Carl von Ossietzky Universität Oldenburg

Oldenburg Discussion Papers in Economics

Better Off On Their Own?

How Peer Effects Determine International Patterns of the Mathematics Gender Achievement Gap

Bernhard C. Dannemann

V – 433-20

November 2020

Department of Economics
University of Oldenburg, D-26111 Oldenburg

Better Off On Their Own? How Peer Effects Determine International Patterns of the Mathematics Gender Achievement Gap *

Bernhard C. Dannemann†

This Version: November 2020

Abstract

This paper applies recent spatial regression techniques in peer effects estimation to a sample of 33 countries in the IAE's TIMSS 2015 study in order to quantify the gender achievement gap in eighth grade mathematics. Based on an education production function setting and controlling for the mediating influences of student-, parent-, teacher-, and school-level factors, a significant, but small gender achievement gap amounting to 10% of a country-level standard deviation is confirmed in the standard linear model. I model endogenous and exogenous peer effects based on the workhorse linear-in-means model as well as on homophily-based dyadic relations between students, both with controls for group unobservables. The results show that the effect size increases to 12-14%, depending on the underlying socio-matrix. However, the partitioned impacts suggest even larger effects when considering spillovers within the classroom. These could amount to as much as 38% of a standard deviation, depending on the underlying dyadic peer weights, with the linear-in-means model possibly overstating the magnitude of classroom externalities.

Keywords: Gender Gap, Education Production Function, Human Capital, Peer Effects

JEL Classification Numbers: I20, I24, J16, J24

*I would like to thank Jürgen Bitzer and Erkan Gören for helpful comments and suggestions. All remaining errors are my own.

†Carl von Ossietzky University Oldenburg, School of Computing Science, Business Administration, Economics, and Law (Faculty II), Institute of Economics, Building A5, 26111 Oldenburg, Germany, Tel.: +49-441-798-4006,
website: https://uol.de/en/economics/applied-macroeconomics/team/bernhard-dannemann/,
e-mail: bernhard.dannemann@uni-oldenburg.de

1 Introduction

Despite longstanding efforts at promoting gender equality in various domains of politics, the economy, and the educational system – as formulated, for example, in the Millennium Development Goals of 2000 and monitored by the Global Gender Gap Report (GGGR) since 2006 – the aims of gender parity, equality of opportunities, and equal achievement across scholastic domains have still not yet been achieved completely. This is true of both developing as well as developed countries to differing degrees and creates considerable challenges for educational policies (World Economic Forum, 2020). One striking example of this is that even in the year 2020, and throughout developed countries, women are still underrepresented in certain science, technology, engineering and mathematics (STEM) fields (for instance, computer science, mathematics, or statistics), while they have achieved parity in others, such as biology and chemistry (Cheryan et al., 2017; Wang and Degol, 2017). Especially in the face of technology-rich environments and skills shortage, not making full use of one half of the population in these fields seems inefficient and likely to perpetuate the gender income gap. Despite the fact that most modern countries have established women's freedom to choose and pursue a career as a fundamental right, many of these countries are far from actually achieving gender parity of graduates in STEM programmes and other technology-rich skill domains (World Economic Forum, 2020).

One possible explanation for this pattern is that certain convictions, beliefs, and role models persist that drive occupational choices and career goals of female students. It has been reported that gender differences in mathematics skills at high achievement levels (Ellison and Swanson, 2010) are a driving force in different subject choices even within STEM fields (Friedman-Sokuler and Justman, 2016). Interestingly, it is specifically in the subject of mathematics that female students are prone to gender-specific stereotyping by family and peers (Ganley et al., 2013; Stoet and Geary, 2012) as well as by teachers (Robinson-Cimpian et al., 2014), with adverse effects on their performance. The gender gap in mathematics has been the subject of wide-ranging research since the 1990s, most of it based on results from international student achievement tests (Lindberg et al., 2010; Else-Quest et al., 2010; Guiso et al., 2008), national evaluation programs (Reilly et al., 2015), and teacher-grade-reports (Voyer and Voyer, 2014). However, the results compiled in these meta-studies differ widely across countries and regions, and appear inconclusive, with negligibly small overall effect sizes.

Accounting for these unobserved and latent factors or cultural traits in the evaluation of gender achievement gaps poses a challenge for empiricists, as the gender gap should be determined net of all possibly confounding factors. Assessing mathematics achievement and its determinants across a sample of economically and culturally diverse countries is thus crucial for understanding patterns of the gender achievement gap. The Trends In Mathematics and Science Study (TIMSS) provides a comprehensive data set containing measures of student achievement, as well as possible determinants on the individual, parent, school, and teacher levels. Further, the rigorous school and classroom sampling in TIMSS renders it suitable for the within-classroom modeling of peer effects.[1] The list of participating countries, regions, and provinces is extensive and, more important, heterogeneous with regard to cultural and socio-demographic variables, but also educational inputs, performance, framework, and efficiency.

Based on the most recent 2015 wave of TIMSS, I model student achievement as the outcome of an augmented education production function, with standard inputs controlling for student- and parent-level variables, as well as classroom- and subject-specific teacher fixed effects. Based on social network matrices embodying the sociological concepts of *status* and *value homophily*, following Dannemann (2020), I select relevant peers from the classroom population to include exogenous and endogenous peer effects. This methodology is applied to a heterogeneous sample of 33 countries or regions[2], yielding a total of $165,980$ individual student test score observations from the 33 cross-sections. The augmented education production function is then estimated for each of these country or region cross-sections, using a spatial auto-regressive model and maximum likelihood estimation (Lee, 2007). These approaches have gained popularity in the estimation of peer effects, as they allow for separate identification of different types of peer effects (as discussed in Manski, 1993) and the implementation of social network matrices to describe dyadic relations between students based on different approaches to weighting (Lin, 2010; Bramoullé et al., 2009; Boucher et al., 2014; Lin, 2015).The results from the individual country cross-sections are then analyzed to assess the gender achievement gap in mathematics, separated into the direct effects on the individual and spillovers to classmates.

As a baseline scenario, education production functions without peer effects are estimated based on a standard linear model (SLM) for all 33 cross-sections. Holding constant the effect of standard inputs of the EPF, such as student characteristics, parental background, and teacher fixed effects (see, for instance, Todd and Wolpin, 2003, 2007; Hanushek and Woessmann, 2011, for detailed information), a significant gender achievement gap in favor of boys is found for the vast majority of countries, which exceeds the magnitude of raw gender differences in mathematics achievement. However, this first approach does not consider the influence of peer relations, which could lead to biased estimates of the gender gap.

Accordingly, in a second step, the prior results are replicated based on an EPF augmented by peer effects arising at the classroom level. Peer effects are modeled using two different approaches. First, the linear-in-means (LiM) model, which assigns uniform weights to all students of a class as the potential peer group represents a well-established form of modeling in the literature on peer effects estimation. Based on this specification of the social network matrices, a widening gender gap in the presence of peer effects is observed, which suggests that the baseline model without peer effects provides a negatively biased estimate of the share of inequality attributable to student gender. Further, the spatial framework allows for the separate discussion of direct and indirect impacts. The results suggest that the spillovers attributable to student gender exceed the direct effects in magnitude. However, the precision of the LiM model estimates varies widely across

[1]The estimation of peer effects is considered problematic for survey-based student assessments, due to the random sampling of classroom members representing the peer group (Micklewright et al., 2012).

[2]These include 3 African, 18 Asian, 5 European, 4 North American, 2 South American, and 2 Oceanian countries or regions. See Table 1 for detailed information.

countries and regions, casting doubt on their reliability.

A further approach to assign weights to dyadic relationships between students is the use of a social network matrix based on a combination of *status homophily* and *value homophily* (see Dannemann, 2020). Here, the dyadic tendency to interact and cooperate is determined based on visual characteristics (that is, gender, age, and migratory background), but also on values and beliefs of the students (for example, sense of school belonging, confidence in mathematics). The two concepts correspond to short- and long-run dynamics in the grouping behavior of students, which have been observed and confirmed in various sociological studies (McPherson et al., 2001; Currarini et al., 2009; Shrum et al., 1988; Louch, 2000; Moreland and Beach, 1992; Tsai et al., 2016). The results based on the *homophily* social network matrix suggest a gender gap that is slightly smaller than in the LiM case, but still considerably larger than in the standard linear model. This is mainly attributable to the indirect effects, which are estimated with more precision, but which also decreased considerably in terms of magnitude.

The remainder of the paper is organized as follows. Section 2 reviews the relevant literature on the presence, magnitude, and possible determinants of the gender achievement gap in mathematics as well as the role of peers in this context. Section 3 outlines the empirical investigation together with the construction of sociology-based social network matrices and a detailed discussion of the main control variables and their respective transmission channels. Section 4 presents the main results and puts them into the context of existing meta-analyses. Section 5 summarizes the main findings and concludes.

2 Theoretical Background: Gender Inequality in Education

At first glance, presumably the greatest progress for gender equality has been achieved in the domain of educational attainment. In the latest GGGR, the World Economic Forum (2020) states that in education more than 95% of the gender parity gap has already been closed, meaning that men and women should differ little regarding access to education and educational attainment. However, this perspective lacks detail, by focusing on enrollment and overall attainment,[3] rather than taking different fields, subjects, and opportunities into consideration. Accordingly, achieving gender parity in basic education is only a prerequisite for reducing gender differences in educational performance. The number of male and female students in the TIMSS data set suggests the presence of gender parity in all participating countries. However, this is a statistical artifact, as the study's sampling design produces largely similar shares of male and female students on the national level for each individual cross-section. However, these show the interesting pattern of a very heterogeneous distribution across classrooms. Unequal gender composition of classes could affect the ways students interact with their classmates and thus be associated with heterogeneous educational outcomes by gender. Figure 1 clearly highlights the different patterns of inequality in gender parity and the raw gender achievement gap (that is, the difference between the average achievement of male and female students) across countries

[3]As measured, for example, by average years of schooling, as compiled by Barro and Lee (2013). However, these data show hardly any difference in educational attainment between the male and female population of developed economies.

and regions. The Middle East countries (with the exception of Turkey) consistently show high segregation by gender across classes (more than 70% inequality), but are predominantly characterized by a negative achievement gap, ranging between 10 and 20% of a country-level standard deviation, implying that girls outperform boys. The majority of remaining countries fall substantially below 40% classroom segregation by gender. Despite their low level of segregation, the American and Oceanian countries in particular show a positive gender gap. For Europe, the figure suggests the absence of a significant gender gap, while the evidence for African and Asian countries appears mixed.

Figure 1 about here

If anything, this bi-variate relation between lacking gender parity and the achievement gap appears to be negative. This would suggest the preposterous implication that female students benefit from segregated classes, or that they are *"better off on their own"*. This idea is disproven by evidence from Ireland, where the historically high level of gender segregation is accompanied by no notable reductions in the gender gap in mathematics (Doris et al., 2013). While some countries show very low levels of classroom-level polarization, others have very segregated classes for cultural, religious, or institutional reasons (making Ireland a unique example in Europe, see Doris et al., 2013).[4] Nevertheless, the figure suggests that classroom composition with regard to gender affects the gender achievement gap to at least some extent.

Only a few studies exist that directly address the role of peers and especially peer gender in the context of the gender achievement gap. Reportedly, gender-based differences in susceptibility to peer influences exists, that is, the intra-gender spillover is more pronounced than the inter-gender effect when considering the grade point average (GPA) of students (Hsieh and Lin, 2017). Similar results exists for other outcomes, such as smoking, alcohol or drug use, obesity, or participation in extracurricular activities (Hsieh and Lin, 2017; Clark and Lohéac, 2007; Trogdon et al., 2008; Knifsend et al., 2018; Kooreman, 2007), suggesting that gender-based differences in group formation behavior could indeed affect the presence and magnitude of the gender gap in scholastic achievement. However, only a handful of studies empirically address the endogenous formation of peer groups based on attributes such as gender (the theoretical foundations are laid in Qu and Lee (2015); Hsieh and Lee (2016), and an empirical application is provided in Hsieh and Lin (2017)).

Early evidence points to the absence of overall gender differences in mathematics (Hyde et al., 1990; Felson and Trudeau, 1991). A commonly proposed explanation for these minor gender differences is gendered parenting, a style of upbringing that may include discouraging girls from engaging in mathematics. It has been argued, for example, that boys outperform girls in school mathematics as a result of increased efforts and motivation, because they expect this will lead to better occupational opportunities (a phenomenon known as the gender stratification hypothesis, see Baker and Perkins Jones, 1993). The gender gap would, according to this theory, be driven by incentives and female labor market participation.

[4] Saudi Arabia is the only country in the TIMSS data set that is characterized by comprehensive gender segregation, which makes the estimation of peer effects and their effect on the gender achievement gap inapplicable. Thus, Saudi Arabia is excluded from the empirical analysis.

Even in more recent research, the evidence regarding the gender gap has remained inconclusive. For example, a comprehensive meta-analysis of findings since 1990 by Lindberg et al. (2010) finds a gender gap in mathematics with a Cohen's d effect size measure of $d = 0.05$, which the authors deem not to constitute a systematic overall gender difference. However, they conclude that their results point toward significant, yet small differences in intra-group variability. This implies that the group of boys in school contains higher shares of high- and low-performing students. However, when considering studies from different regions of the world, effect sizes vary considerably, for example, with higher differences in Africa and lower differences in European countries. A common approach for explaining these cross-country patterns is that these differences disappear in more developed countries that place a greater emphasis on gender-equal policies and culture (Guiso et al., 2008). This pattern is consistent with the results presented by Else-Quest et al. (2010), who conduct an international meta-analytic study based on the coinciding 2003 waves of TIMSS and PISA. While they find no evidence for the presence of a gender gap in mathematics for the 46 country TIMSS sample ($d = -0.01$), the PISA sample (40 countries) shows a significant difference in performance ($d = 0.11$). Further, the authors find sizable differences in self-confidence and affect towards mathematics, especially in the PISA part of the sample, which supports the hypothesis of differences in upbringing and socialization being the main cause of gender gaps in educational performance.

Several studies based on United States National Assessment of Educational Progress (NAEP) and confirmed by a meta study also report a persistent small but positive gender gap persists between 1990 and 2011, with a significantly higher share of high-achieving male students (Reilly et al., 2015). Notwithstanding, a meta-analysis on the gender gap, measured by teacher grades, concludes that female students outperform their male peers on average across various subjects, including language, mathematics, and science courses (Voyer and Voyer, 2014). This finding suggests that differences in attitudes and behavior might indeed play a crucial role in determining gender-based differences in achievement. For example, female students have been observed to receive higher grades than their male counterparts as a consequence of higher self-discipline. While this provides an advantage in the classroom context throughout the school year (Duckworth and Seligman, 2006), it is considered less useful in the direct situation of standardized achievement tests (Walton and Spencer, 2009).

