 6.1 Definition of variables

### 6.1.1 Main variables of interest

The outcome measure of student achievement used in this paper is a standardization of the first ‘plausible value’ in mathematics provided in the TIMSS data. The focus on mathematics permits the isolation of gender peer effects from direct ability peer effects that may be correlated with gender, as discussed in Section 3.

The plausible values method allows for reliable achievement estimates from brief assessments by using all data available, including student characteristics, to impute scores from estimated ability distributions, as explained by Martin et al. (2016). While plausible values should not be used to estimate individual ability, they provide good measures on an aggregated scale. In addition, to facilitate interpretation, we have standardized the plausible values with a mean of zero and a standard deviation of one.

As discussed, peer effects appear to affect girls and boys differently. Hence, we include a variable for students’ gender. For modeling purposes and based on the data available, we use the TIMSS variable sex, separated on girls and boys, to categorize students' gender.

The proportion of girls is calculated on a classroom basis. In these proportions, the individual student is excluded to focus on peer effects. Thus, compared to girls, boys are exposed to a larger share of girls and vice versa, as Guryan et al. (2009) explain. Also in this case, we do not distinguish between sex and gender.

In evaluating peer effects on new immigrants, we focus on those immigrated when they were six years old or later and whose parents are both foreign-born. We retrieve this data from the TIMSS parental questionnaire. Although this definition restricts the sample of new immigrants, a concern discussed in Section 8.2, students with limited experience of Swedish school are distinguished, as primary school begins at age six. Furthermore, due to the recent prevalence of immigration from non-European countries discussed in Section 4, most of these new immigrants will have had limited command of the Swedish language, the language of instruction in most Swedish schools and the language of assessment. Our definition of 'new immigrant' captures this language disadvantage. A final benefit of the definition is that it resembles the one used in the Swedish Education Act (Skolverket, 2016b).

### 6.1.2 Control variables

Among student-level controls, students' age is included, as several studies suggest that age at school entry affects achievement extending from early primary school to the labor market. A relatively higher age, implying, on average, a higher level of maturity, is often found to affect positively a variety of outcomes, although also contrasting evidence exists (Lavy et al., 2012).

The age of classmates is also relevant, and hence included, as the ability differences explained by differences in age in turn may affect classroom peers. Particularly strong effects have been found of so-called 'repeaters,' who have repeated a grade level or delayed their start of primary school (Lavy et al., 2012).

Mother's and father's highest levels of completed education are included in the model as dummy variables to control for family income and potential educational aspirations that are transferred onto the child. Previous studies have shown both parents' educational level to have effects on the child's school performance (McEwan, 2003; Fruehwirth, 2013; Vigdor and Nechyba, 2007). Moreover, parents whose level of education is at any of the extremes of the distribution appear to have the largest impact on achievement (Ohinata and van Ours, 2016).

The highest levels of education of classmates' parents are also relevant to include. This aspect impacts classmates in a similar manner as discussed, and these effects may indirectly affect also the individual student through the peer environment.

Among classroom-level controls, the number of students in each class is included, as supported by substantial research. While also opposing evidence exists, tendencies for achievement to decrease with increases in class size appear particularly large in grade level four and among disadvantaged students (Angrist and Lavy, 1999; Diette and Oyelere, 2014; Burke and Sass, 2013). To account for effects that may be especially large or small at certain class sizes, also a squared term is included.

The significant effect that teachers have on their students is controlled for by including a dummy variable for teacher experience. Hanushek et al. (2003) find that student achievement is in a systematic manner only related to teachers' experience below three years in the profession, while additional years of experience or a more advanced degree do not impact students' achievement significantly. The dummy variable is therefore defined so as to include teachers with a maximum of two years of work experience in the profession.

## 6.2 Model specification

We define a contemporaneous model of student achievement on the basis of Todd and Wolpin's (2003) outline of a contemporaneous educational production function, where the achievement outcome ($A_{ija}$) of the individual student $i$ in family $j$ at age $a$, typically a test score, is explained as a function of parent-supplied ($\mathbf{F}_{ija}$), and school-supplied ($\mathbf{S}_{ija}$) inputs at age $a$, and an error term ($\epsilon_{ija}$). An adapted version of this function is shown in Equation 6.1:

$$A_{ija} = A_a(\mathbf{F}_{ija}, \mathbf{S}_{ija}) + \epsilon_{ija} \tag{6.1}$$

We further adapt this basic production function first by distinguishing not only student-specific variables, denoted $i$ in line with Todd and Wolpin (2003), but also classroom-level variables, denoted $c$, and school-level variables, subscripted $s$. We then identify the share of girls in the classroom excluding the individual student as a separate explanatory variable $Share_{ics}$. As our outcome variable, $A_{ics}$, we use standardized plausible values from the TIMSS tests in mathematics.

On the level of each student, we control for the age and gender of the individual student, whether the student is a new immigrant, an interaction term between gender and the new immigrant variable, the student's mother's and father's levels of education, mothers' and fathers' levels of education among classmates excluding the individual student, and the average age of classmates' excluding the student him- or herself. These student-level controls are included in the vector $\mathbf{X}_{ics}$.

On the classroom level, we control for the number of students in the class, as well as a squared term for this number, and for teachers' experience as discussed in Section 6.1.2. The classroom-level controls are included in the vector $\mathbf{Y}_{cs}$.

Furthermore, to control for between-school variations in unobserved factors such as socioeconomic background, we include school-level fixed effects, $\gamma_s$. We therefore choose to compare classrooms within each school, using within-school variation in the share of girls in each classroom. To do this, we restrict our sample to schools with at least two classes in grade level four. The consequences of this restriction are discussed in Section 6.3. This structure leads to the linear-in-means model presented in Equation 6.2, the results of which are presented in Table 7.1.

$$A_{ics} = \alpha_0 + \beta_1 Share_{ics} + \delta_1 \mathbf{X}_{ics} + \delta_2 \mathbf{Y}_{cs} + \gamma_s + \epsilon_{ics} \tag{6.2}$$

The heterogeneous effects of the share of girls among classroom peers are accounted for by including a pair of interaction terms. The gender ($G_{ics}$) of the individual student, and whether he or she is a new immigrant ($I_{ics}$) are interacted with the $Share_{ics}$ variable. This makes it possible to estimate separate sets of coefficients for the impact of the share of girls on girls and boys, and new immigrants and non-immigrants respectively. Regression results based on the resulting model, 6.3, are shown in Table 7.2.

$$\begin{aligned} A_{ics} = \alpha_0 + \beta_1 Share_{ics} + \beta_2 (G_{ics} \times Share_{ics}) + \beta_3 (I_{ics} \times Share_{ics}) \\ + \delta_1 \mathbf{X}_{ics} + \delta_2 \mathbf{Y}_{cs} + \gamma_s + \epsilon_{ics} \end{aligned} \tag{6.3}$$

Last, we account also for the particular effect of being both girl and new immigrant, relative to new immigrant boys, by including an additional interaction term ($G_{ics} \times I_{ics} \times Share_{ics}$). This final specification is presented in Equation 6.4, and the results based on the model are presented in Table 7.3.

$$\begin{aligned} A_{ics} = \alpha_0 + \beta_1 Share_{ics} + \beta_2 (G_{ics} \times Share_{ics}) + \beta_3 (I_{ics} \times Share_{ics}) \\ + \beta_4 (G_{ics} \times I_{ics} \times Share_{ics}) + \delta_1 \mathbf{X}_{ics} + \delta_2 \mathbf{Y}_{cs} + \gamma_s + \epsilon_{ics} \end{aligned} \tag{6.4}$$

For our hypotheses, listed in Section 3, to be supported, the corresponding coefficients must be positive and statistically significant at the 5 percent level. Specifically, for Hypothesis 1 to be supported, these requirements apply to coefficient $\beta_1$ in Equation 6.2, which indicates the average effect across demographic groups. For Hypothesis 2 and Hypothesis 3 to be supported, the same requirements apply to $\beta_2$ and $\beta_3$ in Equation 6.3, where heterogeneous effects based on gender and new immigrant status are accounted for. Last, for Hypothesis 4 to be supported, the criteria defined apply to $\beta_4$ in Equation 6.4.

