Gender Gaps in Student Academic Achievement and Inequality

Shu-Ling Tsai (Institute of Sociology, Academia Sinica, Taipei)
Michael L. Smith (Economics Institute, Czech Academy of Sciences, Prague)
Robert M. Hauser (Department of Sociology, University of Wisconsin, Madison)

Abstract

Results from international large-scale assessments, such as PISA surveys, suggest that boys do better in math and science, whereas girls do better in reading. How do gender gaps vary across subjects, when estimated simultaneously? Building on the work of Tsai, Smith, and Hauser (2017), we answer this question by applying a multilevel MIMIC model that enables us to estimate gender gaps in two ways: gender differences in the effects of observed family and school factors on math, science, and reading scores; and the "adjusted" gender gaps in test scores across all three subjects after controlling for observables. We apply the model to 2012 PISA data of students aged 15-16 and enrolled in 9th or 10th grade in three East Asian (Japan, South Korea, and Taiwan) and three Western countries (USA, Germany, and the Czech Republic) that represent both similar and different types of school systems. Our findings indicate that the gender gap in math or science achievement in Western countries, favoring boys, does not necessarily apply to the East Asian countries examined here, while all three East Asian countries exhibit similar features of gender reading gaps in the 10th grade. There is evidence indicating that background and school factors impact boys' and girls' achievement in a similar way in USA, Japan, Korea, Taiwan, and the Czech Republic, but not in Germany. Overall, gender differences in family and school influences do not account for gender differences in academic achievement in any of the six countries.

Paper accepted for publication in *Research in the Sociology of Education*.

Introduction

Referring to student performance in course grades, Legewie and DiPrete (2012: 463) stated that "Today, boys generally underperform relative to girls in schools throughout the industrialized world." But does this apply the same way for industrialized Western and East Asian countries, as well as across subjects and grades? Recent research on gender gaps in standardized achievement test scores among adolescent students has provided mixed conclusions, with some evidence favoring boys and some favoring girls. In particular, results from international large-scale assessments suggest that boys surpass girls in math and science performance in most countries, whereas girls do better in reading in all countries participating in PISA surveys (OECD 2015). Scholars have suggested that the size of the gender gap in math, science, or reading varies widely across countries, and that greater variance in boys' achievement compared to girls' is not universal (e.g., Bedard and Cho 2010; Dickerson, McIntosh, and Valente 2015; Fryer and Levitt 2010; Guiso et al. 2008; Machin and Pekkarinen 2008; Marks 2008; Penner 2008).

In this article, we use 2012 PISA data to assess gender gaps in math, science, and reading achievement and inequality among students aged 15-16 and enrolled in the 9th or 10th grades in three East Asian countries (Japan, South Korea, and Taiwan) and three Western countries (USA, Germany, and the Czech Republic) that represent both similar and different types of school systems. From a global perspective, the East Asian countries exhibit a number of commonalities, such as a low degree of tracking before upper secondary education, a standardized system of secondary education without dead-end tracks, the use of centralized tests as the mechanism of selection into higher levels of education, and the high cultural expectations by both schools and parents for student achievement. The educational systems of East Asian countries also have high levels of student achievement and low levels of dispersion compared to other regions of the world (Park 2010; Park and Sandefur 2006), qualities that makes them ideal candidates for comparative research on gender gaps in academic achievement.

By contrast, Germany and the Czech Republic were selected as examples of a German model of education that sharply contrasts with the East Asian model. The German model has deep historical roots across Central Europe, characterized by early educational selection into different tracks of lower secondary education, the presence of dead-end tracks in upper secondary education, and vocationally specific study programs that provide pupils with focused labor market opportunities but not credentials for tertiary education. As a contrast to both the East Asian and German models of education, we also include the United States in our study, with its distinctive system of comprehensive secondary schools characterized by relatively little institutional differentiation but substantial within-school tracking. Key similarities and differences in all six educational systems are summarized in Appendix Table A-1.

We focus on these different educational systems in order to shed light on the degree to which differences in school systems may account for cross-national variation in the effects of family and school factors between boys and girls and across subjects. While it would be ideal to extend our study to a broader number of countries, our analytical approach is both newly developed and computationally demanding, making a broader study prohibitively difficult. We detail our methodology so our approach can be replicated on other country datasets. But by focusing on a smaller number of leading industrialized nations with sharply different educational systems, we may be able to identify similarities and differences in the pattern of gender gaps that could be relevant to a broader set of nations.