Furthermore, the differences in test scores and grades suggest that teacher grades could be biased with regard to student gender. Accordingly, Robinson-Cimpian et al. (2014) find that in the subject of mathematics and after accounting for differences in achievement and teacher ratings of effort, teacher perceptions in kindergarten initially favor of girls, but shift to favor boys in later grades. As they conclude "[...] teachers rate girls on par with similarly achieving boys only if they perceive those girls as working harder and behaving better than those boys." (ibid., p.1275). These behavioral differences are reinforced by the stereotype threat – a latent perception that mathematics is not a "girls' subject", leading to psychological stress and reduced test performance (Stoet and Geary, 2012; Ganley et al., 2013; Walton and Spencer, 2009).

As the aforementioned meta-studies illustrate, a widely varying pattern of differences exists for the gender gap in mathematics across countries, resulting in inconclusive results for overall measures, which should thus be interpreted with caution. As Else-Quest et al. (2010) acknowledge, the results depend not only on the source of the underlying student achievement test data (that is, TIMSS, PIRLS, PISA, or others), but also on the data quality, which has improved with more recent waves through considerable improvements in the representativity, coverage and detail of the studies. Likewise, according to Fryer and Levitt (2010), a possible explanation for the presence of a negative gender gap is that the selection of students of both genders might not be representative across international student achievement tests, as they show in a comparison to the United States NAEP data.

In this paper, I provide an overview of the magnitude of the gender achievement gap across 33 countries and regions participating in TIMSS 2015, which provides sampling of complete classrooms. Relying on the most recent wave of the student achievement test ensures data quality and representativity (Else-Quest et al., 2010; Fryer and Levitt, 2010). Along with the standard inputs of education production functions, I control for the confounding effects of classroom peers by including *endogenous* and *exogenous* peer effects in a spatial-autoregressive model with group unobservables (Lee, 2007). Apart from the workhorse LiM model, I present an application of dyadic peer weights based on the sociological concept of *homophily* introduced in Dannemann (2020) to attain a more relevant peer weighting (Hsieh and Lin, 2017).

3 Empirical Approach

3.1 Econometric Model

In order to evaluate individual student achievement, I estimate an EPF as specified in Equation (1) using a set of standard inputs to control for individual and parental characteristics as well as teacher- and class-level heterogeneity (Hanushek and Woessmann, 2011, 2017). To consider the effect of peers and classroom structure, peer effects are explicitly modeled in the regression as social network matrices in a spatial model.

$$O_{ics} = \alpha + \rho \sum_{j \in c} w_{ij} O_{jcs} + \sum_{k=1}^{K} \beta^{(k)} X_{ics}^{(k)} + \sum_{k=1}^{K} \theta^{(k)} \left(\sum_{j \in c} w_{ij} X_{jcs}^{(k)} \right) + \boldsymbol{P}_{ics} \boldsymbol{\delta} + \boldsymbol{T}_{\boldsymbol{cs}} \boldsymbol{\gamma} + u_{ics} \quad (1)$$
$$\text{where } u_{ics} = \lambda \sum_{j \in c} w_{ij}^{(Adj.)} u_{jcs} + \varepsilon_{ics}$$

The spatial model described in Equation (1) is a general nested spatial model (GNSM), which contains spatial dependency in the dependent variable, independent variables, and the error term. Generally, spatial interactions in the dependent variable correspond to *endogenous* peer effects, that is, the achievement of the individual is affected by the achievement of peers. In contrast, the effect of changes in peers' independent variables is the *exogenous* peer effect; that is, individual achievement is partly determined by relevant peer attributes and characteristics (Manski, 1993). This approach allows for the separate identification of *endogenous* and *exogenous* peer effects while controlling for group unobservables (that is, *correlated effects*), as is shown in Lee (2007).

In Equation (1), α denotes the intercept. The main dependent variable O_{ics} is the student-level mathematics test score of student i in school s and class c. To ensure comparability across the different countries and regions, the variable is standardized for each cross-section by subtracting the mean and dividing by the standard deviation. The coefficient ρ is the *endogenous* peer effect parameter; the term $\sum_{j \in c} w_{ij} O_{jcs}$ refers to the weighted sum of peer math scores. The dyadic weights w_{ij} represent the strength of the relation between student i and j, as measured by the corresponding social network matrix (see the section below for details on the construction of weights). The individual characteristics of the respective student are included in the variables $X^{(k)}_{ics}$ with $k = \{1, \ldots, K\}$ and the associated coefficient $\beta^{(k)}$. These variables are the students gender, age, migratory background, language skills, and perceived bullying (see the discussion below). Further, the influence of the K attributes of classmates or peers on individual performance, that is, the *exogenous peer effect*, is measured by the weighted sum of these attributes, $\sum_{j \in c} w_{ij} X^{(k)}_{jcs}$, using the respective definition of peer weights and the associated coefficients $\theta^{(k)}$. The vector $\boldsymbol{P}_{ics}$ with its coefficient vector $\boldsymbol{\delta}$ represents variables of the parents that are also specific to the individual i. To control for the *correlated* effects arising from unobserved heterogeneity, such as, teacher characteristics, school resources, and infrastructural features of the location, a set of teacher-dummies ($\boldsymbol{T}_{cs}$) specific to school s and class c together with their coefficient vector $\boldsymbol{\gamma}$ is included in the regression. It is worth noting that in TIMSS 2015, there is at least one teacher per class and subject. Accordingly, using teacher fixed effects is more fine grained than relying on class fixed effects.

The residuals u_{ics} are modeled to contain spatial dependency on the classroom level based on the LiM social network matrix (with uniform weights across peers) through the term $\lambda \sum_{j \in c} w^{(Adj.)}_{ij} u_{jcs}$. The motivation for including spatial dependency in the error term among classmates is to account for any remaining unobserved common shocks or factors occurring at the classroom level, which would otherwise induce auto-correlation and invalid inferences (Lin, 2010; LeSage and Pace, 2009). The remaining error term ε_{ics} is then an i.i.d. stochastic term.

Equation (1) is estimated for each of the 33 countries/regions in the sample individually. Accordingly, all coefficients are specific to the 33 distinct cross-sections. The spatial regression model relies on maximum likelihood estimation, as the estimation of models with spatial dependence in the dependent variable using ordinary least squares entails inconsistent estimation of parameters and standard errors (see LeSage and Pace, 2009). All weights matrices are row-normalized to enhance estimability and to facilitate comparability of the estimated effects. Row-normalizing ensures that the weights for each observation sum up to one and thus corresponds to a weighted averaging (Darmofal, 2015). Subsequently, the individual country/region-specific results for the gender indicator variable are employed in a random effects meta-regression model to obtain an overall measure of the effect size. The coefficients are weighted by the associated sample size and standard errors.

While in the standard linear model, the estimated regression coefficients can be interpreted directly as marginal effects, this does not apply to spatial models. In these, as represented in Equation (1) above, the partial derivative of the achievement of individual i with respect to variables of individual j is non-zero. Accordingly, a matrix of partial derivatives for each variable can be constructed, where the main diagonal corresponds to the derivative of student i with respect to the variable of i. This corresponds to the *direct impact*, which includes feedback mechanisms from peers. The off-diagonal elements are then the partial derivative of student i's achievement with respect to the variable of student $j \neq i$, which are the indirect effects specific to the respective student. To obtain a pooled effect, the average effect of the non-zero off-diagonal elements is calculated, which is then the average *indirect impact*. The *total impact* is the sum of *direct* and *indirect impacts* (see, for example LeSage and Pace, 2011, for further discussion). Calculating partitioned effects to assess effect sizes provides the opportunity to identify possible transmission channels of peer effects and gives additional insight on how the gender achievement gap differs across countries. For parsimony, only the partitioned impacts for the gender variable are reported. All other results are available upon request.[5]

3.2 Data and Variables

Main dependent variable. The test score data is obtained from the Trends in Mathematics and Science Study (TIMSS) 2015, a project of the International Association for the Evaluation of Educational Achievement (IAE). TIMSS is designed as a paper-based student assessment, where the language of test is usually the official language of the respective country or sub-region. Regarding the use of calculators in the eighth grade mathematics assessment, the IAE allows each country to decide individually whether this is permitted. The idea is to not give any (dis)advantage to students (not) accustomed to using calculators in mathematics lessons (Grønmo et al., 2015). Student test scores both on the individual as well as the aggregate level have been accepted as a measure of general cognitive ability based on a significant correlation with results from IQ tests (Rindermann, 2007).[6]

Student gender. An indicator variable for male student gender is included in the regression. Differences between male and female students have been observed in both cognitive and non-cognitive abilities. Non-cognitive factors, such as diligence, self-efficacy, or motivation, which play a vital role in determining scholastic outcomes, are found to be affected by gender-specific differences in upbringing, cultural environment, and social conventions (Guiso et al., 2008; Fryer and Levitt, 2010; Ganley et al., 2013; Stoet and Geary, 2012), as well as by how students are perceived, treated, and graded by their teachers (Robinson-Cimpian et al., 2014). Especially stereotypes regarding the roles, occupations, and skills of women are still present in various domains, with implications on educational and occupational choices (Aesaert and van Braak, 2015; Imdorf et al., 2015). Despite major efforts since the adoption of the Millennium Development Goals to attain equality of opportunity for both genders (Sultana, 2008), there are still countries with no universal access to education, especially for girls, as is reflected, for instance, by differences in enrollment rates (Megahed and Lack, 2011;

[5] While this paper focuses on the interpretation and discussion of the partitioned effects of the gender variable, both the estimated regression coefficients and the partitioned impacts for the remaining variables are provided in the online supplemental material to this paper.

[6] Overall, the TIMSS study, as opposed to, for example, OECD's PISA, is aimed slightly more at application and knowledge, rather than reasoning. Consequently, international student tests in general and TIMSS in particular have been criticized for promoting rote or *parrot-fashion* learning (Felson and Trudeau, 1991; Hanushek and Woessmann, 2008, 2011).

World Economic Forum, 2015). Regarding cognitive abilities, which reflect actual classroom performance, gender differences may be perpetuated when gender-specific curricular needs are neglected, for example, when there is a lack of physical activity as a balancing factor for boys to increase their ability to concentrate and stay focused (Cöster et al., 2018). This is consistent with findings of Lavy and Schlosser (2011), who, for the United States, find the incidence of disruption and violence to decrease as the share of girls in a classroom is increased.

Further student-level variables. First, student age controls for the fact that TIMSS 2015 is a grade-level study specific to eighth-grade students.[7] Differences in student age within countries can be broadly distinguished into being caused by either *timing of enrollment*, or by *grade repetition*. Regarding the timing, delayed enrollment may be beneficial due to, for example, higher readiness for school, better health, or gains in maturity (Bedard and Dhuey, 2006; McEwan and Shapiro, 2007; Brown and Park, 2002; Glewee and Jacoby, 1995). Grade repetition (or the opposite, grade acceleration or skipping) is suggestive of the student's abilities (Hughes et al., 2017) and circumstances, for example, nutrition, state of health, major issues, or other important factors (Warne, 2017; Kretschmann et al., 2016; Warne and Liu, 2017).

Second, it has been observed that for immigrants, both the socio-economic status, as well as the educational and economic opportunities differ across countries and cultures, which accounts for different speeds of assimilation and affects human capital accumulation (Djajić, 2003). To control for this issue, I include variables referring to the student's first- or second-generation immigrant background, as well as to the frequency of use of the language of test at home. A higher share of migrants in classrooms has been found to have adverse impacts on the educational performance of natives (Brunello and Rocco, 2013) and has thus been the subject of heated discussion in policy and research. Further, education received in the home country is found to affect the performance in the destination country (Hanushek and Woessmann, 2012) and the efforts to acquire sufficient language skills. Successfully acquiring language skills is determined by students' motivation, which again correlates positively with socio-economic status (Kormos and Kiddle, 2013).

Third and last, to consider the sense of well-being for the individual students and the general social atmosphere within a classroom, a control for the student's perception of bullying activities in the everyday school context is included. The phenomenon of school bullying, which occurs across a wide range of countries, affects the behavior and performance of students throughout virtually all grades and for diverse reasons (Bekiari et al., 2017; Ponzo, 2013). It exerts a negative influence on student achievement, causing psychological distress, increased absenteeism and, reduced school attendance (Harel-Fisch et al., 2011).

Parent-level variables. The highest educational attainment of the parents is included as a compound measure for various aspects of child-rearing and parental involvement (Johnsen et al., 2018). Parental education serves as a proxy for the innate abilities of the child, as a strong heritability of intelligence from parents

[7] It differs in this respect from the OECD's PISA study, which is targeted at a 15-year-old student population and is therefore heterogeneous in terms of grade. There is significant variation in student age both between and within countries (see Table 1).

to their offspring is observed (Anger and Heineck, 2010). In addition, parental education also determines parental efforts, for example, by increasing the focus on child health and nutrition (Currie, 2009), increasing parents' involvement in their children's academic activities (Fan and Chen, 2001; Davis-Keane, 2005), and increasing the value parents place on education *per se* (Angelucci et al., 2010; Rindermann, 2007). Further, it is often associated with a better socio-economic status and thus a better provision of material resources (Schuetz et al., 2008). Similar to the student-level variable, a variable controlling for the immigration background of both parents is included to account for the effect of second generation immigrants. As stated by Djajić (2003), different speeds of assimilation can give rise to conflicts among immigrant children and their parents, with unclear implications on schooling efforts and academic performance. As a further measure of resources employed in the acquisition of cognitive skills, proxies for household economic and educational resources are included (Parcel and Dufur, 2001). Such are the number of books (Fuchs and Woessmann, 2007) or digital devices (Hanushek et al., 2015) at home, or the provision of basic study supplies such as a desk or an own room. These variables proxy for a basic level of wealth or economic resources, and for cultural and technological affinity of the parents.

Correlated effects. Research has shown that school systems differ across countries in their capability to compensate for differences in students' socio-economic status. This is of particular relevance, as it has been observed that the variation in socio-economic background within schools and classrooms is higher in developing countries (Gustafsson et al., 2018). Schools' endowment with educational and instructional resources and infrastructure, which differ across both countries and schools, could also affect student outcomes (see Hanushek and Woessmann, 2017, for a comprehensive discussion). For example, it has been observed for a sample of British schools that school poverty negatively affects students' educational achievement (Nieuwenhuis, 2018). At the teacher level, it is well-documented that teacher characteristics partly determine the efficiency of the learning environment. While it has been shown that especially beginner teachers perform significantly worse than their more experienced counterparts (Rivkin et al., 2005), reasearch has had difficulties identifying the influence of other observable teacher characteristics (Hanushek and Rivkin, 2010).