## 6.3 Critical evaluation of the model

### 6.3.1 Core assumptions

Precise and unbiased estimations based on the model presented in Section 6.2 require a number of assumptions. First, the contemporaneous model will only generate correct estimates under relatively strict conditions, described in Section 2.4.2. Bias may occur, for instance, if historical inputs, such as the share of girls in classrooms before fourth grade, still affect students' outcomes in fourth grade. The use of school-level fixed effects nevertheless partially addresses this *omitted variables* problem to the extent that variation in historical inputs is substantially smaller within schools than between schools. Still, the lack of historical data is cause for caution.

School-level fixed effects also mitigate the *selection problem* discussed in Section 2.4.2. As between-school variation is controlled for, nonrandom allocation of students to schools is allowed. Within schools, between classrooms, however, student allocation is assumed to be random. This assumption will be examined further in Section 7.4.1.

Nonetheless, the possibility that unobserved aspects such as teacher instruction vary with the classroom gender composition may still confound estimates. This factor is, however, difficult to control for as no common measure of teachers is systematically related to students' outcomes other than limited teacher experience (Hanushek et al., 2003). Such effects also arguably do not bias the magnitudes of our estimates, but rather complicate their interpretation.

The *reflection problem* (Section 2.4.2) is also a problem more concerning interpretation than unbiased estimation. The use of gender rather than ability as our explanatory variable nevertheless limits this problem substantially, as gender effects cannot create the feedback loops that are possible with ability peer effects. In addition, as mathematics ability appears not to differ systematically between the genders in the sample (see Table 6.1), gender effects are unlikely to mask direct ability effects.

### 6.3.2 Model interpretation

The school-level fixed effects approach used restricts the data to schools with at least two classes in grade level four, omitting smaller schools and schools without such separation into several classes. To the extent that omitted schools differ systematically compared to the schools included in factors relevant to student achievement, the approach may limit the possibility of extrapolating estimates to evaluate single-classroom schools.

Furthermore, while fixed effects at the school level allow for distinct average test scores at each school, the coefficient estimates of the impact of the share of girls on these scores are common across schools. Although the effects of classroom gender composition may vary between the four demographic groups identified, each group is hence assumed to be affected in a similar manner regardless of school. Nevertheless, the coefficient estimates may advantageously be interpreted as average effects across schools. This assumption will receive further attention in Section 7.4.2.

Missing data points also necessitate further attention. While some sources of attrition, as discussed in Section 5.2, are random, other sources may not be. The likelihood that a student's parents have not filled in the TIMSS parental questionnaire may be higher for new immigrant students, as their parents may be less able to manage a form provided in the Swedish language. Omitted answers to some questions in the student questionnaire may also be more probable among new immigrants due to these students' own language limitations. However, as noted in Section 5.2, student participation rates after exclusions and attrition are high, resulting in relatively modest magnitudes of any potential biases related to systematic attrition.

Attention has also been assigned to potential outliers. Among test scores in mathematics, no outliers exist, partially due to the plausible values specification. The share of girls among classroom peers is in rare cases very large or small, but represents existing classroom compositions and has hence been kept unadjusted. Also students' ages are kept without modification, as ability effects exist across age spans (Lavy et al., 2012; Rivkin and Schiman, 2015), and these ages are controlled for in the model.

An outlier in the number of students per class, however, has been removed from the dataset. The particular class, with only two students, is a clear outlier and risks confounding coefficient interpretations. Especially large classes, however, are more common and are kept in the data, as it is not possible to assume that these classes are not in fact classes where peers interact on a daily basis.

### 6.3.3 Implications for the sample in use

The data limitations and classifications discussed above reduce the size of the sample used when estimating the specifications outlined in Section 6.2. Compared to the full TIMSS dataset, the average achievement of students remaining in the final sample is moderately higher.

Testing a variety of sample specifications, we find that the fixed effects estimation, where schools with only one class in grade level four are excluded, contributes to this difference, as schools with more than one class have a marginally higher average achievement in our sample. However, most of the difference in average achievement between the original dataset and the final sample used can be attributed to data attrition. In particular, by classifying students on the basis of their immigration status, we exclude students whose parents have not provided this data in TIMSS questionnaires. This exclusion explains a large share of the difference. This attrition, therefore, appears to be systematic.

Although omitted answers in teacher questionnaires also contribute to a relatively large proportion of the total attrition, this type of attrition appears not to be systematic and hence does not influence our estimations substantially.

Figure 6.1 provides a comparison of the distribution of standardized test scores in mathematics, based on estimated plausible values, for the original TIMSS dataset and the subset sample used in this paper. The kernel density plot displays the proportion of students receiving test scores within different intervals in a similar manner to a histogram.

Figure 6.1: Comparison of distributions of test scores in mathematics

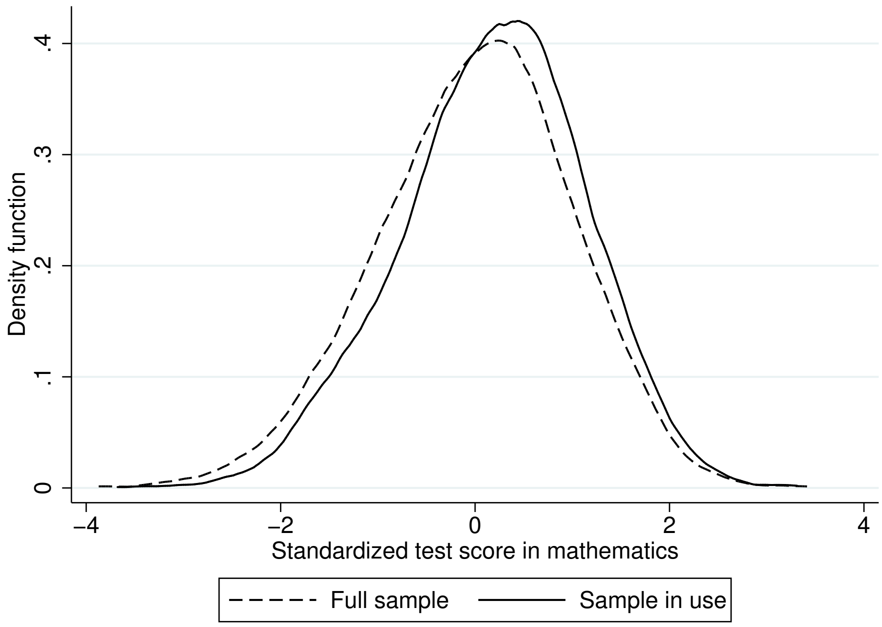


Table 6.1 displays test scores in mathematics, in standardized plausible values, for the used sample. The scores are indicated for each of four mutually exclusive and

jointly exhaustive demographic groups, such as new immigrant girls, as well as for the whole sample in use. Hence, it is possible to observe that gender differences are small, while differences between new immigrants and non-immigrants (defined as all students who are not defined as 'new immigrants') are notable.