Analytically, we employ a multilevel MIMIC (multiple indicators and multiple causes) model to assess gender gaps in academic achievement in two different ways: gender differences in the overall effect of family and school factors on math, science, and reading; and “adjusted” gender gaps in math, science and reading scores, taking into account family and school factors specified in the model. The key analytic contribution of the multilevel MIMIC model is that it enables the simultaneous estimation of how reading, math, and science achievement respond to variation in family background, among schools, and between genders and countries; other statistical approaches do not allow this simultaneous estimation, which typically forces scholars to focus on only one measure of academic achievement at a time.

Our analysis is built upon the work of Tsai, Smith, and Hauser (2017), in which the multilevel MIMIC model is developed and applied to the above six countries. In that study, the authors demonstrate that the impact of family background and school attended on math, science, and reading performance can vary greatly within a country, and that in some countries, performance inequality is largest in subjects with the highest mean scores, suggesting a modest tradeoff between inequality and efficiency. They also found that in Taiwan and Korea, family and school factors impact math performance to a greater degree than reading, and impact reading performance more than science. By contrast, in Japan and the United States, family and school effects are stronger on science performance than on math or reading performance, and in Germany and the Czech Republic, the family and school effects on math and science are larger than those on reading performance. Tsai, Smith, and Hauser's (2017) cross-national comparisons, however, do not consider the potential impact of gender, which is a key dimension of inequality in student academic performance in OECD and partner nations (OECD 2015).

Previous Research

Three main types of explanation have been proposed to account for differential gender gaps by subject. A first popular explanation is biological: men have greater aptitude and interest in math and science by nature, as do women in reading. Biological arguments are based in genetic, hormonal, and neurological considerations (see, e.g., Ceci, Williams, and

Bernett 2009; Penner 2008). However, the biological explanation cannot account for gender parity in math and science performance in some countries like Norway and Sweden, let alone Israeli girls' outperformance of boys. Nor can it explain the variations in the size of the gender gap in achievement test scores across individuals with different characteristics, both within and between countries. It is thus likely that gender gaps in student academic performance depend on an interaction between nature and nurture, rather than nature alone.

A second prominent type of explanation is cultural, based in gender ideologies and beliefs that have implications for math, science, and reading ability. According to this view, cultural influences steer boys toward non-literary activities and define reading as a feminine characteristic, whereas mathematical and scientific activities are defined as masculine. In particular, psychologists suggest that gender stereotypes are widely held by parents and teachers, and these produce stereotype threats to students. In turn adolescent students reproduce gender-typed behaviors in terms of course taking, college plans, and career choices, through psychological mechanisms such as self-efficacy (Bandura 1986) or achievement motivation based on expectancy-value (Eccles 2011ab). Psychologists also suggest that due to cultural shifts toward gender equity today, the two sexes are similar on most psychological and cognitive variables, with the consequence that the gender gap in math or science performance has largely closed, and that gender differences in course taking are trivial in all subjects, except physics, during high school years (e.g., Hyde 2005; Hyde and Linn 2006; Hyde et al. 2008; Lindberg et al. 2010).

Sociological research focuses on a third explanation, namely social or structural effects. Sociologists are keen to attribute the observed gender gap as an outcome of the social production of gender differences in the life-long process of social stratification, in which families, schools, and labor markets are all involved (see, e.g., DiPrete and Buchmann 2013; Xie and Shauman 2003). Sociologists recognize the importance of culture in gender socialization, but most view gender ideologies as external constraints, rather than internalized preferences; even discussions of cultural influences typically focus more on how stereotypes yield biases that underestimate women's ability or overestimate men's, and less on how internalized gender ideals restrains women's aspirations and achievements (England 2016). In the sociological literature, beliefs and stereotypes about status characteristics, not just gender *per se*, are believed to produce profound effects on individual outcomes. It has been found that beliefs about gender status – plus the expectations for rewards and costs that they produce – affect individuals' performances on cognitive tests (Lovaglia et al. 1998) and self-perceptions of performance (Correll 2004). Along this vein, Penner (2008) suggests that the size of the gender gap in math is determined by gender status and incentive structures within which students decide whether to pursue more math education or math-intensive fields.