3.3 Social Network Matrices

Peer effects are usually approximated by calculating averages of variables across a predefined group of potential peers. In this study, common class membership is chosen as the peer definition to exploit this specific feature of the TIMSS data set and is represented by the so-called adjacency matrix $\boldsymbol{W}^{(Adj.)}$, as formalized in Equation (2).

$$\boldsymbol{W}^{(Adj.)} := (w_{ij})_{n\times n}, \text{ where } w_{ij} = \begin{cases} 1 & if\ c_i = c_j \\ 0 & if\ c_i \neq c_j \\ 0 & if\ i = j \end{cases} \tag{2}$$

Here, students i and j are assigned a dyadic weight of 1 if they are members of the same class c and 0 otherwise. Of course, the individual under investigation should not be considered in the calculation of averages and is thus assigned zero weight, that is, the adjacency matrix corresponds to calculating the mean of the peer group, except for individual i, using uniform weights for all peers. Even though this approach has been commonly employed in the literature on peer effects in the past, it fails to account for the relevance of different peers in the peer group (see the discussion in Angrist, 2014; Lin, 2015). This is especially relevant in the case of a sample that is heterogeneous in terms of culture, values, ethnicity, and class composition. Furthermore, the countries differ substantially in terms of educational systems and educators' approaches to classroom organization. Considering student decisions in group formation could provide further insights into peer relations and help in the selection of relevant peers (which appears important given the concerns raised by Angrist, 2014), even though the sensitivity of spatial models to matrix choices is reported to be rather low (LeSage and Pace, 2011).

Following Dannemann (2020), I construct similarity-based social network matrices to account for the two forms of *homophily* that could arise in different stages of group formation within the classroom (McPherson et al., 2001; Di Stefano et al., 2015). First, *status homophily* corresponds to students forming peer groups based on visual characteristics in the short run, that is, in the early phase after being assigned to classes. Accordingly, dyadic similarity between students is calculated based on the attributes gender, age, and migratory background (Shrum et al., 1988; Currarini et al., 2009; Tsai et al., 2016). Second, *value homophily* considers values and beliefs of students in the decision to cooperate and thus corresponds to a process of group formation that is oriented more towards the long run, that is, the groups that have formed some time after students were assigned to the class (Freeman et al., 2017). Here, dyadic similarity is based on students' self-reports of whether they like or value the subject mathematics, and how they assess their feeling of school belonging. To attain a holistic measure, *combined homophily* incorporates the rationale of both of the aforementioned concepts. It is used to proxy for the idea that groups that formed through *status homophily* in the initial phase have subsequently changed through *value homophily* after students have got to know each other.

$$\boldsymbol{W}^{(Hom.)} := (S_{ij})_{n\times n}, \text{ where } S_{ij} = \frac{1}{M}\sum_{m=1}^{M} s_{ij}^{(m)}(c) \tag{3}$$

In Equation (3), $\boldsymbol{W}^{(Hom.)}$ is the *combined homophily* classroom social network matrix. For all n observations of each sample, S_{ij} are the dyadic similarities of students i and j in classroom c, which correspond to the average across the individual similarities $s_{ij}^{(m)}(c)$ of the M variables employed in the calculation. This compound dyadic similarity measure S_{ij} ranges between zero and one and represents the relative importance of the peers across the classroom. This implies that peer weights reflect a cardinal scale. The individual similarities are then calculated as shown in Equation (4), following the concept of the Gower (1971) similarity measure.

$$s_{ij}^{(m)}(c) = 1 - \frac{|x_i^{(m)}(c) - x_j^{(m)}(c)|}{\max(x^{(m)}(c)) - \min(x^{(m)}(c))} \quad (4)$$

Here, $x_i^{(m)}(c)$ and $x_j^{(m)}(c)$ are the values on metric variables $x^{(m)}$ for students i and j in classroom c. The expression $\big(\max(x^{(m)}(c)) - \min(x^{(m)}(c))\big)$ denotes the range within the classroom c of variable $x^{(m)}(c)$. By definition, the similarity between the two students with the minimum and maximum value on variable $x^(m)$ within the classroom is zero, whereas two students with identical values have a similarity of one. The similarities for dichotomous variables are defined as shown in Equation (5). If students i and j are similar on the dichotomous variable, for example *migratory background*, they are assigned a similarity of one. If they differ, their similarity is accordingly set to zero.

$$s_{ij}^{(m)}(c) := \begin{cases} 1 \text{ if } x_i^{(m)}(c) = x_j^{(m)}(c) \\ 0 \text{ otherwise} \end{cases} \quad (5)$$

The *combined homophily* matrix $\boldsymbol{W}^{(Hom.)}$ differs from the adjacency matrix $\boldsymbol{W}^{(Adj.)}$ by allowing the relative importance of peers to vary (that is, not all peers in the classroom exert equally strong effects) and by relaxing the implicit assumption of full network density, which implies for the adjacency matrix that all possible connections within the network exist at full strength. Instead, more weight is shifted to the more relevant peers (in terms of likelihood to interact and cooperate, as has been suggested by sociological studies, such as Freeman et al., 2017; Tsai et al., 2016).

3.4 Descriptive Statistics

Out of the 45 countries and regions/sub-regions included in the TIMSS data set, 33 countries/regions are part of the empirical sample (see Table 1 for additional details). The study includes sub-samples for two African, six North or South American, 11 Middle East, seven Asian, five European, as well as two Oceanian countries and regions, and thus provides student-level data on achievement and various individual- and parent-level variables for a total of $163,752$ eighth-grade students. The complete list of included countries/regions with their sample sizes is shown in Table 1.

Figure 2 about here

The sample of countries and regions is rather heterogeneous in terms of student achievement in mathematics, as can be seen from Figure 2, which shows the countries' deviation from 500 test score points, which is often interpreted as the OECD average performance and is fairly close to the sample mean. The dashed lines indicate the range of one sample standard deviation (approximately 60 points), implying the presence of 5 relatively high-performance countries (the East Asian countries Hong Kong, Japan, South Korea, Taiwan, and Singapore) as well as 7 low-achieving countries (the four Middle East countries Egypt, Oman, Kuwait, and Jordan, the African countries South Africa and Botswana, as well as Buenos Aires (Argentina)).

4 Main Empirical Results

For each of the 33 countries/regions and the two distinct types of social network matrices, the education production function is estimated to assess the magnitude of the gender achievement gap in mathematics. As discussed above, the partitioned impacts are analyzed in the following to attain a sense of the pattern and magnitude of the gender achievement gap. Below the individual country/region-specific results, an overall measure is presented that is based on a random effects restricted maximum likelihood (REML) meta regression.

Direct Impacts. As a first step, it is worthwhile to look at the direct impacts, as they are most suitable for comparison with other studies. Table 2 shows the results for the direct impacts of male student gender on mathematics test performance. It is divided into three panels, where the left one (a) shows the results for the standard linear model (or "OLS" estimates) without peer effects. The middle one (b) shows the partitioned impacts for the linear-in-means model, and the right one (c) shows those of the *combined homophily* model. Each panel includes a plot and a list of the direct impact effect sizes with their associated 90% confidence interval. Last, each final column shows the relative weight of the regression in the computation of the overall measure (indicated as a diamond at the bottom of the graph), which is based on the standard error and the number of observations of the individual regressions.

Table 2 about here

In Panel (a), the coefficients for the baseline estimates from the standard linear model without peer effects are shown. The results are straightforward and unambiguous: As opposed to the initial picture emerging from the stylized facts presented in the prior section (especially Figure 1), only Egypt, Abu Dhabi, and South Korea show a negative gender gap. Instead, the majority of countries/regions (27 out of 33, with 3 insignificant), including the previously negative Middle East countries, shows a significant and positive gender gap (with the exception of Oman, for which no significant difference is found). Most of these estimates range between 5% and 15% of a standard deviation of the dependent variable. It is striking that especially Iran, Jordan, and Kuwait, the countries with the highest levels of classroom gender segregation, are estimated very imprecisely, as becomes apparent from the wide confidence intervals. Interestingly, for the North and South American regions, all estimated gender gaps in mathematics exceed the overall average direct effect for the SLM, which is estimated at 0.10 standard deviations of the student achievement variable.

For the direct effects in the linear in means model in Panel (b), that is, upon the inclusion of peer effects based on the unweighted classroom average of peers, it is mostly the effect sizes that are affected, but not the signs. Especially the American countries and regions show clearly increased magnitudes of the estimated gender gap. Particularly for the Middle East countries, the pattern of the gender gap is noteworthy. Although with the exception of Dubai, the countries are characterized by a negative raw gender gap (see Figure 1), the estimate from the SLM model as well as the direct effects in the augmented models point to the existence of a positive gender gap, as most effects are shifted positively, as is the case for Abu Dhabi. While in the SLM case, a negative gender achievement gap of -0.13 standard deviations is found, it is increased to positive 0.10 standard deviations of the dependent variable upon the inclusion of peer effects in the EPF. A noteworthy exception is the decrease in the coefficient for Iran, which is accompanied by a reduction in precision, causing it to lose statistical significance at conventional levels. More generally, as compared to the SLM, Egypt is now the only country with a significant and negative gender gap. Overall, 26 of the 33 cross-sections find a positive direct effect. For 6 countries (Botswana, Abu Dhabi, Iran, Jordan, Oman, and South Korea), no significant difference in mathematics performance for boys and girls is found, although it is striking that especially the effects for Iran and Jordan are estimated very imprecisely, as is indicated by the very large confidence interval. Accordingly, the overall measure suggests a gender gap of 0.14 standard deviations in the dependent variable.

Switching to the *combined homophily* social network matrix in Panel (c), which corresponds to considering *status* and *value homophily* in peer relations, brings only small changes to the overall pattern. The majority of regressions continue to show positive and significant direct effects of male gender (25 out of 33). While significance vanishes for Thailand and Norway, the direct effect for Iran is estimated more precisely and gains significance at the 10% level. It is also striking that for Canada and its province Ontario (but not for Quebec), the magnitude of the effect is decreased substantially, but remains significant at the 1% level.

To summarize, relying on *homophily*-based peer effects in the calculation of the gender achievement gap reduces the magnitude of the direct effects in many cases and results in an overall measure of 0.12 standard deviations. This is relatively sizable, compared to effect sizes found in meta-analyses (Hyde et al., 1990; Lindberg et al., 2010; Reilly et al., 2015). Especially the European and American countries show the sobering picture of a clearly positive gender gap, which is particularly pronounced in Ireland, Italy, Argentina (Buenos Aires), and Chile. This finding confirms that significant differences still persist in mathematics achievement, which contradicts the statement that differences in math scores would disappear in the presence of a gender-equal culture (as found by Guiso et al., 2008, based on the full gender gap index) or that European and North American countries have largely achieved gender equality in education, as suggested by the education sub-index (World Economic Forum, 2020). However, this sub-index, for which the focus is on enrollment and literacy, might be too superficial to capture differences across subjects or gender-based stereotyping (Ganley et al., 2013; Stoet and Geary, 2012). The negative implications of these role models and stereotypes are differences in subjects of university study and occupational choices, with low shares of women in STEM fields even in highly developed countries (Cheryan et al., 2017; Wang and Degol, 2017; Friedman-Sokuler and Justman, 2016). Thus, the findings on international achievements in promoting gender equality in education presented as part of the GGGR bear the risk of giving a false sense of security.

Indirect Impacts. In contrast, the indirect impacts (or classroom spillovers) presented in Table 3 are markedly more diverse and in both panels show no clear overall pattern. Note that in the absence of peer effects, no indirect effects can be calculated, thus the panel with the SLM results is dropped.

Panel (a) with the LiM results contains 12 regressions with positive and significant spillover effects. However, 5 of these (South Africa, Canada, Ontario, Oman, and Thailand) show effect sizes of more than one standard deviation, which appears unreasonably large.[8] 6 countries (United Arab Emirates, Abu Dhabi, Dubai, Egypt, Qatar, and New Zealand) show a significantly negative spillover effect. Further, 15 regressions show no significant spillover effect for the gender variable. Nevertheless, the overall measure amounts to 0.38 standard deviations, which is very sizable but is presumably driven by the large positive outliers.[9]

Table 3 about here

In Panel (b), the results are even more ambiguous. In only 10 cases does the spillover effect remain positive and significant (Quebec, Malaysia, and Oman lose significance, and England gains significance, although only at the 10% level). Further, the magnitude of the prior very large outliers is reduced considerably (for example, the coefficient for Canada decreases from 3.97 in LiM to 1.70, which is, however, still considered large). Interestingly, five of the ten Middle East countries show a negative and significant spillover effect, while not one is positive, implying that increasing the share of boys reduces the performance of all classmates. New Zealand is no longer associated with a significant spillover effect, while Singapore turns significantly negative at the 10% level, yielding a total of seven cases with negative spillovers. In the case of *combined homophily*, the overall measure is calculated as 0.13 standard deviations, but it is not statistically significantly different from zero at the 10% level, with the confidence interval barely containing the zero.[10]

Both specifications with peer effects reveal a negative classroom externality for the United Arab Emirates (including Abu Dhabi and Dubai), Qatar, Iran, and Egypt, suggesting that classes benefit from an increased number of girls, presumably through an improved learning atmosphere. One possible explanation for this observation could be changes in the cultural and religious environment of these primarily Muslim countries. The literature has found religiosity rather than political institutions to influence educational outcomes (Cooray and Potrafke, 2011; Østby et al., 2016), mainly as a consequence of bans on certain intellectual and scientific pursuits, as well as the presence of religiosity in all spheres of everyday life (Stoet and Geary, 2017). Despite Muslim women being identified as among the least educated population groups in prior cohorts, more recent generations have succeeded in catching up and significantly reducing the education gap, both with respect to Muslim men as well as to women of other religions, especially in the Gulf States. This effect has been particularly strong in countries where it has been accompanied by high economic development (McClendon

[8] Sawilowsky (2009) discusses rules of thumb for assessing effect sizes. Effect sizes larger than one standard deviation are considered *large*, while effect sizes greater than two standard deviations are considered *huge*.

[9] The overall effect decreases to 0.11 $[-0.05,\ 0.27]$ standard deviations if the countries with an indirect effect size larger than two $S.D.$ (Canada, Ontario, and South Africa) are removed from the sample.