Table 6.1: Mathematics achievement by demographic group

| | Boys | Girls | Total | n |
|---|---|---|---|---|
| New immigrants | -0.752 | -0.715 | -0.733 | 47 |
| | (0.793) | (0.760) | (0.768) | |
| Non-immigrants | 0.200 | 0.204 | 0.202 | 2362 |
| | (0.933) | (0.955) | (0.944) | |
| Total | 0.181 | 0.186 | 0.184 | 2409 |
| | (0.940) | (0.960) | (0.950) | |
| n | 1181 | 1228 | 2409 | |

Standardized mean plausible values, with standard deviations in parentheses. n = Number of observations.

Table 6.2 further provides summary statistics for the remaining variables of interest, as defined in Section 6.1.1. Corresponding statistics for remaining control variables are included in Appendix A.1.

Table 6.2: Summary statistics

| | Mean | Std | Min | Max |
|---|---|---|---|---|
| Girl share | 0.500 | 0.112 | 0.185 | 1.000 |
| Girl | 0.510 | 0.500 | 0.000 | 1.000 |
| New immigrant | 0.020 | 0.138 | 0.000 | 1.000 |
| Girl × New immigrant | 0.010 | 0.099 | 0.000 | 1.000 |

# 7 Results

The estimation of the models presented in Section 6.2 proceeds in three steps. First, we apply the linear-in-means model presented in Equation 6.2, where we estimate the effect of the share of girls in the classroom excluding an individual student, in decimal form, on the standardized TIMSS mathematics test score of that student.

Second, we estimate Equation 6.3, where we use interaction terms to separate the heterogeneous effects of girl share depending on students' gender and whether they are new immigrants.

Third, we apply Equation 6.4, which features also the interaction effect of girl share on new immigrant girls distinct from new immigrant boys.

In all estimations, we proceed in steps, first (1) with only the main explanatory variables included, then (2) with school-level fixed effects to measure only within-school variation (further discussed in Section 6.3), and finally (3) with both school-level fixed effects and the controls outlined in Section 6.2, including parental education as a measure of socioeconomic background. All three resulting tables are also presented in extended versions with all control variables displayed in Appendix A.

## 7.1 Linear-in-means

The results of the linear-in-means Equation 6.2 are presented in Table 7.1. They indicate that the impact of the share of girls among classroom peers on student achievement is statistically insignificant. The naive estimation (1), where test scores are regressed only on the share of girls, also explains very little of the total variation in test scores, as evidenced by the low $R^2$ value. However, including school-level fixed effects, as has been done in estimation (2), significantly improves the model fit, suggesting that between-school variation is important to control for. Finally, control variables further add to the estimation power of the model. Specifically, students' age, parental education, and whether the student is a new immigrant are significant.

Table 7.1: Regression of share of girls in the classroom on mathematics test scores

| | (1) | (2) | (3) |
|---|---|---|---|
| Girl share | 0.104 | -0.159 | 0.015 |
| | (0.349) | (0.503) | (0.490) |
| School-level fixed effects | No | Yes | Yes |
| Controls | No | No | Yes |
| Number of observations | 2409 | 2409 | 2409 |
| Number of classroom clusters | 63 | 63 | 63 |
| $R^2$ | 0.000 | 0.150 | 0.240 |

Robust standard errors, adjusted for clustering at the school level, are presented in parentheses.
*** Significant at the 1% level; ** Significant at the 5% level; * Significant at the 10% level

Although marginally positive, the coefficient estimate of *Girl share* in the complete specification (3) is not statistically significant. Hence, we reject Hypothesis 1, that a large share of girls among classroom peers impacts positively on students' test scores in mathematics.

## 7.2 Heterogeneous effects

The results of the heterogeneous specification in Equation 6.3 are presented in Table 7.2. The $R^2$ values indicate that this model has greater explanatory power than the previous, and although the $R^2$ value is relatively low in the naive model (1) without school-level fixed effects and controls, it appears that the share of girls among classroom peers is one of many explanatory factors behind student achievement in mathematics for certain groups.

However, we do not find statistically significant coefficients in the most complete specification (3), suggesting that there is no positive average effect of girl share neither on boys nor on girls, and neither on new immigrants nor on non-immigrants, at least not viewed as distinct groups without allowing for intersectional effects. Hence, we reject Hypothesis 2 and Hypothesis 3 concerning the effects on girls and new immigrants, respectively.

Table 7.2: Regression of share of girls in the classroom on mathematics test scores by group without intersection

| | (1) | (2) | (3) |
|---|---|---|---|
| Girl share | 0.139 | -0.123 | -0.326 |
| | (0.336) | (0.509) | (0.518) |
| Girl × (Girl share) | 0.052 | 0.023 | 0.623 |
| | (0.076) | (0.082) | (0.431) |
| New immigrant × (Girl share) | -1.727*** | -1.286*** | -0.165 |
| | (0.304) | (0.199) | (1.163) |
| School-level fixed effects | No | Yes | Yes |
| Controls | No | No | Yes |
| Number of observations | 2409 | 2409 | 2409 |
| Number of classroom clusters | 63 | 63 | 63 |
| $R^2$ | 0.018 | 0.159 | 0.241 |

Robust standard errors, adjusted for clustering at the school level, are presented in parentheses.
*** Significant at the 1% level; ** Significant at the 5% level; * Significant at the 10% level

## 7.3 Intersectional effects

In Table 7.3, results of the intersectional specification in Equation 6.4, with intersectional effects on new immigrant girls, are presented. Based on estimation (3), we find that the effect on new immigrant girls, relative to new immigrant boys, is statistically significant, positive, and relatively large. Hence, we find support for Hypothesis 4.

The effect of being both girl and new immigrant, in contrast to being a new immigrant boy, is 3.6 standard deviations for a change from 0 percent to 100 percent girl share, or 0.036 standard deviations for each percentage point change. The total effect for new immigrant girls is, however, smaller. A shift from 0 percent to 100 percent girls among classroom peers, with these estimates, noting that all component parts are

Table 7.3: Regression of share of girls in the classroom on mathematics test scores by group with intersection

| | (1) | (2) | (3) |
|---|---|---|---|
| Girl share | 0.140 | -0.123 | -0.280 |
| | (0.335) | (0.509) | (0.518) |
| Girl × (Girl share) | 0.043 | 0.023 | 0.558 |
| | (0.078) | (0.084) | (0.439) |
| New immigrant × (Girl share) | -1.971*** | -1.278*** | -1.772 |
| | (0.401) | (0.323) | (1.350) |
| Girl × New immigrant × (Girl share) | 0.424 | -0.014 | 3.601** |
| | (0.471) | (0.460) | (1.634) |
| School-level fixed effects | No | Yes | Yes |
| Controls | No | No | Yes |
| Number of observations | 2409 | 2409 | 2409 |
| Number of classroom clusters | 63 | 63 | 63 |
| $R^2$ | 0.018 | 0.159 | 0.242 |

Robust standard errors, adjusted for clustering at the school level, are presented in parentheses.
*** Significant at the 1% level; ** Significant at the 5% level; * Significant at the 10% level

not significant, would imply a test score increase of $(-0.280+0.558-1.772+3.601=)$ 2.107 standard deviations, or 0.021 standard deviations per percentage point of girls. These results suggest that the new immigrant girl group is affected beyond what can be explained by their immigrant status alone—that students at the intersection of the girl and new immigrant identities are particularly susceptible to the gender peer influences studied.