To explain why incentive structures at the national level are decisive for academic performance at the individual level, some view education as an investment decision, with a focus on how the family influences an individual's human capital formation. There are at least two possibilities. One is that parents largely influence students' educational expectations, which in turn determine their efforts in school. Another possibility is that students first develop interests in certain subjects by nature and then base their own career expectations and effects in school on these interests. Whether students are pushed into their academic achievements by aggregate social influences or realize these achievements by their own purposive choice is a debated issue that has not yet been settled (see, e.g., Andrew and Hauser 2012; Legewie and DiPrete 2014; Morgan, Gelbgister, and Weeden 2013). While educational and career expectations are strongly related to educational outcomes, irrespective of gender, Marks (2008) finds that gender differences in these expectations do *not* account for gender gaps in reading and math, as observed in the 2000 PISA data.

Many scholars of gender argue for the primacy of structural factors in causing gender inequality or equity in academic achievement. Baker and Jones (1993), for example, maintain that girls' poorer math achievement and more negative math attitude are the result of societal gender stratification. The authors test a "gender stratification hypothesis," according to which gender parity in educational and occupational opportunities among adults leads to gender parity in math performances among eighth-grade students. Similar gender stratification hypotheses have been tested by researchers in other disciplines (e.g., Dickerson, McIntosh, and Valente 2015; Else-Quest, Hyde, and Linn 2010; Guiso et al. 2008; Riegle-Crumb 2005; Stoet and Geary 2013). Using numerous measures of gender equity in the home and family, in higher education, in the labor force, and in the political domains, this strand of research seeks to establish a significant link from the structure of opportunities for men and women to the performance of children in school. However, the evidence is mixed and inconclusive. For example, Guiso et al. (2008) find in 2003 PISA data that the gender gap in math is strongly correlated with various measures of gender equality, whereas Stoet and Geary (2013) find *no* evidence in four PISA assessments (2000, 2003, 2006, and 2009) that gender disparities in math and reading are related to national gender equality indicators.

Another important strand of research focuses on school factors and attributes the gender gap in student performance to the institutional arrangements of national educational systems. For example, the within-country gender gap in reading has been attributed to school practices, on the hypothesis that boys are more sensitive than girls to school resources that create a learning-oriented environment (Legewie and DiPrete 2012), and that the gender-specific formation of study plans is shaped by the local environment of high schools (Legewie and DiPrete 2014). In international comparisons, it is fashionable to argue that cross-national variations in the size of the gender gap are due to the way students are allocated into different types of schools or tracking programs. Yet there are contradictory findings in the literature. For example, Marks' (2008) comparison of 31 countries participating in the 2000 PISA survey indicates that school system factors are associated with the gender gap in reading but *not* in math, whereas Bedard and Cho (2010) find in the 1995, 1999, and 2003 TIMSS data that the degree of academic tracking is correlated with the gender gap in math across developed countries. That is, countries with more heavily tracked school systems tend to have larger female-male test score gaps, while countries with less institutional differentiation have smaller gender gaps, at least in math.

Set-up and hypotheses

Our work diverges from other gender-gap studies in several ways. First, we use more recent PISA data (2012) to examine gender differences in math, science, and reading achievement and inequality at the same time, compared to previous research that typically looks at the gender gap in just one or two subjects examined side-by-side. Second, we gauge not only the "observed" gender gaps in three subjects, but the "adjusted' gender gap across all three subjects after controlling students' observable family and school characteristics. Third, we test for gender similarities and differences in the effects of observed family and school factors on student academic performance, within the framework of the multilevel MIMIC model, in which family and school effects are estimated simultaneously for the three subjects. This allows us to explicitly test whether the several academic achievement constructs respond similarly to variation in family background and variation among schools for each of the populations examined. We can thus achieve strict comparability in estimating the size of the gender gap across subjects, which would not be the case if the gender gaps were estimated separately. Fourth, we compare 10th graders in six highly industrialized countries that have somewhat similar and different school systems and cultural norms, and thus extend the work of Tsai, Smith and Hauser (2017) to include effects of gender. Finally, we also compare patterns of gender gaps across 9th and 10th graders for three countries where data are available

(Taiwan, the Czech Republic, and Germany). While the German and Czech educational systems already exhibit early tracking in the lower secondary level, in the case of Taiwan the transition from 9th to 10th grade denotes the first critical movement across major institutional divisions – that is, from lower to upper secondary education – with rules of student recruitment shifting from equal access to meritocratic selection.