[10] Likewise, the overall effect decreases to 0.09 $[-0.03,\ 0.21]$ standard deviations, if according to the prior rule, the country with an indirect effect size larger than two $S.D.$, that is, Ontario (Canada), is removed from the sample.

et al., 2018).

Total Impacts. Table 4 shows the results for the total impacts of male student gender on mathematics test performance. In addition to the prior tables, it replicates the point estimates for the direct and indirect impacts for the sake of comparison.

In Panel (a), positive and significant estimates for direct, indirect, and total impacts are found in 11 cross-sections (South Africa, Canada, Ontario, Quebec, United States, Malaysia, Thailand, Taiwan, Ireland, Norway, and Australia). This implies the existence of a positive gender achievement gap, which is further perpetuated by a positive classroom externality. Only England, Italy, and Hong Kong show a positive and significant direct and total impact, but an insignificant indirect impact. Here, increases in the individual student's performance affect classmates only through the *endogenous* peer effect, that is, through improved scholastic performance, yet the gender variable has no distinct *exogenous* effect on the peer outcomes.

Table 4 about here

In Oman, the positive and significant total impact is driven by the positive and significant indirect impact. The direct effect itself is not significant at conventional levels. It is worth mentioning that the indirect impact amounts to 1.063 standard deviations in the dependent variable, which is again unusually large and points towards a rather biased estimation (see Note 8). In the 8 regressions, Buenos Aires, Chile, Bahrain, Kuwait, Turkey, Japan, Singapore, and Norway (eighth grade), a positive and significant direct impact is governed by the insignificance of the indirect impact, entailing an insignificant total impact.

Six countries are characterized by a significantly negative total impact: First, Abu Dhabi shows a significant and negative indirect, yet an insignificant direct impact. Here, the performance of boys does not differ significantly from that of girls, however, they create a negative classroom externality. The United Arab Emirates, Dubai, Qatar, and New Zealand show the interesting pattern of a positive direct impact that is outweighed by a significant and negative indirect effect. While for these countries, a positive gender achievement gap for the individual is confirmed, the resulting negative externality is stronger, suggesting that larger shares of male students are harmful to the classroom atmosphere, as suggested by Lavy and Schlosser (2011).

The overall total effect across the 33 regressions is estimated to be 0.52 standard deviations of the dependent variable and is considerably larger than comparable magnitudes found in the literature (Lindberg et al., 2010, $d = 0.05$) (Else-Quest et al., 2010, $d = -0.01$ for TIMSS and $d = 0.11$ for PISA 2003) (Guiso et al., 2008, approx. 10.5 points / 2% of mean score[11]).

At first glance, switching to the *combined homophily* social network matrix in Panel (b) entails only minor changes in the signs and significance of the results. For Buenos Aires, Chile, Kuwait, Japan, Singapore,

[11] Guiso et al. (2008) report no descriptive statistics on the estimation sample. As their study is based on the 2003 wave of PISA, it can be assumed that the gap of 10.5 corresponds to approximately $10 - 15$ % of a country-level standard deviation.

and Norway (eighth grade) the prior insignificant total impact turns positive and significant. This is caused mainly by higher precision in the estimation of the indirect impact, which, however, remains insignificant in four of the five cases, with the exception of Singapore, where the indirect effect becomes significant at the 10% level. Similarly, the precision of Iran's indirect impact of student gender is substantially increased, causing it to gain statistical significance at the 1% level. In spite of the direct impact turning significant and positive, although only at the 10% level, the total impact is significant and negative as well, although at the 5% level. Further, two countries with a positive total impact in the LiM model, Oman and Quebec, no longer show a significant total impact upon switching to the *combined homophily* model, presumably due to notable reductions in the point estimate for the indirect effect, causing it to lose significance at conventional levels. The direct effect remains virtually unchanged (positive for Quebec and insignificant for Oman). Nevertheless, the magnitude of indirect impacts is heavily reduced as compared to the LiM model for the majority of countries (see, for example, Canada, South Africa, or Thailand). Interestingly, some countries (such as Australia or Norway (eighth grade)) even show modest increases in the magnitude of the spillover. However, the dispersion of the total effects is reduced considerably as compared to the LiM model. These findings are further highlighted by the overall total impact, which is reduced to half its magnitude. Accordingly, the total effect of male student gender on student achievement in mathematics across the 33-country sample amounts to 0.26 standard deviations on average.

5 Conclusion

This paper has presented evidence for the presence of a gender achievement gap in eighth grade mathematics among a total of $163,752$ students in 33 participating countries and regions of the IAE's 2015 wave of the Trends in Mathematics and Science Study (TIMSS). In an augmented education production function, the student-level estimates for each country/region consider the confounding effects of individual- (that is, gender, age, migratory background, perceived bullying) or parent-level variables (for example, level of education, migratory background, or household resources). Further, I control for the influence of *endogenous* and *exogenous* peer effects as well as spatially clustered errors in a spatial auto-regressive model, which helps in further reducing omitted variable bias on the estimated coefficient of the student gender variable. This model controls implicitly for school- and teacher-level factors through the inclusion of teacher fixed effects to capture the influence of group unobservables (Lee, 2007). The results show that the direct impacts of the student gender variable range between 0.12 and 0.14 standard deviations of the mathematics test score variable when including peer effects and are thus considered moderate but significant. Further, the indirect effects calculated from the spatial model suggest that classroom externalities are sizable and considerably larger than the direct effects (0.38 S.D. in the LiM, 0.13 S.D., but insignificant, in the *combined homophily* model). Considering the total impacts of student gender suggests a gender achievement gap of 0.52 standard deviations in the LiM model, which decreases to 0.26 standard deviations once the effect of *status* and *value homophily* in peer relations (see Dannemann, 2020, for a detailed discussion of the homophily network matrices) is considered.

The main insight from the evaluation of the gender achievement gap across 33 countries and regions is that a significant difference between the mathematics performance of male and female students persists, even in highly developed countries. Furthermore, accounting for the externalities arising from classroom peers suggests that the gap could even be significantly larger than prior studies suggest (Lindberg et al., 2010). The focus on partitioned impacts comes with twofold implications: First, considering direct impacts is broadly in line with the magnitudes of the gender achievement gap typically found in prior meta-analyses. However, the magnitude of the gap increases when peer effects are considered, suggesting that the omission of *endogenous* and *exogenous* peer effects entails a negative bias on the coefficient of the gender variable. Second, the results from indirect and total impacts suggest that classroom externalities and spillovers could be important drivers of gender inequality in education, which have not been explicitly addressed in prior research. Comparing results from the LiM and *combined homophily* models highlights the importance of selecting relevant peers in the estimation. Using the LiM, which implicitly assumes maximum network density, as the unweighted peer measure entails imprecise estimation of spillovers (as is shown in Bramoullé et al., 2009) and an overstated gender gap overall. This finding points to the need for further research on social dynamics within schools and classrooms, which might be at the core of gender differences in certain subjects. While notable formal models for endogenously determined social matrices exist (Qu and Lee, 2015; Hsieh and Lee, 2016), their application is still limited (Hsieh and Lin, 2017). In fact, this study has highlighted the presence of different patterns in gender gap spillovers, yet remains short on detailed explanations for the emergence of these differences, which are thus left for future research. For example, observing actual classroom relations in an experimental setting could provide educators with valuable information on how students of different genders interact and thus provide a starting point for teacher intervention to optimize the learning atmosphere and inhibit negative externalities.

Overall, several policy implications can be drawn from this paper. For the European, American, and Oceanian countries, the largely positive and significant direct effects underscore the need for educational policy intervention if the goal of gender equality is to be achieved. It is noteworthy that the magnitude of the peer effects is affected strongly by the underlying social network matrix, suggesting that these externalities are driven by the group formation behavior of students and could thus be moderated through teacher intervention. However, these results are rather imprecise and indicate the need for further research. Accordingly, they should be interpreted with caution. Especially for the Middle East region, a large discrepancy between gender parity, the SLM, and the augmented models with peer effects is found. Regardless of the massive gains in participation and attainment of girls in education (as is shown in World Economic Forum, 2015, 2020), gender segregation is still existent across a wide area and substantially exceeds that in other regions of the world. The negative spillovers suggest that reducing segregation by promoting gender-mixed classes could be a key to lowering the gender gap in mathematics and fostering gender equality in education. In light of this, considering raw gender gaps as stylized facts, as presented in Figure 1, is misleading and conceals the fact that female students in these countries are indeed not *"better off on their own"*. Strategies for increasing female mathematics achievement could aim, for example, at promoting classroom gender parity to mitigate negative externalities arising from a high share of male students in the classroom.

References

Aesaert, K. and van Braak, J. (2015). Gender and Socioeconomic Related Differences in Performance Based ICT Competences. *Computers & Education*, 84:8–25.

Angelucci, M., De Giorgi, G., Rangel, M. A., and Rasul, I. (2010). Family Networks and School Enrollment: Evidence From a Randomized Social Experiment. *Journal of Public Economics*, 94:197–221.

Anger, S. and Heineck, G. (2010). Do Smart Parents Raise Smart Children? The Intergenerational Transmission of Cognitive Abilities. *Journal of Population Economics*, 23:1255–1282.

Angrist, J. D. (2014). The Perils of Peer Effects. *Labour Economics*, 30:98–108.

Baker, D. P. and Perkins Jones, D. (1993). Creating Gender Equality: Cross-national Gender Stratification and Mathematical Performance. *Sociology of Education*, 66(2):91–103.

Barro, R. J. and Lee, J. W. (2013). A New Data Set of Educational Attainment in the World, 1950–2010. *Journal of Development Economics*, 104:184–198.

Bedard, K. and Dhuey, E. (2006). The Persistence of Early Childhood Maturity: International Evidence of Long-Run Age Effects. *Quarterly Journal of Economics*, 121(4):1437–1472.

Bekiari, A., Pachi, V., and Hasanagas, N. (2017). Investigating Bullying Determinants and Typologies with Social Network Analysis. *Journal of Computer and Communication*, 5:11–27.

Boucher, V., Bramoullé, Y., Djebbari, H., and Fortin, B. (2014). Do Peers Affect Student Achievement? Evidence from Canada Using Group Size Variation. *Journal of Applied Econometrics*, 29:91–109.

Bramoullé, Y., Djebbari, H., and Fortin, B. (2009). Identification of peer effects through social networks. *Journal of Econometrics*, 150:41–55.

Brown, P. H. and Park, A. (2002). Education and Poverty in Rural China. *Economics of Education Review*, 21:523–541.

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*, 32:234–246.

Cheryan, S., Ziegler, S. A., Montoya, A. K., and Jiang, L. (2017). Why Are Some STEM Fields More Gender Balanced Than Others? *Psychological Bulletin*, 143(1):1–35.

Clark, A. E. and Lohéac, Y. (2007). "It wasn't me, it was them!" Social Influence in Risky Behavior by Adolescents. *Journal of Health Economics*, 26(4):763–784.

Cooray, A. and Potrafke, N. (2011). Gender Inequality in Education: Political Institutions or Culture and Religion? *European Journal of Political Economy*, 27(2):268–280.

Cöster, M. E., Fritz, J., Karlsson, C., Rosengren, B. E., and Karlsson, M. K. (2018). Extended Physical Education in Children Aged 6–15 Years Was Associated With Improved Academic Achievement in Boys. *Acta Pædiatrica*, 107:1083–1087.

Currarini, S., Jackson, M. O., and Pin, P. (2009). An Economic Model of Friendship: Homophily, Minorities, and Segregation. *Econometrica*, 77(4):1003–1045.

Currie, J. (2009). Healthy, Wealthy, and Wise: Socioeconomic Status, Poor Health in Childhood, and Human Capital Development. *Journal of Economic Literature*, 47(1):87–122.

Dannemann, B. C. (2020). Peer Effects in Secondary Education: Evidence from Trends in Mathematics and Science Study 2015 based on Homophily. *Oldenburg Discussion Papers in Economics*, V–428-20.

Darmofal, D. (2015). *Spatial Analysis for the Social Sciences*. Cambridge University Press.

Davis-Keane, P. E. (2005). The Influence of Parent Education and Family Income on Child Achievement: The Indirect Role of Parental Expectations and the Home Environment. *Journal of Family Psychology*, 19(2):294–304.

Di Stefano, A., Scatà, M., La Corte, A., Liò, P., Catania, E., Guardo, E., and Pagano, S. (2015). Quantifying the Role of Homophily in Human Cooperation Using Multiplex Evolutionary Game Theory. *PLOS ONE*, 10(10):1–21.

Djajić, S. (2003). Assimilation of Immigrants: Implications for Human Capital Accumulation of the Second Generation. *Journal of Population Economics*, 16(4):831–845.

Doris, A., O'Neill, D., and Sweetman, O. (2013). Gender, Single-sex Schooling and Maths Achievement. *Economics of Education Review*, 35:104–119.

Duckworth, A. L. and Seligman, M. E. P. (2006). Self-Discipline Gives Girls the Edge: Gender in Self-Discipline, Grades, and Achievement Test Scores. *Journal of Educational Psychology*, 98(1):198–208.

Ellison, G. and Swanson, A. (2010). The Gender Gap in Secondary School Mathematics at High Achievement Levels: Evidence from the American Mathematics Competitions. *Journal of Economic Perspectives*, 24(2):109–28.

Else-Quest, N. M., Hyde, J. S., and Linn, M. L. (2010). Cross-National Patterns of Gender Differences in Mathematics: A Meta-Analysis. *Psychological Bulletin*, 136(1):103–127.

Fan, X. and Chen, M. (2001). Parental Involvement and Students' Academic Achievement: A Meta-Analysis. *Educational Psychology Review*, 13(1):1–22.

Felson, R. B. and Trudeau, L. (1991). Gender Differences in Mathematics Performance. *Social Psychology Quarterly*, 54(2):113–126.

Freeman, S., Theobald, R., Crowe, A. J., and Wenderoth, M. P. (2017). Likes Attract: Students self-sort in a Classroom by Gender, Demography, and Academic Characteristics. *Active Learning in Higher Education*, 18(2):115–126.

Friedman-Sokuler, N. and Justman, M. (2016). Gender Streaming and Prior Achievement in High School Science and Mathematics. *Economics of Education Review*, 53:230–253.

Fryer, R. G. and Levitt, S. D. (2010). An Empirical Analysis of the Gender Gap in Mathematics. *American Economic Journal: Applied Economics*, 2(2):210–240.

Fuchs, T. and Woessmann, L. (2007). What Accounts for International Differences in Student Performance? A Re-Examination using PISA Data. *Empirical Economics*, 32:433–464.