## 7.4 Sensitivity analysis

To evaluate the sensitivity of our estimates to variations in model specification and the underlying data, additional regressions and tests will be performed. First, the question of whether students are randomly allocated to classes within schools will be addressed. Second, we examine potential variations between schools in the effects of girl share on the different demographic groups.

### 7.4.1 Random allocation of students to classes

In order to avoid the selection problem described in Section 2.4.2, random allocation of students to classes within schools has to be assumed. To reinforce this assumption, a test of whether each classroom gender composition is significantly different from the school gender composition is conducted. We reject the hypothesis that students are randomly allocated with regards to gender for nine schools (14.3 percent of our initial school sample) at a significance level of 10 percent. A reestimated regression excluding these schools is included in Table 7.4.

Table 7.4: Regression of share of girls in the classroom on mathematics test scores by group with intersection (restricted sample)

| | (1) | (2) | (3) |
|---|---|---|---|
| Girl share | 0.211 | -0.152 | -0.513 |
| | (0.333) | (0.517) | (0.457) |
| Girl × (Girl share) | 0.016 | -0.021 | 0.401 |
| | (0.075) | (0.077) | (0.418) |
| New immigrant × (Girl share) | -2.157*** | -1.393*** | -2.015 |
| | (0.385) | (0.319) | (1.419) |
| Girl × New immigrant × (Girl share) | 0.654 | 0.204 | 3.628** |
| | (0.431) | (0.412) | (1.696) |
| School-level fixed effects | No | Yes | Yes |
| Controls | No | No | Yes |
| Number of observations | 2149 | 2149 | 2149 |
| Number of classroom clusters | 54 | 54 | 54 |
| $R^2$ | 0.021 | 0.155 | 0.241 |

Robust standard errors, adjusted for clustering at the school level, are presented in parentheses.
*** Significant at the 1% level; ** Significant at the 5% level; * Significant at the 10% level

The results, however, show only small deviations from the estimates found using the initial school sample (Table 7.3) and no effect on the statistical significance of the included variables. Furthermore, as statistical indication of nonrandom allocation is expected in 10 percent of schools due to the 10 percent significance level, nonrandom allocation appears moderate. Hence, it can be concluded that our estimates are relatively robust to the level of nonrandom allocation of girls between classrooms that appears present in Swedish schools.

### 7.4.2 Variations in effects between schools

As described in Section 6.3.2, our estimates of the heterogeneous effects of the share of girls among classroom peers on mathematics scores for new immigrant and non-immigrant girls and boys should be interpreted as average effects across schools. Nonetheless, in this section, we allow for distinct effects between schools and test for the significance of such disparities in the effects in different schools.

Although the between-school differences are statistically significant, as measured by an F-test, the inclusion of fixed effects interactions to regard them in the model does not result in noteworthy increases in the explanatory power (the $R^2$ value) of the model. We are therefore able to conclude that the average effects across schools express the influences on the four demographic groups studied well.

# 8 Discussion

## 8.1 Interpretation of results

Before allowing for heterogeneous effects, the results of the regressions based on the linear-in-means model with all control variables included show no significant effect of the share of girls in the classroom on average mathematics test scores. Hypothesis 1 of this paper can therefore be rejected. Furthermore, the results of the regressions using the first extension of the model, allowing for heterogeneous effects based on gender and immigration status, show no significant effect on neither girls' nor new immigrants' test scores in mathematics. Hence, Hypothesis 2 and Hypothesis 3 are also rejected.

In contrast, the results of the regression on the final model show that the share of girls in the classroom has a statistically significant and positive effect on individuals who are at the intersection of the two identities girl and new immigrant, in contrast to new immigrant boys. Hypothesis 4 of this paper, namely that new immigrant girls see greater positive effects of a higher share of girls in the classroom than do new immigrant boys, thus appears supported. The magnitude of the estimated effect is also economically significant. For instance, changing one boy for one girl in a classroom of twenty students that initially has an equal amount of boys and girls would result in a test score increase of 0.11 standard deviations for students who are both new immigrants and girls.

The evidence that new immigrant girls are especially affected is in line with the findings in intersectionality literature, although the additional effect of interacting identities, beyond a simple aggregate, has not been evaluated with our econometric methods. Intersectionality theory, nonetheless, predicts that including the additional effect is likely to further increase the total effects observed. This would imply that the direction of our estimates is correct while the magnitudes are downward biased.

The first results of this paper, which indicate that the share of girls in the classroom has no effect on average class achievement, are not in line with previous research. This can indicate two things. First, the absence of a gender gap in mathematics achievement among Swedish elementary school students makes the classroom gender composition less significant. In other words, in a Swedish context, a change in gender composition does not mask an actual change in ability but instead the two effects can be separated. Nonetheless, it is worth noting that previous research shows that even in countries in which boys are relatively higher achieving in mathematics, increasing the share of girls affects classroom average ability positively. This suggests that the classroom share of girls may be significant *per se*, and not only through ability differences.

Second, the effect of gender composition could be less significant in Sweden if gender norms affecting mathematics achievement are relatively less present in this context. If gender norms to a smaller extent are affecting mathematics achievement, increasing the number of girls in a classroom should not directly affect the mathematics test scores.

Furthermore, our evidence that new immigrant girls are more susceptible to differences in classroom gender composition is somewhat in line with previous peer effect literature, although an intersectional perspective has to our knowledge not been recognized in this literature. The mechanisms behind new immigrant girls' higher susceptibility to gender composition can be analyzed by considering the effect of internalized gender norms. As previous literature indicates (see Section 2.1.3), gender norms appear to affect student achievement in other countries and contexts. With experience from these other countries, new immigrant girls can therefore encounter a stronger contrast between their internalized gender norms and the norms they face in Swedish schools—a contrast that is increasing with the number of girls in the classroom not conforming to the internalized gender conceptions. As a result, the internalized gender norms among new immigrant girls may weaken when increasing the share of girls in the classroom, which allows for a positive effect on their mathematics achievement.

## 8.2 Method and data considerations

Data exclusions and attrition may complicate attempts to interpret the estimated coefficients. In particular, as discussed in Section 5.2, recent immigrants unable to comprehend enough Swedish to participate in TIMSS are excluded. Furthermore, among the students excluded as their parental questionnaire is not filled in, new immigrants may be over-represented, as the Swedish language abilities of their parents are often limited. These factors are also likely to contribute to the small sample size of new immigrants, which in turn makes coefficients that are economically significant unlikely to be statistically significant. In our case, this reasoning may explain the lack of statistical significance of the effects of girl share on new immigrant boys.

Nonetheless, if the observed peer effects originate in differences in internalized gender norms between new immigrants and other students, there is little reason to suspect that these effects do not apply also to the excluded groups. In contrast, the most recently immigrated girls may have been less affected by the relatively weaker gender norms in Sweden and hence experience greater effects of girl share among peers. If so, our estimates for new immigrant girls are downward biased.

Furthermore, the specification of measurement variables is likely to be important to the estimates found. TIMSS test scores are a beneficial measure of ability because they allow for comparisons across countries and over time, and because the tests are corrected externally, avoiding direct teacher biases. However, a single test is a noisy measurement, and other outcome variables, such as classroom participation, may indicate additional mechanisms of peer influence.

The fixed effects specification, where between-school average effects on the four demographic groups used are estimated (see Section 6.3.2), provides an overview of effects which may be further examined. With a systematic categorization of schools into more homogeneous types, it may be possible to analyze disparities in gender peer effects between school types. Such an analysis could provide further insights into the mechanisms behind the observed effects.

## 8.3 Policy implications

From a policy view, heterogeneous effects allow for non-zero sum implications. In linear-in-means specifications, reallocation of resources can only facilitate the achievement of one group by inhibiting that of another group. Heterogeneous effects instead make possible that one group may benefit from resource redistribution while other groups are unaffected. Hence, in our context, targeted policies can aid the achievement of new immigrant girls and provide aggregate benefits to Swedish fourth-graders even without an increase in total resources.

Our findings suggest that internalized gender norms among new immigrant girls, which affect mathematics achievement negatively, may be weakened if these girls interact with a greater range of girls who do not conform to one stereotype. With further analysis of whether the observed effects differ in magnitude between schools, policy measures to facilitate this interaction between girls may be targeted towards the schools where these effects are most significant.

# 9 Conclusion

This paper has contributed an intersectional dimension to the study of how classroom gender composition impacts mathematics achievement for different groups of students depending on their characteristics. In contrast to many studies using non-Swedish populations, we have found no significant average effects of girl share on Swedish fourth-grade students. However, with regards to new immigrant girls, relative to new immigrant boys, we have provided evidence of a statistically significant and relatively large positive effect of increasing the share of girls in the classroom—an effect that cannot be explained by their gender or immigrant status separately.

We have argued that these results may be explained by relatively weak gender norms with regards to mathematics achievement among most students in Swedish grade four, and stronger internalized gender norms among new immigrant girls. These internalized norms may weaken the performance of new immigrant girls in mathematics but are themselves weakened through the interaction with a range of girls who do not all conform to one stereotype.

Based on these findings, we conclude that addressing the specific context of new immigrant girls and mitigating gender norms among them may effectively enhance primary school achievement in mathematics. In addition, intersectional analyses and targeted initiatives also in other contexts—with other groups, outcome measures, and countries—may provide for substantial improvements in outcomes.

## 9.1 Future work

In future work, a heterogeneous analysis of how the effects of the proportion of girls among classroom peers on the different groups depend on school type would be informative. Such an analysis, categorizing schools with regards to dimensions such as geographic area or amount of funding, would allow for improved targeting of measures to improve student achievement.

A further investigation into the mechanisms behind the effects observed would also be beneficial. More specifically, separating the effects of achievement norms, disruption, and other explanatory factors would inform both research and policy.

In addition, the specification of peer groups may be evaluated. Although classes have been shown to capture peer effects better than whole grade levels in school, a more narrow definition, such as bench neighbors in class, may yield additional insights.

Lastly, the interpretation of estimates would be aided by the conducting of a similar study with a larger dataset. Registry data, where more background variables that also account for historical inputs can be controlled for, would additionally facilitate more precise estimations.

# Bibliography


Acker, J. (2006). Inequality regimes: Gender, class, and race in organizations, *Gender and Society* **20**(4): 441–464.

Ammermueller, A. and Pischke, J.-S. (2009). Peer effects in European primary schools: Evidence from the Progress in International Reading Literacy Study, *Journal of Labor Economics* **27**(3): 315–348.

Angrist, J. D. and Lavy, V. (1999). Using Maimonides' Rule to estimate the effect of class size on scholastic achievement, *Quarterly Journal of Economics* **114**(2): 533–575.

Booij, A. S., Leuven, E., and Oosterbeek, H. (2016). Ability peer effects in university: Evidence from a randomized experiment, *Review of Economic Studies* **84**(2): 547–578.

Browne, I. and Misra, J. (2003). The intersection of gender and race in the labor market, *Annual Review of Sociology* **29**: 487–513.

Brunello, G. and Rocco, L. (2013). The effect of immigration on the school performance of natives: Cross country evidence using PISA test scores, *Economics of Education Review* **32**(1): 234–246.

Burke, M. A. and Sass, T. R. (2013). Classroom peer effects and student achievement, *Journal of Labor Economics* **31**(1): 51–82.

Coleman, J. S. (1966). *Equality of Educational Opportunity*, Department of Health, Education, and Welfare, Office of Education, Washington D.C.

Cooley, J. (2010). Classroom peer effects, *in* S. N. Durlauf and L. E. Blum (eds), *The New Palgrave Dictionary of Economics*, Palgrave Macmillan. Available from: http://www.dictionaryofeconomics.com/article?id=pde2010_C000613 [Accessed March 3, 2017].

Crenshaw, K. (1989). Demarginalizing the intersection of race and sex: A black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics, *University of Chicago Legal Forum* **1989**(1): 139–167.

Diette, T. M. and Oyelere, R. U. (2014). Gender and race heterogeneity: The impact of students with limited English on native students' performance, *American Economic Review* **104**(5): 412–417.

Entorf, H. and Lauk, M. (2008). Peer effects, social multipliers and migrants at school: An international comparison, *Journal of Ethnic and Migration Studies* **34**(4): 633–654.

Fertig, M. (2003). Who's to blame? The determinants of German students' achievement in the PISA 2000 study, *Report 739*, Institute for the Study of Labor, Bonn.

Foster, G. (2006). It's not your peers, and it's not your friends: Some progress toward understanding the educational peer effect mechanism, *Journal of Public Economics* **90**(8–9): 1455–1475.

Fruehwirth, J. C. (2013). Identifying peer achievement spillovers: Implications for desegregation and the achievement gap, *Quantitative Economics* **4**(1): 85–124.

Gottfried, M. A. and Graves, J. (2014). Peer effects and policy: The relationship between classroom gender composition and student achievement in early elementary school, *B.E. Journal of Economic Analysis and Policy* **14**(3): 937–977.

Gould, E. D., Lavy, V., and Paserman, M. D. (2009). Does immigration affect the long-term educational outcomes of natives? Quasi-experimental evidence, *Economic Journal* **119**(540): 1243–1269.

Griffith, A. L. and Rask, K. N. (2014). Peer effects in higher education: A look at heterogeneous impacts, *Economics of Education Review* **39**: 65–77.

Guryan, J., Kroft, K., and Notowidigdo, M. J. (2009). Peer effects in the workplace: Evidence from random groupings in professional golf tournaments, *American Economic Journal: Applied Economics* **1**(4): 34–68.

Hanushek, E. A., Kain, J. F., Markman, J. M., and Rivkin, S. G. (2003). Does peer ability affect student achievement?, *Journal of Applied Econometrics* **18**(5): 527–544.