Ganley, C. M., Mingle, L. A., Ryan, A. M., Ryan, K., Vasilyeva, M., and Perry, M. (2013). An Examination of Stereotype Threat Effects on Girls' Mathematics Performance. *Developmental psychology*, 49(10):1886–1897.

Glewee, P. and Jacoby, H. G. (1995). An Economic Analysis Of Delayed Primary School Enrollment In A Low Income Country. *Review of Economics and Statistics*, 77(1):156–169.

Gower, J. C. (1971). A General Coefficient of Similarity and Some of Its Properties. *Biometrics*, 27(4):857–871.

Grønmo, L. S., Lindquist, M., Arora, A., and Mullis, I. V. (2015). TIMSS 2015 Mathematics Framework. *TIMSS*, pages 11–27.

Guiso, L., Monte, F., Sapienza, P., and Zingales, L. (2008). Culture, Gender, and Math. *Science*, 320:1164–1165.

Gustafsson, J.-E., Nilsen, T., and Hansen, K. Y. (2018). School Characteristics Moderating the Relation Between Student Socio-Economic Status and Mathematics Achievement in Grade 8. Evidence from 50 Countries in TIMSS 2011. *Studies in Educational Evaluation*, 57:16–30.

Hanushek, E. A. and Rivkin, S. G. (2010). Generalizations about Using Value-Added Measures of Teacher Quality. *American Economic Review*, 100(2):267–271.

Hanushek, E. A., Schwerdt, G., Wiederhold, S., and Woessmann, L. (2015). Returns to Skills Around the World: Evidence from PIAAC. *European Economic Review*, 73:103–130.

Hanushek, E. A. and Woessmann, L. (2008). The Role of Cognitive Skills in Economic Development. *Journal of Economic Literature*, 46(3):607–68.

Hanushek, E. A. and Woessmann, L. (2011). The Economics of International Differences in Educational Achievement. In Hanushek, E. A., Machin, S., and Woessmann, L., editors, *Handbook of the Economics of Education. Band 3*, volume 26, pages 89–200. Elsevier.

Hanushek, E. A. and Woessmann, L. (2012). Do Better Schools Lead to More Growth? Cognitive Skills, Economic Outcomes, and Causation. *Journal of Economic Growth*, 17:267–321.

Hanushek, E. A. and Woessmann, L. (2017). School Resources and Student Achievement: A Review of Cross-Country Economic Research. In *Cognitive Abilities and Educational Outcomes*, pages 149–171. Springer.

Harel-Fisch, Y., Walsh, S. D., Fogel-Grinvald, H., Amitai, G., Pickett, W., Molcho, M., Due, P., Gaspar de Matos, M., Craig, W., and Members of the HSBC Violence and Injury Prevention Focus Group (2011). Negative School Perceptions and Involvement in School Bullying: A Universal Relationship Across 40 Countries. *Journal of Adolescence*, 34(4):639–652.

Hsieh, C.-S. and Lee, L.-F. (2016). A Social Interactions Model with Endogenous Friendship Formation and Selectivity. *Journal of Applied Econometrics*, 31:301–319.

Hsieh, C.-S. and Lin, X. (2017). Gender and Racial Peer Effects with Endogenous Network Formation. *Regional Science and Urban Economics*, 67:135–147.

Hughes, J. N., Cao, Q., West, S. G., Smith, P. A., and Cerda, C. (2017). Effect of Retention in Elementary Grades on Dropping Out of School Early. *Journal of School Psychology*, 65:11–27.

Hyde, J. S., Fennema, E., and Lamon, S. J. (1990). Gender Differences in Mathematics Performance: A Meta-Analysis. *Psychological Bulletin*, 107(2):139–155.

Imdorf, C., Hegna, K., Eberhard, V., and Doray, P. (2015). Educational Systems and Gender Segregation in Education: A Three-Country Comparison of Germany, Norway and Canada. In *Gender Segregation in Vocational Education*. Emerald Group Publishing Limited.

Johnsen, A., Bjørknes, R., Iversen, A. C., and Sandbæk, M. (2018). School Competence among Adolescents in Low-Income Families: Does Parenting Style Matter? *Journal of Child and Family Structure*, 27:2285–2294.

Knifsend, C. A., Camacho-Thompson, D. E., Juvonen, J., and Graham, S. (2018). Friends in Activities, School-related Affect, and Academic Outcomes in Diverse Middle Schools. *Journal of Youth and Adolescence*, 47:1208–1220.

Kooreman, P. (2007). Time, Money, Peers, and Parents; Some Data and Theories on Teenage Behavior. *Journal of Population Economics*, 20(1):9–33.

Kormos, J. and Kiddle, T. (2013). The Role of Socio-Economic Factors in Motivation to Learn English as a Foreign Language: The Case of Chile. *System*, 41(2):399–412.

Kretschmann, J., Vock, M., Lüdtke, O., and Gronostaj, A. (2016). Skipping to the Bigger Pond: Examining Gender Differences in Students' Psychosocial Development after Early Acceleration. *Contemporary Educational Psychology*, 46:195–207.

Lavy, V. and Schlosser, A. (2011). Mechanisms and Impacts of Gender Peer Effects at School. *American Economic Journal: Applied Economics*, 3(2):1–33.

Lee, L.-F. (2007). Identification and Estimation of Econometric Models with Group Interactions, Contextual Factors and Fixed Effects. *Journal of Econometrics*, 140:333–374.

LeSage, J. P. and Pace, R. (2009). *Introduction to Spatial Econometrics*. Statistics: A Series of Textbooks and Monographs. CRC Press.

LeSage, J. P. and Pace, R. K. (2011). Pitfalls in Higher Order Model Extensions of Basic Spatial Regression Methodology. *The Review of Regional Studies*, 41(1):13–26.

Lin, X. (2010). Identifying peer effects in student academic achievement by spatial autoregressive models with group unobservables. *Journal of Labor Economics*, 28(4):825–860.

Lin, X. (2015). Utilizing Spatial Autoregressive Models to Identify Peer Effects Among Adolescents. *Empirical Economics*, 49:929–960.

Lindberg, S. M., Hyde, J. S., and Petersen, J. L. (2010). New Trends in Gender and Mathematics Performance: A Meta-Analysis. *Psychological Bulletin*, 136(6):1123–135.

Louch, H. (2000). Personal Network Integration: Transitivity and Homophily in Strong-Tie Relations. *Social Networks*, 22:45–64.

Manski, C. F. (1993). Identification of Endogenous Social Effects: The Reflection Problem. *Review of Economic Studies*, 60(3):531–542.

McClendon, D., Hackett, C., Potančoková, M., Stonawski, M., and Skirbekk, V. (2018). Women's Education in the Muslim World. *Population and Development Review*, 44(2):311–342.

McEwan, P. J. and Shapiro, J. S. (2007). The Benefits of Delayed Primary School Enrollment – Discontinuity Estimates Using Exact Birth Dates. *The Journal of Human Resources*, 43(1):1–28.

McPherson, J. M., Smith-Lovin, L., and Cook, J. M. (2001). Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology*, 27:415–444.

Megahed, N. and Lack, S. (2011). Colonial Legacy, Women's Rights and Gender-Educational Inequality in the Arab World with particular Reference to Egypt and Tunisia. *International Review of Education*, 57(3-4):397–418.

Micklewright, J., Schnepf, S. V., and Silva, P. N. (2012). Peer Effects and Measurement Error: The Impact of Sampling Variation in School Survey Data (Evidence from PISA. *Economics of Education Review*, 31:1136–142.

Moreland, R. L. and Beach, S. R. (1992). Exposure Effects in the Classroom: The Development of Affinity Among Students. *Journal of Experimental Social Psychology*, 28:255–276.

Nieuwenhuis, J. (2018). The Interaction Between School Poverty and Agreeableness in Predicting Educational Achievement. *Personality and Individual Differences*, 127:85–88.

Østby, G., Urdal, H., and Rudolfsen, I. (2016). What Is Driving Gender Equality in Secondary Education? Evidence from 57 Developing Countries, 1970–2010. *Education Research International*, 2016.

Parcel, T. L. and Dufur, M. J. (2001). Capital at Home and at School: Effects on Student Achievement. *Social Forces*, 79(3):881–911.

Ponzo, M. (2013). Does Bullying Reduce Educational Achievement? An Evaluation Using Matching Estimators. *Journal of Policy Modeling*, 35:1057–1078.

Qu, X. and Lee, L.-F. (2015). Estimating a Spatial Autoregressive Model with an Endogenous Spatial Weight Matrix. *Journal of Econometrics*, 184:209–232.

Reilly, D., Neumann, D. L., and Andrews, G. (2015). Sex Differences in Mathematics and Science Achievement: A Meta-Analysis of National Assessment of Educational Progress Assessments. *Journal of Educational Psychology*, 107(3):645–662.

Reynal-Querol, M. (2002). Ethnicity, Political Systems, and Civil Wars. *Journal of Conflict Resolution*, 46(1):29–54.

Rindermann, H. (2007). The g-Factor of International Cognitive Ability Comparisons: The Homogeneity of Results in PISA, TIMSS, PIRLS and IQ-Tests Across Nations. *European Journal of Personality*, 21:667–706.

Rivkin, S. G., Hanushek, E. A., and Kain, J. F. (2005). Teachers, Schools, and Academic Achievement. *Econometrica*, 73(2):417–458.

Robinson-Cimpian, J. P., Lubienski, S. T., Ganley, C. M., and Copur-Gencturk, Y. (2014). Teachers' Perceptions of Students' Mathematics Proficiency May Exacerbate Early Gender Gaps in Achievement. *Developmental Psychology*, 50(4):1262–1281.

Sawilowsky, S. S. (2009). New Effect Size Rules of Thumb. *Journal of Modern Applied Statistical Methods*, 8(2):597–599.

Schuetz, G., Ursprung, H. W., and Woessmann, L. (2008). Education Policy and Equality of Opportunity. *Kyklos*, 61:279–308.

Shrum, W., Cheek, N. H., and Hunter, S. M. (1988). Friendship in School: Gender and Racial Homophily. *Sociology of Education*, 61(4):227–239.

Stoet, G. and Geary, D. C. (2012). Can Stereotype Threat Explain the Gender Gap in Mathematics Performance and Achievement? *Review of General Psychology*, 16(1):92–102.

Stoet, G. and Geary, D. C. (2017). Students in Countries with Higher Levels of Religiosity Perform Lower in Science and Mathematics. *Intelligence*, 62:71–78.

Sultana, R. G. (2008). The Girls' Education Initiative in Egypt. *United Nations Children's Fund*.

Todd, P. E. and Wolpin, K. I. (2003). On the Specification and Estimation of the Production Function for Cognitive Achievement. *The Economic Journal*, 113(485):F3–F33.

Todd, P. E. and Wolpin, K. I. (2007). The Production of Cognitive Achievement in Children: Home, School, and Racial Test Score Gaps. *Journal of Human Capital*, 1(1):91–136.

Trogdon, J. G., Nonnemaker, J., and Pais, J. (2008). Peer Effects in Adolescent Overweight. *Journal of Health Economics*, 27(5):1388–1399.

Tsai, J., Valente, T. W., Miller, K. A., de la Haye, K., Pickering, T. A., and Cockburn, M. G. (2016). Friendship Networks and Sun Safety Behavior Among Children. *Network Science*, 4(3):314–335.

Voyer, D. and Voyer, S. D. (2014). Gender Differences in Scholastic Achievement: A Meta-Analysis. *Psychological Bulletin*, 140(4):1174–1204.

Walton, G. M. and Spencer, S. J. (2009). Latent Ability: Grades and Test Scores Systematically Underestimate the Intellectual Ability of Negatively Stereotyped Students. *Psychological Science*, 20(9):1132–1139.

Wang, M.-T. and Degol, J. L. (2017). Gender Gap in Science, Technology, Engineering, and Mathematics (STEM): Current Knowledge, Implications for Practice, Policy, and Future Directions. *Educational Psychology Review*, 29:119–140.

Warne, R. T. (2017). Possible Economic Benefits of Full-Grade Acceleration. *Journal of School Psychology*, 65:54–68.

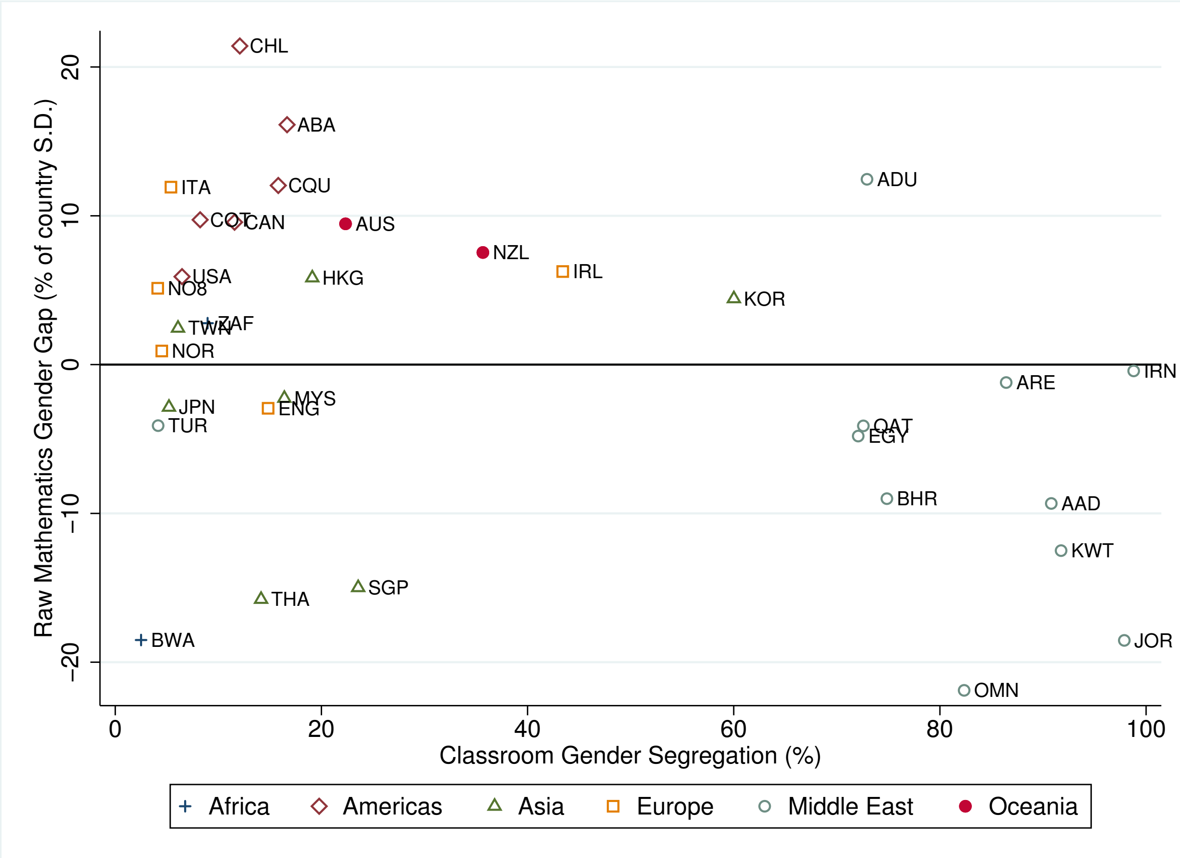
Warne, R. T. and Liu, J. K. (2017). Income Differences Among Grade Skippers and Non-Grade Skippers Across Genders in the Terman Sample, 1936–1976. *Learning and Instruction*, 47:1–12.