Hoxby, C. M. (2000). Peer effects in the classroom: Learning from gender and race variation. NBER Working Paper 7867. National Bureau of Economic Research, Cambridge, MA. Available from: http://www.nber.org/papers/w7867 [Accessed February 15, 2017].

Hoxby, C. M. and Weingarth, G. (2005). Taking race out of the equation: School reassignment and the structure of peer effects. NBER Conference Paper. National Bureau of Economic Research, Cambridge, MA. Available from: https://www.pausd.org/sites/default/files/pdf-faqs/attachments/TakingRaceOutOfTheEquation.pdf [Accessed February 15, 2017].

IEA (2016). TIMSS 2015 assessment. Available from: http://rms.iea-dpc.org [Accessed March 9, 2017].

Inzlicht, M. and Ben-Zeev, T. (2000). A threatening intellectual environment: Why females are susceptible to experiencing problem-solving deficits in the presence of males, *Psychological Science* **11**(5): 365–371.

Jensen, P. and Rasmussen, A. W. (2011). The effect of immigrant concentration in schools on native and immigrant children's reading and math skills, *Economics of Education Review* **30**(6): 1503–1515.

Kang, C. (2007). Academic interactions among classroom peers: A cross-country comparison using TIMSS, *Applied Economics* **39**(12): 1531–1544.

Lavy, V., Paserman, M. D., and Schlosser, A. (2012). Inside the black box of ability peer effects: Evidence from variation in the proportion of low achievers in the classroom, *Economic Journal* **122**(559): 208–237.

Lavy, V. and Schlosser, A. (2011). Mechanisms and impacts of gender peer effects at school, *American Economic Journal: Applied Economics* **3**(2): 1–33.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem, *Review of Economic Studies* **60**(3): 531–542.

Manski, C. F. (2000). Economic analysis of social interactions, *Journal of Economic Perspectives* **14**(3): 115–136.

Martin, M. O., Mullis, I. V. S., and Hooper, M. (eds) (2016). *Methods and Procedures in TIMSS 2015*, TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College, Chestnut Hill, MA.

McEwan, P. J. (2003). Peer effects on student achievement: Evidence from Chile, *Economics of Education Review* **22**(2): 131–141.

Muzzatti, B. and Agnoli, F. (2007). Gender and mathematics: Attitudes and stereotype threat susceptibility in Italian children, *Developmental Psychology* **43**(3): 747–759.

OECD (2015). Improving schools in Sweden: An OECD perspective, *Report*, OECD Publishing, Paris.

Ohinata, A. and van Ours, J. C. (2016). Quantile peer effects of immigrant children at primary schools, *Labour* **30**(2): 135–157.

Oosterbeek, H. and van Ewijk, R. (2014). Gender peer effects in university: Evidence from a randomized experiment, *Economics of Education Review* **38**: 51–63.

Raitano, M. and Vona, F. (2013). Peer heterogeneity, school tracking and students' performances: Evidence from PISA 2006, *Applied Economics* **45**(32): 4516–4532.

Rivkin, S. G. and Schiman, J. C. (2015). Instruction time, classroom quality, and academic achievement, *Economic Journal* **125**(588): F425–F448.

Sacerdote, B. (2001). Peer effects with random assignment: Results for Dartmouth roommates, *Quarterly Journal of Economics* **116**(2): 681–704.

Sacerdote, B. (2011). Peer effects in education: How might they work, how big are they and how much do we know thus far?, *in* E. A. Hanushek, S. Machin, and L. Woessmann (eds), *Handbook of the Economics of Education*, Vol. 3, North-Holland, Amsterdam, pp. 249–277.

SCB (2013). Andelen barn som invandrar till Sverige ökar. Available from: http://www.scb.se/sv_/Hitta-statistik/Artiklar/Andelen-barn-som-invandrar-till-Sverige-okar/ [Accessed May 9, 2017].

Schneeweis, N. and Winter-Ebmer, R. (2007). Peer effects in Austrian schools, *Empirical Economics* **32**(2-3): 387–409.

Schneeweis, N. and Zweimüller, M. (2012). Girls, girls, girls: Gender composition and female school choice, *Economics of Education Review* **31**(4): 482–500.

Skolverket (2011). Resursfördelning till grundskolan – rektorers perspektiv, *Report*, Skolverket, Stockholm.

Skolverket (2013). Kommunernas resursfördelning till grundskolor, *Report*, Skolverket, Stockholm.

Skolverket (2015). Skolverkets lägesbedömning 2015, *Report*, Skolverket, Stockholm.

Skolverket (2016a). Grundskola. Available from: https://www.skolverket.se/skolformer/grundskoleutbildning/grundskola/grundskola-1.196222 [Accessed May 6, 2017].

Skolverket (2016b). Nyinvandrade elever i grundskolan, *Report*, Skolverket, Stockholm.

Sund, K. (2009). Estimating peer effects in Swedish high school using school, teacher, and student fixed effects, *Economics of Education Review* **28**(3): 329–336.

Todd, P. E. and Wolpin, K. I. (2003). On the specification and estimation of the production function for cognitive achievement, *Economic Journal* **113**(485): F3–F33.

Vigdor, J. and Nechyba, T. (2007). Peer effects in North Carolina public schools, *in* L. Woessmann and P. E. Peterson (eds), *Schools and the Equal Opportunity Problem*, MIT Press, Cambridge, MA, pp. 73–101.

Whitmore, D. (2005). Resource and peer impacts on girls' academic achievement: Evidence from a randomized experiment, *American Economic Review* **95**(2): 199–203.

Wolter, I. B. and Hannover, B. (2016). Gender role self-concept at school start and its impact on academic self-concept and performance in mathematics and reading, *European Journal of Developmental Psychology* **13**(6): 681–703.

Zimmerman, D. J. (2003). Peer effects in academic outcomes: Evidence from a natural experiment, *Review of Economics and Statistics* **85**(1): 9–23.