World Economic Forum (2015). The Global Gender Gap Report 2015. Retrieved from http://www3.weforum.org/docs/GGGR2015/cover.pdf, last accessed: July 22, 2020.

World Economic Forum (2020). The Global Gender Gap Report 2020. Retrieved from http://www3.weforum.org/docs/WEF_GGGR_2020.pdf, last accessed: July 22, 2020.

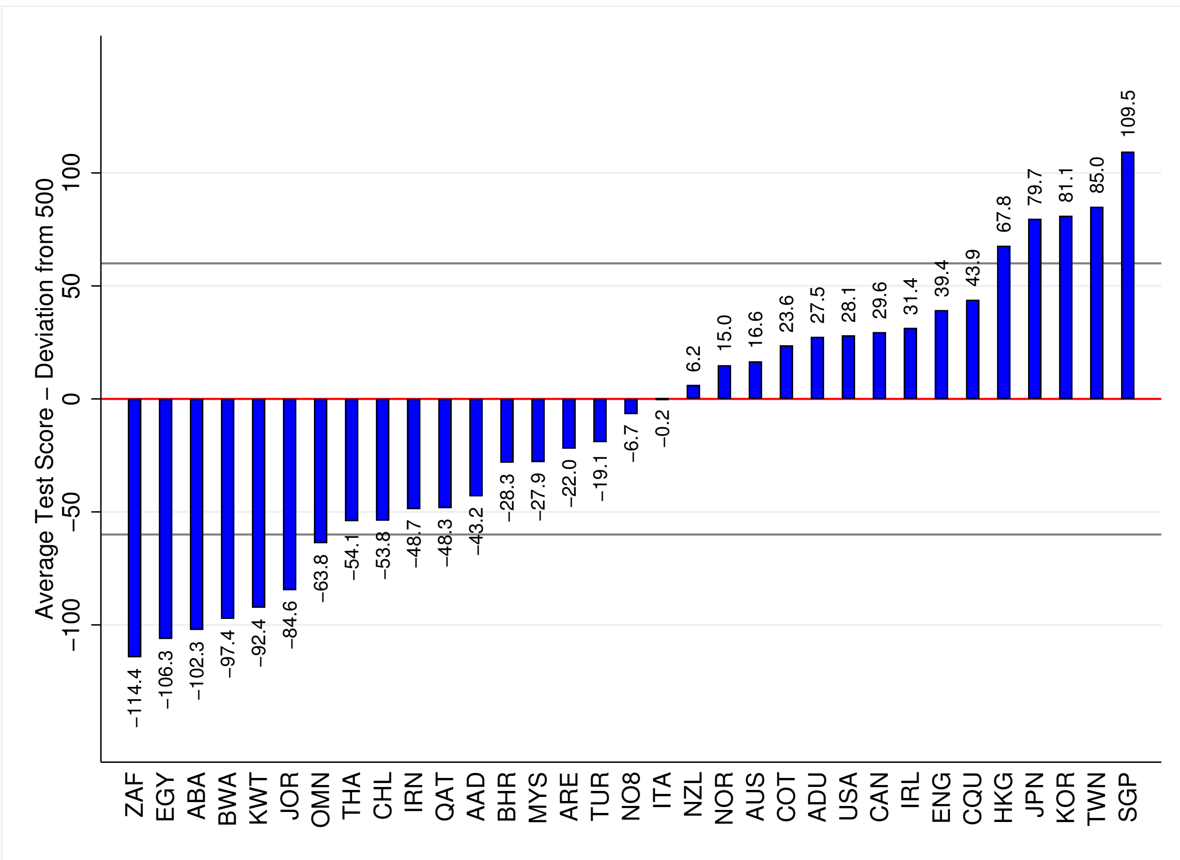
Figures and Tables

Figure 1: Raw Achievement Gender Gap and Classroom Gender Segregation



Notes: The 33 countries from the regression sample are shown. *Raw Mathematics Gender Gap (% of country S.D.)* corresponds to the raw difference between achievement of boys and girls, divided by each country's standard deviation of average mathematics achievement (see Table 1 for details). *Classroom Segregation (%)* is based on the Reynal-Querol (2002) polarization index and measures how gender-segregated classrooms in the country are on average (Segregation of genders j in classroom c is calculated as $Seg_c = 1 - RQ_c = \sum_{j=1}^{2} 4(1-\pi_{jc})^2\pi_{jc}$, where a value of 1 corresponds to 100 % segregation and 0 is full gender parity). Note that in contrast to the gender parity index, no distinction is made with regard to over-accomplishment of gender parity; classes with either no girls or no boys are both assigned maximum inequality.

Figure 2: Average Achievement in Mathematics



Notes: Deviation of the country-level average of the TIMSS 2015 mathematics test score from 500 points (benchmark, zero line). The range of ±1 sample standard deviation is indicated by the dashed lines (approximately 60 points).

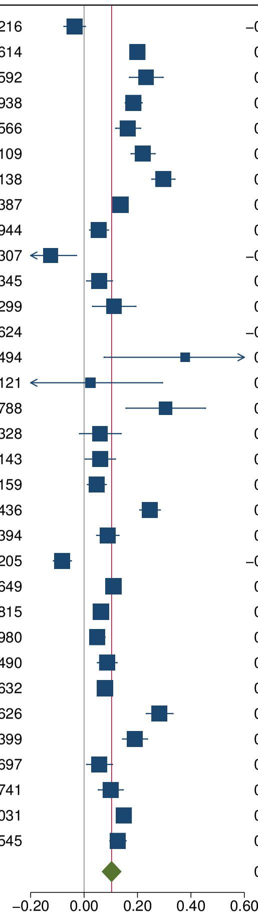
Table 1: List of Participating Countries and/or Sub-Regions

Country	ISO3C	Region	Average Avg. Age	Share of Migrants (%)	Math Score Average	Math Score Boys	Math Score Girls	Gender Segregation (%)	Number of Observations
Botswana	BWA	Africa	15.53 (0.92)	7.12	407.72 (78.94)	400.09 (83.60)	414.70 (73.74)	2.50	4,216
South Africa	ZAF	Africa	15.61 (1.13)	4.65	385.86 (78.57)	387.02 (79.15)	384.84 (78.06)	8.96	6,614
Buenes Aires, Argentina	ABA	Americas	13.96 (0.71)	10.64	414.19 (79.41)	420.94 (79.03)	408.14 (79.31)	16.67	1,592
Canada	CAN	Americas	13.97 (0.44)	14.74	535.61 (65.99)	538.84 (68.19)	532.53 (63.67)	11.57	6,938
Ontario, Canada	COT	Americas	13.81 (0.32)	15.57	523.14 (66.89)	526.44 (69.78)	519.93 (63.83)	8.23	3,566
Quebec, Canada	CQU	Americas	14.16 (0.49)	13.43	552.85 (60.69)	556.64 (61.90)	549.33 (59.34)	15.82	3,109
Chile	CHL	Americas	14.23 (0.64)	3.26	446.99 (78.50)	455.18 (77.80)	438.37 (78.33)	12.08	3,138
United States	USA	Americas	14.24 (0.47)	6.39	520.27 (78.79)	522.65 (80.41)	518.00 (77.17)	6.48	8,387
United Arab Emirates	ARE	Middle East	13.90 (0.77)	36.39	464.69 (94.31)	464.11 (99.97)	465.25 (88.47)	86.43	11,944
Abu Dhabi, UAE	AAD	Middle East	13.86 (0.73)	29.10	440.09 (91.42)	435.84 (98.46)	444.37 (83.54)	90.81	3,307
Dubai, UAE	ADU	Middle East	13.93 (0.75)	47.36	505.59 (93.37)	511.38 (94.55)	499.75 (91.81)	72.93	4,345
Bahrain	BHR	Middle East	13.94 (0.61)	23.72	466.71 (76.48)	463.09 (83.25)	469.98 (69.68)	74.86	3,299
Egypt	EGY	Middle East	14.07 (0.53)	6.62	406.24 (87.50)	404.20 (87.24)	408.40 (87.74)	72.07	4,624
Iran	IRN	Middle East	14.13 (0.41)	2.22	448.83 (91.59)	448.64 (95.41)	449.03 (87.43)	98.80	5,494
Jordan	JOR	Middle East	13.83 (0.42)	8.89	392.65 (81.13)	384.13 (83.88)	399.17 (78.36)	97.88	5,121
Kuwait	KWT	Middle East	13.74 (0.68)	19.91	391.10 (79.49)	385.87 (86.50)	395.81 (72.32)	91.75	2,788
Oman	OMN	Middle East	13.87 (0.86)	19.97	412.15 (90.42)	402.39 (95.19)	422.19 (84.09)	82.35	6,328
Qatar	QAT	Middle East	14.04 (0.81)	45.15	444.26 (96.35)	442.30 (100.08)	446.27 (92.36)	72.57	4,143
Turkey	TUR	Middle East	13.92 (0.57)	2.06	462.42 (99.66)	460.40 (100.02)	464.49 (99.28)	4.16	5,159
Hong Kong	HKG	Asia	14.27 (0.70)	22.40	590.81 (75.46)	592.96 (80.51)	588.56 (69.74)	19.11	3,436
Japan	JPN	Asia	14.45 (0.28)	1.05	587.00 (85.86)	585.74 (87.33)	588.21 (84.42)	5.21	4,394
South Korea	KOR	Asia	14.40 (0.32)	0.59	604.48 (81.91)	606.27 (85.55)	602.65 (77.98)	60.01	5,205
Malaysia	MYS	Asia	14.34 (0.33)	1.17	504.80 (85.68)	503.78 (89.94)	505.72 (81.65)	16.40	8,649
Singapore	SGP	Asia	14.36 (0.47)	13.45	616.77 (80.66)	610.81 (84.71)	622.90 (75.80)	23.57	5,815
Thailand	THA	Asia	14.41 (0.50)	4.03	449.66 (96.03)	441.39 (99.77)	456.54 (92.25)	14.15	4,980
Taiwan	TWN	Asia	14.26 (0.31)	1.53	602.41 (93.88)	603.55 (96.51)	601.25 (91.12)	6.09	5,490
England, UK	ENG	Europe	14.07 (0.30)	10.04	533.36 (73.83)	532.24 (74.28)	534.41 (73.42)	14.83	3,632
Ireland	IRL	Europe	14.43 (0.41)	15.21	528.97 (68.27)	531.20 (70.51)	526.93 (66.11)	43.41	3,626
Italy	ITA	Europe	13.79 (0.47)	7.04	501.71 (69.42)	505.93 (70.44)	497.65 (68.20)	5.40	3,399
Norway	NOR	Europe	14.72 (0.31)	8.98	517.39 (66.13)	517.70 (66.64)	517.09 (65.65)	4.51	3,697
Norway, 8th Grade	NO8	Europe	13.72 (0.32)	9.15	494.11 (61.14)	495.70 (61.73)	492.56 (60.55)	4.12	3,741
Australia	AUS	Oceania	13.97 (0.45)	15.30	514.57 (83.23)	518.63 (84.14)	510.75 (82.20)	22.34	7,031
New Zealand	NZL	Oceania	14.06 (0.34)	17.30	504.72 (83.42)	508.11 (86.69)	501.83 (80.43)	35.66	6,545
$\sum$									163,752

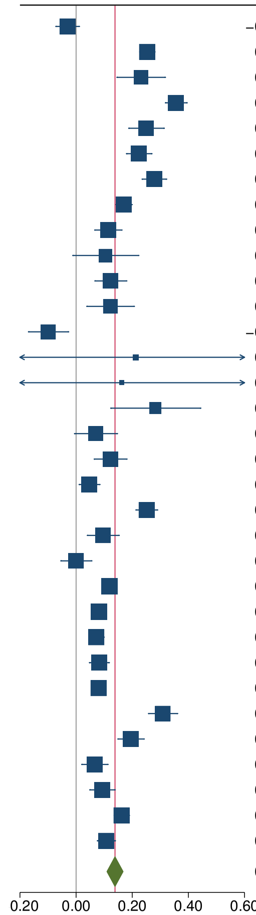
Notes: Standard deviations are reported in parentheses
12 countries/regions are excluded, mostly due to missing data on student values and beliefs. These are Georgia, Hungary, Israel, Kazakhstan, Lebanon, Lithuania, Malta, Morocco, Russia, Saudi Arabia, Slovenia, and Sweden.