# A Appendix

## A.1 Summary statistics

Table A.1: Summary statistics for remaining control variables

| | Mean | Std | Min | Max |
|---|---|---|---|---|
| Student's age | 10.754 | 0.298 | 9.67 | 13.42 |
| Mother: Did not go to school | 0.007 | 0.084 | 0.000 | 1.000 |
| Mother: Some primary school | 0.038 | 0.191 | 0.000 | 1.000 |
| Mother: Upper secondary school | 0.191 | 0.393 | 0.000 | 1.000 |
| Mother: Bachelor's or equivalent | 0.242 | 0.428 | 0.000 | 1.000 |
| Mother: Post secondary, less than Bachelor's | 0.335 | 0.472 | 0.000 | 1.000 |
| Mother: Postgraduate degree | 0.182 | 0.386 | 0.000 | 1.000 |
| Mother: Not applicable | 0.005 | 0.073 | 0.000 | 1.000 |
| Father: Did not go to school | 0.006 | 0.079 | 0.000 | 1.000 |
| Father: Some primary school | 0.056 | 0.230 | 0.000 | 1.000 |
| Father: Upper secondary school | 0.276 | 0.447 | 0.000 | 1.000 |
| Father: Post secondary, less than Bachelor's | 0.263 | 0.440 | 0.000 | 1.000 |
| Father: Bachelor's or equivalent | 0.189 | 0.392 | 0.000 | 1.000 |
| Father: Postgraduate degree | 0.199 | 0.400 | 0.000 | 1.000 |
| Father: Not applicable | 0.011 | 0.103 | 0.000 | 1.000 |
| Classmates' mothers' share: Did not go to school | 0.011 | 0.043 | 0.000 | 0.444 |
| Classmates' mothers' share: Some primary school | 0.033 | 0.044 | 0.000 | 0.222 |
| Classmates' mothers' share: Upper secondary school | 0.189 | 0.130 | 0.000 | 1.000 |
| Classmates' mothers' share: Post secondary, less than Bachelor's | 0.193 | 0.096 | 0.000 | 0.476 |
| Classmates' mothers' share: Bachelor's or equivalent | 0.333 | 0.130 | 0.000 | 0.750 |
| Classmates' mothers' share: Postgraduate degree | 0.180 | 0.138 | 0.000 | 0.720 |
| Classmates' mothers' share: Not applicable | 0.005 | 0.022 | 0.000 | 0.222 |
| Classmates' fathers' share: Did not go to school | 0.007 | 0.028 | 0.000 | 0.286 |
| Classmates' fathers' share: Some primary school | 0.049 | 0.060 | 0.000 | 0.333 |
| Classmates' fathers' share: Upper secondary school | 0.273 | 0.148 | 0.000 | 0.727 |
| Classmates' fathers' share: Post secondary, less than Bachelor's | 0.210 | 0.097 | 0.000 | 0.500 |
| Classmates' fathers' share: Bachelor's or equivalent | 0.191 | 0.112 | 0.000 | 0.545 |
| Classmates' fathers' share: Postgraduate degree | 0.194 | 0.152 | 0.000 | 0.760 |
| Classmates' fathers' share: Not applicable | 0.018 | 0.037 | 0.000 | 0.286 |
| Class size | 32.402 | 13.511 | 7.000 | 62.000 |
| Teacher experience $< 3$ years | 0.115 | 0.320 | 0.000 | 1.000 |
| Classmates' average age | 10.761 | 0.075 | 10.562 | 11.039 |

## A.2 Linear-in-means

Extended version of Table 7.1.

Table A.2: Regression of share of girls in the classroom on mathematics test scores

| | (1) | (2) | (3) |
|---|---|---|---|
| Girl share | 0.104 | -0.159 | 0.015 |
| | (0.349) | (0.503) | (0.490) |
| Student's age | | | 0.320*** |
| | | | (0.080) |
| Girl | | | -0.005 |
| | | | (0.044) |
| New immigrant | | | -0.564** |
| | | | (0.221) |
| Girl × New immigrant | | | -0.063 |
| | | | (0.264) |
| Mother: Did not go to school | | | 0.098 |
| | | | (0.634) |
| Mother: Some primary school | | | 0.302 |
| | | | (0.396) |
| Mother: Upper secondary school | | | 0.454 |
| | | | (0.417) |
| Mother: Post secondary, less than Bachelor's | | | 0.517 |
| | | | (0.413) |
| Mother: Bachelor's or equivalent | | | 0.691 |
| | | | (0.422) |
| Mother: Postgraduate degree | | | 0.818* |
| | | | (0.433) |
| Mother: Not applicable | | | 0.000 |
| | | | (.) |
| Father: Did not go to school | | | -0.188 |
| | | | (0.410) |
| Father: Some primary school | | | 0.027 |
| | | | (0.230) |
| Father: Upper secondary school | | | 0.148 |
| | | | (0.229) |
| Father: Post secondary, less than Bachelor's | | | 0.177 |
| | | | (0.219) |
| Father: Bachelor's or equivalent | | | 0.388 |
| | | | (0.244) |
| Father: Postgraduate degree | | | 0.437* |
| | | | (0.241) |
| Father: Not applicable | | | 0.000 |
| | | | (.) |
| Classmates' mothers' share: Did not go to school | | | -0.185 |
| | | | (1.096) |
| Classmates' mothers' share: Some primary school | | | 0.484 |
| | | | (1.926) |
| Classmates' mothers' share: Upper secondary school | | | -0.533 |
| | | | (1.354) |
| Classmates' mothers' share: Post secondary, less than Bachelor's | | | -0.035 |
| | | | (1.868) |
| Classmates' mothers' share: Bachelor's or equivalent | | | -0.023 |

| | | | |
|---|---|---|---|
| | | | (1.541) |
| Classmates' mothers' share: Postgraduate degree | | | 0.630 |
| | | | (1.931) |
| Classmates' mothers' share: Not applicable | | | 0.212 |
| | | | (1.730) |
| Classmates' fathers' share: Did not go to school | | | -1.213 |
| | | | (1.413) |
| Classmates' fathers' share: Some primary school | | | 2.593*** |
| | | | (0.919) |
| Classmates' fathers' share: Upper secondary school | | | 0.391 |
| | | | (0.583) |
| Classmates' fathers' share: Post secondary, less than Bachelor's | | | -0.149 |
| | | | (0.713) |
| Classmates' fathers' share: Bachelor's or equivalent | | | 0.718 |
| | | | (0.720) |
| Classmates' fathers' share: Postgraduate degree | | | -0.077 |
| | | | (0.799) |
| Classmates' fathers' share: Not applicable | | | -1.824 |
| | | | (1.148) |
| Class size | | | 0.006 |
| | | | (0.022) |
| (Class size)$^2$ | | | 0.000 |
| | | | (0.000) |
| Teacher experience $<$ 3 years | | | -0.073 |
| | | | (0.052) |
| Classmates' average age | | | 0.639 |
| | | | (0.528) |
| School-level fixed effects | No | Yes | Yes |
| Controls | No | No | Yes |
| Number of observations | 2409 | 2409 | 2409 |
| Number of classroom clusters | 63 | 63 | 63 |
| $R^2$ | 0.000 | 0.150 | 0.240 |

Robust standard errors, adjusted for clustering at the school level, are presented in parentheses.
*** Significant at the 1% level; ** Significant at the 5% level; * Significant at the 10% level

## A.3 Heterogeneous effects

Extended version of Table 7.2.

Table A.3: Regression of share of girls in the classroom on mathematics test scores by group without intersection

| | (1) | (2) | (3) |
|---|---|---|---|
| Girl share | 0.139 | -0.123 | -0.326 |
| | (0.336) | (0.509) | (0.518) |
| Girl × (Girl share) | 0.052 | 0.023 | 0.623 |
| | (0.076) | (0.082) | (0.431) |
| New immigrant × (Girl share) | -1.727*** | -1.286*** | -0.165 |
| | (0.304) | (0.199) | (1.163) |
| Student's age | | | 0.319*** |
| | | | (0.080) |
| Girl | | | -0.315 |
| | | | (0.225) |
| New immigrant | | | -0.484 |
| | | | (0.669) |
| Girl × New immigrant | | | -0.060 |
| | | | (0.250) |
| Mother: Did not go to school | | | 0.108 |
| | | | (0.626) |
| Mother: Some primary school | | | 0.301 |
| | | | (0.389) |
| Mother: Upper secondary school | | | 0.452 |
| | | | (0.412) |
| Mother: Post secondary, less than Bachelor's | | | 0.513 |
| | | | (0.409) |
| Mother: Bachelor's or equivalent | | | 0.687 |
| | | | (0.418) |
| Mother: Postgraduate degree | | | 0.816* |
| | | | (0.428) |
| Mother: Not applicable | | | 0.000 |
| | | | (.) |
| Father: Did not go to school | | | -0.209 |
| | | | (0.396) |
| Father: Some primary school | | | 0.026 |
| | | | (0.224) |
| Father: Upper secondary school | | | 0.147 |
| | | | (0.222) |
| Father: Post secondary, less than Bachelor's | | | 0.178 |
| | | | (0.212) |
| Father: Bachelor's or equivalent | | | 0.390 |
| | | | (0.236) |
| Father: Postgraduate degree | | | 0.434* |
| | | | (0.233) |
| Father: Not applicable | | | 0.000 |
| | | | (.) |
| Classmates' mothers' share: Did not go to school | | | -0.354 |
| | | | (1.103) |
| Classmates' mothers' share: Some primary school | | | 0.343 |
| | | | (1.925) |