Table 2: Forest Plot: Student Gender – Direct Impacts

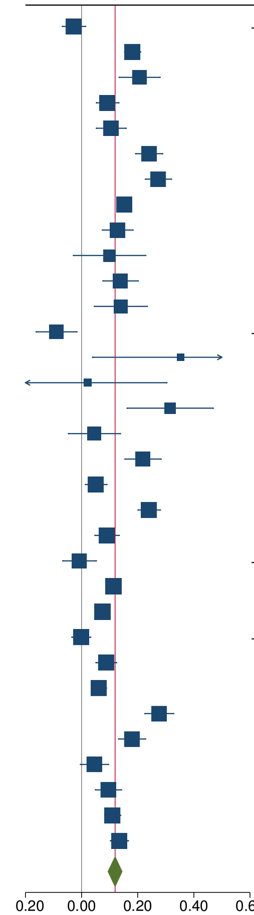
(a) Standard Linear Model



(b) Linear in Means



(c) Combined Homophily

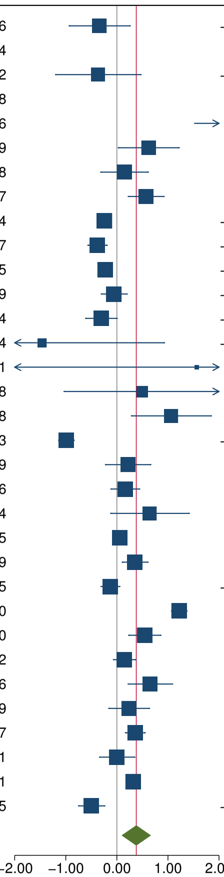


Study	N.Obs.	(a) Effect Size with 90% CI		(a) Weight (%)	(b) Effect Size with 90% CI		(b) Weight (%)	(c) Effect Size with 90% CI		(c) Weight (%)
Botswana	4,216	−0.04 [−0.08, 0.00]		3.22	−0.03 [−0.07, 0.01]		3.36	−0.03 [−0.07, 0.01]		3.39
South Africa	6,614	0.20 [0.18, 0.22]	***	3.29	0.25 [0.23, 0.28]	***	3.49	0.18 [0.15, 0.21]	***	3.55
Buenes Aires, Argentina	1,592	0.23 [0.17, 0.29]	***	3.05	0.23 [0.15, 0.32]	***	2.78	0.21 [0.13, 0.28]	***	2.88
Canada	6,938	0.18 [0.15, 0.22]	***	3.27	0.36 [0.32, 0.39]	***	3.39	0.09 [0.05, 0.13]	***	3.41
Ontario, Canada	3,566	0.16 [0.12, 0.21]	***	3.18	0.25 [0.19, 0.31]	***	3.10	0.10 [0.05, 0.16]	***	3.22
Quebec, Canada	3,109	0.22 [0.18, 0.26]	***	3.19	0.22 [0.18, 0.27]	***	3.32	0.24 [0.19, 0.29]	***	3.30
Chile	3,138	0.30 [0.25, 0.34]	***	3.20	0.28 [0.24, 0.32]	***	3.34	0.27 [0.23, 0.32]	***	3.32
United States	8,387	0.14 [0.11, 0.16]	***	3.31	0.17 [0.14, 0.20]	***	3.47	0.15 [0.13, 0.18]	***	3.58
United Arab Emirates	11,944	0.06 [0.02, 0.09]	***	3.25	0.11 [0.07, 0.16]	***	3.29	0.13 [0.07, 0.18]	***	3.19
Abu Dhabi, UAE	3,307	−0.13 [−0.22, −0.03]	**	2.74	0.10 [−0.01, 0.22]		2.32	0.10 [−0.03, 0.23]		1.96
Dubai, UAE	4,345	0.06 [0.01, 0.10]	**	3.17	0.12 [0.07, 0.18]	***	3.19	0.14 [0.08, 0.20]	***	3.06
Bahrain	3,299	0.11 [0.03, 0.19]	**	2.89	0.12 [0.04, 0.21]	**	2.80	0.14 [0.05, 0.23]	**	2.51
Egypt	4,624	−0.34 [−0.39, −0.29]	***	3.16	−0.10 [−0.17, −0.03]	**	2.99	−0.09 [−0.16, −0.02]	**	2.88
Iran	5,494	0.38 [0.07, 0.68]	**	1.03	0.21 [−0.27, 0.70]		0.34	0.35 [0.04, 0.67]	*	0.58
Jordan	5,121	0.02 [−0.24, 0.29]		1.22	0.16 [−0.50, 0.83]		0.19	0.02 [−0.26, 0.30]		0.70
Kuwait	2,788	0.30 [0.16, 0.45]	***	2.18	0.28 [0.12, 0.44]	***	1.78	0.32 [0.16, 0.47]	***	1.63
Oman	6,328	0.06 [−0.02, 0.14]		2.92	0.07 [−0.01, 0.15]		2.92	0.04 [−0.05, 0.14]		2.54
Qatar	4,143	0.06 [0.00, 0.12]	*	3.11	0.12 [0.06, 0.18]	***	3.17	0.22 [0.15, 0.28]	***	3.03
Turkey	5,159	0.05 [0.01, 0.08]	**	3.25	0.05 [0.01, 0.08]	**	3.41	0.05 [0.01, 0.09]	**	3.44
Hong Kong	3,436	0.25 [0.21, 0.28]	***	3.23	0.25 [0.21, 0.29]	***	3.39	0.24 [0.20, 0.28]	***	3.41
Japan	4,394	0.09 [0.05, 0.13]	***	3.21	0.10 [0.04, 0.15]	***	3.19	0.09 [0.05, 0.13]	***	3.37
South Korea	5,205	−0.08 [−0.11, −0.05]	***	3.26	0.00 [−0.05, 0.05]		3.21	−0.01 [−0.07, 0.05]		3.11
Malaysia	8,649	0.11 [0.09, 0.13]	***	3.31	0.12 [0.10, 0.14]	***	3.53	0.11 [0.09, 0.14]	***	3.61
Singapore	5,815	0.06 [0.04, 0.08]	***	3.31	0.08 [0.06, 0.10]	***	3.52	0.07 [0.05, 0.10]	***	3.61
Thailand	4,980	0.05 [0.02, 0.08]	***	3.28	0.07 [0.05, 0.10]	***	3.49	−0.00 [−0.03, 0.03]		3.50
Taiwan	5,490	0.09 [0.05, 0.12]	***	3.24	0.08 [0.05, 0.12]	***	3.42	0.09 [0.05, 0.12]	***	3.46
England, UK	3,632	0.08 [0.05, 0.10]	***	3.30	0.08 [0.06, 0.10]	***	3.50	0.06 [0.03, 0.09]	***	3.57
Ireland	3,626	0.28 [0.23, 0.33]	***	3.16	0.31 [0.26, 0.36]	***	3.25	0.28 [0.23, 0.33]	***	3.24
Italy	3,399	0.19 [0.14, 0.24]	***	3.18	0.20 [0.15, 0.24]	***	3.30	0.18 [0.13, 0.23]	***	3.30
Norway	3,697	0.06 [0.01, 0.10]	**	3.17	0.07 [0.02, 0.11]	**	3.30	0.04 [−0.00, 0.09]		3.27
Norway, 8th Grade	3,741	0.10 [0.05, 0.15]	***	3.18	0.09 [0.05, 0.14]	***	3.32	0.10 [0.05, 0.14]	***	3.32
Australia	7,031	0.15 [0.12, 0.17]	***	3.29	0.16 [0.14, 0.19]	***	3.49	0.11 [0.08, 0.14]	***	3.55
New Zealand	6,545	0.13 [0.10, 0.16]	***	3.27	0.11 [0.08, 0.14]	***	3.45	0.13 [0.10, 0.17]	***	3.52
Overall		0.10 [0.07, 0.14]			0.14 [0.11, 0.17]			0.12 [0.09, 0.15]		

Notes: Overall effect is obtained from a random-effects REML model. *: Significant at the 10% level. **: Significant at the 5% level. ***: Significant at the 1% level. *N.Obs.* refers to the number of observations. *Effect Size with 90% CI* refers to the direct impact for the student gender variable obtained from the estimated coefficients of Equation (1) alongside with its 90% confidence interval. *Weight (%)* refers to the relative importance of the respective cross-section in the calculation of the overall measure, based on the number of observations and the standard error of the effect. This is visualized in the forest plot, where the square box corresponds to the point estimate of the impact. The size of the box corresponds to the associated weight of the cross-section in the calculation of the overall effect, while the whiskers represent the 90% confidence interval, with arrows at the end if truncated. The diamond at the bottom represents the overall effect, with its center indicating the point estimate and its stretching corresponding to the 90% confidence interval. Note that in Panel (a), the point estimate for Egypt lies outside the axis range.

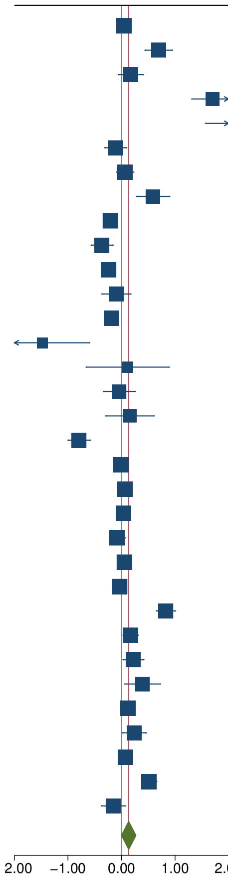
Table 3: Forest Plot: Student Gender – Indirect Impacts

(a) Linear in Means



Study	N.Obs.	Effect Size with 90% CI		Weight (%)
Botswana	4,216	−0.34 [−0.93, 0.25]		3.01
South Africa	6,614	2.46 [2.09, 2.84]	***	3.27
Buenes Aires, Argentina	1,592	−0.37 [−1.20, 0.46]		2.68
Canada	6,938	3.97 [3.51, 4.43]	***	3.17
Ontario, Canada	3,566	2.41 [1.51, 3.30]	***	2.59
Quebec, Canada	3,109	0.62 [0.03, 1.21]	*	3.01
Chile	3,138	0.15 [−0.32, 0.61]		3.18
United States	8,387	0.57 [0.22, 0.92]	***	3.29
United Arab Emirates	11,944	−0.24 [−0.31, −0.17]	***	3.46
Abu Dhabi, UAE	3,307	−0.38 [−0.57, −0.20]	***	3.41
Dubai, UAE	4,345	−0.23 [−0.33, −0.13]	***	3.45
Bahrain	3,299	−0.06 [−0.31, 0.19]		3.37
Egypt	4,624	−0.31 [−0.61, −0.00]	*	3.33
Iran	5,494	−1.46 [−3.84, 0.92]		1.02
Jordan	5,121	1.56 [−5.78, 8.91]		0.15
Kuwait	2,788	0.50 [−1.03, 2.03]		1.73
Oman	6,328	1.06 [0.28, 1.84]	**	2.76
Qatar	4,143	−0.99 [−1.14, −0.84]	***	3.43
Turkey	5,159	0.22 [−0.22, 0.66]		3.20
Hong Kong	3,436	0.16 [−0.12, 0.44]		3.35
Japan	4,394	0.64 [−0.12, 1.41]		2.78
South Korea	5,205	0.06 [−0.07, 0.19]		3.44
Malaysia	8,649	0.35 [0.10, 0.60]	**	3.37
Singapore	5,815	−0.13 [−0.31, 0.05]		3.42
Thailand	4,980	1.22 [1.07, 1.37]	***	3.43
Taiwan	5,490	0.54 [0.23, 0.85]	***	3.33
England, UK	3,632	0.15 [−0.07, 0.36]		3.40
Ireland	3,626	0.65 [0.22, 1.08]	**	3.21
Italy	3,399	0.24 [−0.16, 0.63]		3.25
Norway	3,697	0.36 [0.17, 0.55]	***	3.41
Norway, 8th Grade	3,741	0.00 [−0.34, 0.34]		3.30
Australia	7,031	0.32 [0.19, 0.45]	***	3.44
New Zealand	6,545	−0.50 [−0.75, −0.24]	***	3.37
Overall		0.38 [0.09, 0.67]		

(b) Combined Homophily

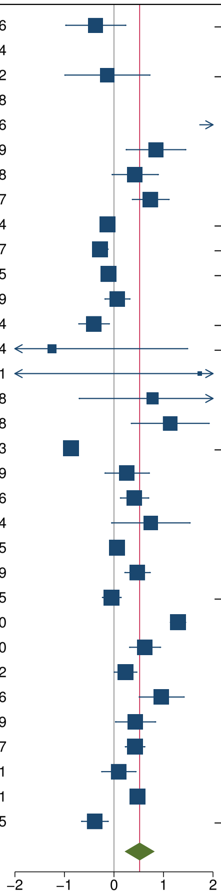


Study	Effect Size with 90% CI		Weight (%)
Botswana	0.04 [−0.08, 0.17]		3.28
South Africa	0.69 [0.44, 0.95]	***	3.05
Buenes Aires, Argentina	0.17 [−0.06, 0.40]		3.10
Canada	1.70 [1.31, 2.09]	***	2.70
Ontario, Canada	2.12 [1.56, 2.68]	***	2.23
Quebec, Canada	−0.11 [−0.31, 0.09]		3.16
Chile	0.07 [−0.09, 0.22]		3.23
United States	0.58 [0.28, 0.89]	***	2.91
United Arab Emirates	−0.21 [−0.28, −0.13]	***	3.34
Abu Dhabi, UAE	−0.37 [−0.57, −0.17]	***	3.16
Dubai, UAE	−0.24 [−0.34, −0.13]	***	3.30
Bahrain	−0.10 [−0.37, 0.17]		3.01
Egypt	−0.19 [−0.32, −0.06]	**	3.28
Iran	−1.48 [−2.35, −0.60]	***	1.50
Jordan	0.11 [−0.66, 0.88]		1.70
Kuwait	−0.04 [−0.34, 0.25]		2.95
Oman	0.15 [−0.30, 0.60]		2.53
Qatar	−0.79 [−1.00, −0.59]	***	3.15
Turkey	−0.01 [−0.14, 0.12]		3.28
Hong Kong	0.07 [−0.02, 0.16]		3.32
Japan	0.03 [−0.06, 0.13]		3.32
South Korea	−0.08 [−0.23, 0.06]		3.25
Malaysia	0.05 [−0.04, 0.14]		3.32
Singapore	−0.03 [−0.06, −0.00]	*	3.36
Thailand	0.83 [0.65, 1.00]	***	3.20
Taiwan	0.17 [0.03, 0.31]	**	3.26
England, UK	0.22 [0.03, 0.41]	*	3.18
Ireland	0.39 [0.06, 0.72]	*	2.86
Italy	0.12 [−0.02, 0.25]		3.27
Norway	0.24 [0.02, 0.45]	*	3.13
Norway, 8th Grade	0.07 [−0.05, 0.19]		3.29
Australia	0.52 [0.38, 0.65]	***	3.27
New Zealand	−0.16 [−0.38, 0.07]		3.11
Overall	0.13 [−0.01, 0.28]		

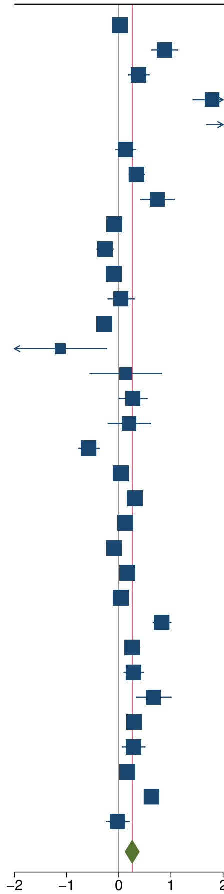
Notes: Overall effect is obtained from a random-effects REML model. *: Significant at the 10% level. **: Significant at the 5% level. ***: Significant at the 1% level. *N.Obs.* refers to the number of observations. *Effect Size with 90% CI* refers to the indirect impact for the student gender variable obtained from the estimated coefficients of Equation (1) alongside with its 90% confidence interval. *Weight (%)* refers to the relative importance of the respective cross-section in the calculation of the overall measure, based on the number of observations and the standard error of the effect. This is visualized in the forest plot, where the square box corresponds to the point estimate of the impact. The size of the box corresponds to the associated weight of the cross-section in the calculation of the overall effect, while the whiskers represent the 90% confidence interval, with arrows at the end if truncated. The diamond at the bottom represents the overall effect, with its center indicating the point estimate and its stretching corresponding to the 90% confidence interval. Note that in Panel (a), the point estimates for South Africa, Canada, and Ontario (Canada) lie outside the axis range. In Panel (b), Ontario (Canada) lies outside the axis range.

Table 4: Forest Plot: Student Gender – Total Impacts

(a) Linear in Means



(b) Combined Homophily



Study	N.Obs.	(a) Effect Size with 90% CI		(a) DI	(a) II	(a) Weight (%)	(b) Effect Size with 90% CI		(b) DI	(b) II	(b) Weight (%)
Botswana	4,216	–0.37 [–0.97, 0.23]		–.031	–.341	3.03	0.01 [–0.10, 0.13]		–.028	.043	3.26
South Africa	6,614	2.72 [2.33, 3.10]	***	.253***	2.465***	3.27	0.87 [0.63, 1.12]	***	.181***	.692***	3.04
Buenes Aires, Argentina	1,592	–0.14 [–0.99, 0.72]		.231***	–.368	2.70	0.38 [0.18, 0.57]	***	.206***	.172	3.14
Canada	6,938	4.33 [3.84, 4.81]	***	.356***	3.97***	3.17	1.79 [1.42, 2.16]	***	.092***	1.7***	2.73
Ontario, Canada	3,566	2.65 [1.72, 3.59]	***	.25***	2.406***	2.60	2.22 [1.68, 2.77]	***	.105***	2.117***	2.24
Quebec, Canada	3,109	0.84 [0.25, 1.44]	**	.224***	.62*	3.04	0.13 [–0.06, 0.31]		.24***	–.111	3.16
Chile	3,138	0.42 [–0.04, 0.89]		.278***	.146	3.19	0.34 [0.20, 0.48]	***	.273***	.065	3.23
United States	8,387	0.74 [0.37, 1.10]	***	.17***	.569***	3.28	0.74 [0.42, 1.05]	***	.152***	.585***	2.87
United Arab Emirates	11,944	–0.13 [–0.18, –0.08]	***	.114***	–.242***	3.44	–0.08 [–0.13, –0.03]	***	.128***	–.209***	3.32
Abu Dhabi, UAE	3,307	–0.28 [–0.43, –0.13]	***	.105	–.383***	3.42	–0.27 [–0.42, –0.12]	***	.099	–.366***	3.21
Dubai, UAE	4,345	–0.11 [–0.19, –0.02]	**	.123***	–.231***	3.44	–0.10 [–0.19, –0.01]	*	.138***	–.238***	3.29
Bahrain	3,299	0.07 [–0.18, 0.31]		.122**	–.056	3.37	0.04 [–0.21, 0.29]		.139**	–.099	3.03
Egypt	4,624	–0.41 [–0.71, –0.10]	**	–.099**	–.307*	3.33	–0.28 [–0.39, –0.17]	***	–.09**	–.189**	3.27
Iran	5,494	–1.25 [–3.97, 1.48]		.213	–1.461	0.90	–1.12 [–2.00, –0.24]	**	.353*	–1.476***	1.47
Jordan	5,121	1.73 [–6.23, 9.68]		.163	1.565	0.14	0.13 [–0.55, 0.81]		.021	.111	1.89
Kuwait	2,788	0.78 [–0.70, 2.27]		.283***	.5	1.88	0.27 [0.01, 0.53]	*	.315***	–.044	2.99
Oman	6,328	1.13 [0.35, 1.91]	**	.07	1.063**	2.80	0.20 [–0.20, 0.60]		.045	.154	2.63
Qatar	4,143	–0.87 [–1.01, –0.73]	***	.122***	–.993***	3.42	–0.58 [–0.77, –0.38]	***	.218***	–.793***	3.15
Turkey	5,159	0.26 [–0.18, 0.71]		.047**	.216	3.21	0.04 [–0.07, 0.15]		.051**	–.01	3.27
Hong Kong	3,436	0.41 [0.13, 0.69]	**	.251***	.16	3.35	0.31 [0.22, 0.40]	***	.24***	.069	3.29
Japan	4,394	0.74 [–0.05, 1.53]		.096***	.644	2.79	0.13 [0.03, 0.22]	**	.09***	.034	3.28
South Korea	5,205	0.06 [–0.07, 0.18]		0	.059	3.43	–0.09 [–0.22, 0.04]		–.008	–.084	3.24
Malaysia	8,649	0.47 [0.22, 0.72]	***	.119***	.353**	3.37	0.17 [0.08, 0.25]	***	.114***	.052	3.29
Singapore	5,815	–0.05 [–0.23, 0.14]		.081***	–.13	3.40	0.04 [0.01, 0.08]	*	.075***	–.033*	3.33
Thailand	4,980	1.29 [1.14, 1.44]	***	.072***	1.217***	3.42	0.83 [0.66, 0.99]	***	–.002	.827***	3.19
Taiwan	5,490	0.62 [0.31, 0.93]	***	.082***	.542***	3.33	0.25 [0.12, 0.39]	***	.087***	.167**	3.24
England, UK	3,632	0.23 [0.01, 0.45]	*	.08***	.148	3.38	0.28 [0.10, 0.46]	**	.061***	.219*	3.16
Ireland	3,626	0.96 [0.51, 1.41]	***	.309***	.651**	3.20	0.67 [0.34, 0.99]	***	.276***	.388*	2.83
Italy	3,399	0.43 [0.03, 0.83]	*	.195***	.236	3.25	0.30 [0.17, 0.42]	***	.179***	.118	3.25
Norway	3,697	0.42 [0.23, 0.61]	***	.066**	.356***	3.40	0.28 [0.07, 0.49]	**	.045	.237*	3.10
Norway, 8th Grade	3,741	0.09 [–0.25, 0.44]		.093***	.002	3.30	0.17 [0.05, 0.28]	**	.095***	.071	3.26
Australia	7,031	0.48 [0.34, 0.62]	***	.162***	.317***	3.42	0.63 [0.50, 0.76]	***	.11***	.515***	3.24
New Zealand	6,545	–0.39 [–0.65, –0.13]	**	.107***	–.498***	3.36	–0.02 [–0.24, 0.19]		.134***	–.156	3.10
Overall		0.52 [0.22, 0.82]					0.26 [0.12, 0.40]				

Notes: Overall effect is obtained from a random-effects REML model. *: Significant at the 10% level. **: Significant at the 5% level. ***: Significant at the 1% level. *N.Obs.* refers to the number of observations. *Effect Size with 90% CI* refers to the total impact for the student gender variable obtained from the estimated coefficients of Equation (1) alongside with its 90% confidence interval. *Weight (%)* refers to the relative importance of the respective cross-section in the calculation of the overall measure, based on the number of observations and the standard error of the effect. This is visualized in the forest plot, where the square box corresponds to the point estimate of the impact. The size of the box corresponds to the associated weight of the cross-section in the calculation of the overall effect, while the whiskers represent the 90% confidence interval, with arrows at the end if truncated. The diamond at the bottom represents the overall effect, with its center indicating the point estimate and its stretching corresponding to the 90% confidence interval. Note that in Panel (a), the point estimates for South Africa, Canada, and Ontario (Canada) lie outside the axis range. In Panel (b), Ontario (Canada) lies outside the axis range.

Zuletzt erschienen /previous publications:

V-433-20 **Bernhard C. Dannemann,** Better Off On Their Own? How Peer Effects Determine International Patterns of the Mathematics Gender Achievement Gap
V-432-20 **Christoph Böhringer, Carolyn Fischer**, Kill Bill or Tax: An Analysis of Alternative CO2 Price Floor Options for EU Member States
V-431-20 **Heinz Welsch**, How Climate-Friendly Behavior Relates to Moral Identity and Identity-Protective Cognition: Evidence from the European Social Surveys
V-430-20 **Christoph Böhringer, Knut Einar Rosendahl,** Europe beyond Coal – An Economic and Climate Impact Assessment
V-429-20 **Oliver Richters**, Modeling the out-of-equilibrium dynamics of bounded rationality and economic constraints
V-428-20 **Bernhard C. Dannemann,** Peer Effects in Secondary Education: Evidence from the 2015 Trends in Mathematics and Science Study Based on Homophily
V-427-19 **Christoph Böhringer, Knut Einar Rosendahl, Halvor Briseid Storrøsten,** Smart hedging against carbon leakage
V-426-19 **Emmanuel Asane-Otoo, Bernhard Dannemann,** Rockets and Feathers Revisited: Asymmetric Retail Fuel Pricing in the Era of Market Transparency
V-425-19 **Heinz Welsch,** Moral Foundations and Voluntary Public Good Provison: The Case of Climate Change
V-424-19 **Gökçe Akın-Olçum, Christoph Böhringer, Thomas Rutherford, Andrew Schreiber,** Economic and Environmental Impacts of a Carbon Adder in New York
V-423-19 **Jasper N. Meya, Paul Neetzow,** Renewable energy policies in federal government systems
V-422-19 **Philipp Biermann, Heinz Welsch**, Changing Conditions, Persistent Mentality: An Anatomy of East German Unhappiness, 1990-2016
V-421-19 **Philipp Biermann, Jürgen Bitzer, Erkan Gören,** The Relationship between Age and Subjective Well-Being: Estimating Within and Between Effects Simultaneously
V-420-19 **Philipp Poppitz,** Multidimensional Inequality and Divergence: The Eurozone Crisis in Retrospect
V-419-19 **Heinz Welsch,** Utilitarian and Ideological Determinants of Attitudes toward Immigration: Germany before and after the "Refugee Crisis"
V-418-19 **Christoph Böhringer**, **Xaquin Garcia-Muros, Mikel González-Eguino**, Greener and Fairer: A Progressive Environmental Tax Reform for Spain
V-417-19 **Heinz Welsch, Martin Binder, Ann-Kathrin Blankenberg**, Pro-environmental norms and subjective well-being: panel evidence from the UK
V-416-18 **Jasper N. Meya,** Environmental Inequality and Economic Valuation
V-415-18 **Christoph Böhringer, Thomas F. Rutherford, Edward J. Balistreri,** Quantifying Disruptive Trade Policies
V-414-18 **Oliver Richters, Andreas Siemoneit**, The contested concept of growth imperatives: Technology and the fear of stagnation
V-413-18 **Carsten Helm, Mathias Mier**, Subsidising Renewables but Taxing Storage? Second-Best Policies with Imperfect Carbon Pricing
V-412-18 **Mathias Mier**, Policy Implications of a World with Renewables,Limited Dispatchability, and Fixed Load
V-411-18 **Klaus Eisenack, Mathias Mier**, Peak-load Pricing with Different Types of Dispatchability
V-410-18 **Christoph Böhringer, Nicholas Rivers,** The energy efficiency rebound effect in general equilibrium
V-409-18 **Oliver Richters, Erhard Glötzl**, Modeling economic forces, power relations, and stock-flow consistency: a general constrained dynamics approach
V-408-18 **Bernhard C. Dannemann, Erkan Gören,** The Educational Burden of ADHD: Evidence From Student Achievement Test Scores

V-407-18 **Jürgen Bitzer, Erkan Gören**, Foreign Aid and Subnational Development: A Grid Cell Analysis
V-406-17 **Christoph Böhringer**, **Jan Schneider**, **Marco Springmann,** Economic and Environmental Impacts of Raising Revenues for Climate Finance from Public Sources
V-405-17 **Erhard Glötzl, Florentin Glötzl, Oliver Richters**, From constrained optimization to constrained dynamics: extending analogies between economics and mechanics
V-404-17 **Heinz, Welsch, Jan Kühling**, How Green Self Image Affects Subjective Well-Being: Pro-Environmental Values as a Social Norm
V-403-17 **Achim Hagen, Jan Schneider**, Boon or Bane? Trade Sanctions and the Stability of International Environmental Agreements
V-402-17 **Erkan Gören**, The Role of Novelty-Seeking Traits in Contemporary Knowledge Creation
V-401-17 **Heinz Welsch, Jan Kühling**, Divided We Stand: Immigration Attitudes, Identity, and Subjective Well-Being
V-400-17 **Christoph Böhringer, Thomas F. Rutherford**, Paris after Trump: An inconvenient insight
V-399-17 **Frank Pothen, Heinz Welsch**, Economic Development and Material Use
V-398-17 **Klaus Eisenack, Marius Paschen,** Designing long-lived investments under uncertain and ongoing change
V-397-16 **Marius Paschen**, The effect of intermittent renewable supply on the forward premium in German electricity markets
V-396-16 **Heinz Welsch, Philipp Biermann**, Poverty is a Public Bad: Panel Evidence from Subjective Well-being Data
V-395-16 **Philipp Biermann**, How Fuel Poverty Affects Subjective Well-Being: Panel Evidence from Germany
V-394-16 **Heinz Welsch**, Electricity Externalities, Siting, and the Energy Mix: A Survey
V-393-16 **Leonhard Kähler, Klaus Eisenack**, Strategic Complements in International Environmental Agreements: a New Island of Stability
V-392-16 **Christoph Böhringer, Xaquin Garcia-Muros, Ignacio Cazcarro, Iñaki Arto**, The Efficiency Cost of Protective Measures in Climate Policy
V-391-16 **Achim Hagen, Juan-Carlos Altamirano-Cabrera, Hans-Peter Weikard**, The Influence of Political Pressure Groups on the Stability of International Environmental Agreements
V-390-16 **Christoph Böhringer, Florian Landis, Miguel Angel Tovar Reaños,** Cost-effectiveness and Incidence of Renewable Energy Promotion in Germany
V-389-16 **Carsten Helm, Mathias Mier,** Efficient diffusion of renewable energies: A roller-coaster ride
V-388-16 **Christoph Böhringer, Jan Schneider, Emmanuel Asane-Otoo,** Trade In Carbon and The Effectiveness of Carbon Tariffs
V-387-16 **Achim Hagen, Leonhard Kähler, Klaus Eisenack,** Transnational Environmental Agreements with Heterogeneous Actors