| | | | |
|---|---|---|---|
| Classmates' mothers' share: Upper secondary school | | | -0.605 |
| | | | (1.354) |
| Classmates' mothers' share: Post secondary, less than Bachelor's | | | -0.189 |
| | | | (1.878) |
| Classmates' mothers' share: Bachelor's or equivalent | | | -0.178 |
| | | | (1.550) |
| Classmates' mothers' share: Postgraduate degree | | | 0.489 |
| | | | (1.937) |
| Classmates' mothers' share: Not applicable | | | -0.017 |
| | | | (1.671) |
| Classmates' fathers' share: Did not go to school | | | -1.267 |
| | | | (1.412) |
| Classmates' fathers' share: Some primary school | | | 2.634*** |
| | | | (0.908) |
| Classmates' fathers' share: Upper secondary school | | | 0.444 |
| | | | (0.577) |
| Classmates' fathers' share: Post secondary, less than Bachelor's | | | -0.047 |
| | | | (0.712) |
| Classmates' fathers' share: Bachelor's or equivalent | | | 0.828 |
| | | | (0.725) |
| Classmates' fathers' share: Postgraduate degree | | | 0.028 |
| | | | (0.808) |
| Classmates' fathers' share: Not applicable | | | -1.706 |
| | | | (1.120) |
| Class size | | | 0.005 |
| | | | (0.022) |
| $(\text{Class size})^2$ | | | 0.000 |
| | | | (0.000) |
| Teacher experience $< 3$ years | | | -0.074 |
| | | | (0.053) |
| Classmates' average age | | | 0.595 |
| | | | (0.530) |
| School-level fixed effects | No | Yes | Yes |
| Controls | No | No | Yes |
| Number of observations | 2409 | 2409 | 2409 |
| Number of classroom clusters | 63 | 63 | 63 |
| $R^2$ | 0.018 | 0.159 | 0.241 |

Robust standard errors, adjusted for clustering at the school level, are presented in parentheses.
*** Significant at the 1% level; ** Significant at the 5% level; * Significant at the 10% level

## A.4 Intersectional effects

Extended version of Table 7.3.

Table A.4: Regression of share of girls in the classroom on mathematics test scores by group with intersection

| | (1) | (2) | (3) |
|---|---|---|---|
| Girl share | 0.140 | -0.123 | -0.280 |
| | (0.335) | (0.509) | (0.518) |
| Girl × (Girl share) | 0.043 | 0.023 | 0.558 |
| | (0.078) | (0.084) | (0.439) |
| New immigrant × (Girl share) | -1.971*** | -1.278*** | -1.772 |
| | (0.401) | (0.323) | (1.350) |
| Girl × New immigrant × (Girl share) | 0.424 | -0.014 | 3.601** |
| | (0.471) | (0.460) | (1.634) |
| Student's age | | | 0.318*** |
| | | | (0.080) |
| Girl | | | -0.283 |
| | | | (0.228) |
| New immigrant | | | 0.291 |
| | | | (0.714) |
| Girl × New immigrant | | | -1.934** |
| | | | (0.864) |
| Mother: Did not go to school | | | 0.146 |
| | | | (0.629) |
| Mother: Some primary school | | | 0.291 |
| | | | (0.388) |
| Mother: Upper secondary school | | | 0.449 |
| | | | (0.413) |
| Mother: Post secondary, less than Bachelor's | | | 0.510 |
| | | | (0.409) |
| Mother: Bachelor's or equivalent | | | 0.683 |
| | | | (0.419) |
| Mother: Postgraduate degree | | | 0.811* |
| | | | (0.428) |
| Mother: Not applicable | | | 0.000 |
| | | | (.) |
| Father: Did not go to school | | | -0.264 |
| | | | (0.397) |
| Father: Some primary school | | | 0.036 |
| | | | (0.223) |
| Father: Upper secondary school | | | 0.158 |
| | | | (0.221) |
| Father: Post secondary, less than Bachelor's | | | 0.189 |
| | | | (0.211) |
| Father: Bachelor's or equivalent | | | 0.400* |
| | | | (0.234) |
| Father: Postgraduate degree | | | 0.445* |
| | | | (0.231) |
| Father: Not applicable | | | 0.000 |
| | | | (.) |
| Classmates' mothers' share: Did not go to school | | | -0.500 |
| | | | (1.108) |

| | | | |
|---|---|---|---|
| Classmates' mothers' share: Some primary school | | | 0.248 |
| | | | (1.914) |
| Classmates' mothers' share: Upper secondary school | | | -0.651 |
| | | | (1.349) |
| Classmates' mothers' share: Post secondary, less than Bachelor's | | | -0.255 |
| | | | (1.870) |
| Classmates' mothers' share: Bachelor's or equivalent | | | -0.243 |
| | | | (1.541) |
| Classmates' mothers' share: Postgraduate degree | | | 0.437 |
| | | | (1.928) |
| Classmates' mothers' share: Not applicable | | | -0.107 |
| | | | (1.680) |
| Classmates' fathers' share: Did not go to school | | | -1.318 |
| | | | (1.389) |
| Classmates' fathers' share: Some primary school | | | 2.645*** |
| | | | (0.907) |
| Classmates' fathers' share: Upper secondary school | | | 0.459 |
| | | | (0.577) |
| Classmates' fathers' share: Post secondary, less than Bachelor's | | | -0.020 |
| | | | (0.709) |
| Classmates' fathers' share: Bachelor's or equivalent | | | 0.857 |
| | | | (0.721) |
| Classmates' fathers' share: Postgraduate degree | | | 0.046 |
| | | | (0.804) |
| Classmates' fathers' share: Not applicable | | | -1.702 |
| | | | (1.115) |
| Class size | | | 0.005 |
| | | | (0.022) |
| (Class size)$^2$ | | | 0.000 |
| | | | (0.000) |
| Teacher experience $<$ 3 years | | | -0.076 |
| | | | (0.053) |
| Classmates' average age | | | 0.595 |
| | | | (0.528) |
| School-level fixed effects | No | Yes | Yes |
| Controls | No | No | Yes |
| Number of observations | 2409 | 2409 | 2409 |
| Number of classroom clusters | 63 | 63 | 63 |
| $R^2$ | 0.018 | 0.159 | 0.242 |

Robust standard errors, adjusted for clustering at the school level, are presented in parentheses.
*** Significant at the 1% level; ** Significant at the 5% level; * Significant at the 10% level