Our study explores three distinct questions about gender gaps in academic achievement. First, how do family background and school factors impact overall academic achievement – modelled as a composite of reading, math and science performance - differently by gender? If school tracks are "gendered" – not simply in the distribution of boys and girls in them, but in terms of the differential effect of tracks on overall academic achievement – we would expect that a dummy variable for that track would have a significantly different effect on boys' achievement compared to girls' achievement. If they are not "gendered," then we would observe relatively homogeneous coefficients of that track on boys and girls; those coefficients would be positive or negative depending on their contribution to latent achievement, regardless of gender. Similarly, family conditions, such as educational resources in the home, could also be distributed to boys and girls differently in ways that impact their overall academic achievement. However, given that we are analyzing highly industrialized societies that nominally practice non-discrimination in the distribution of educational resources by gender, we do not hypothesize large gender differences in school and family effects on overall academic achievement.

Second, how do family background and school factors, taken together, determine achievement in specific subjects – particularly math, science, and reading – differently between boys and girls? Even if we may find that there are not significant gender differences in the size of those effects on overall academic achievement, there may be gender differences in effects between subjects. Past research (Tsai, Smith and Hauser 2017) found that family background and school factors more strongly shape math and science achievement than reading achievement, but did not explore gender differences. We would hypothesize that family and school factors would more strongly effect reading achievement among boys, and math and science achievement among girls, particularly in more stratified school systems like in Germany or the Czech Republic, where boys and girls would be sorted to different types of school with different exposures to these subjects.

Third, after taking into account the effects of social background and schools in the MIMIC model, are gender gaps in reading, math and science scores larger, smaller or the same compared to raw measures of gender gaps in scores? By 'raw' measures, we mean the face value gender differences in PISA test scores that are often reported by the OECD, and which are not adjusted for any differences in the learning or family environments of boys and girls. By contrast, the multilevel MIMIC model is distinctive in the possibility of estimating intercepts in reading, math and science within the same model and adjusted by parameters at the individual and school levels and by error variances. We hypothesize that, in the more stratified school systems, the adjusted gender gaps according to the MIMIC model would shift in the direction favoring boys, compared to the raw gender gaps – due to the way the model takes into account how girls and boys are sorted into different tracks providing more academic versus vocational training, respectively. Even if school factors produce similar effects for boys and girls on overall achievement (the first hypothesis), differences in the allocation of boys and girls into distinct types of schools, which can differ considerably in their mean academic achievement, can yield estimated gender gaps that are smaller than those computed by the OECD. Particularly in highly stratified school systems, where boys are more likely to attend technical or vocational schools that generally underperform compared to academic schools, we expect to find less evidence of underachievement by boys after that sorting process is taken into account. Likewise, we do not expect large differences in adjusted and raw gender gaps in the less stratified school systems in East Asia and the United States.

Methods and Data

We test gender similarities and differences in the determination of academic achievement inequality within the framework of a multilevel MIMIC model. Compared to a conventional regression model, the MIMIC model presents a more parsimonious specification in which proportionality constraints are placed on the entire matrix of coefficients estimated in a single population (Hauser 1972; Hauser and Goldberger 1971). To estimate the family and school effects for each population examined, we employ a two-level MIMIC Model of student academic achievement developed in Tsai, Smith, and Hauser (2017). In algebraic form, the two-level MIMIC model depicted in Figure 1 consists of two levels and two equations in each level, which can be expressed as:

Level 1: Within-school Model

$$\eta_{ij} = \Sigma\beta_{kj}X_{kij} + \xi_{ij} , \tag{1}$$

$$Y_{cij} = \lambda_{cj}\eta_{ij} + \varepsilon_{cij} , \tag{2}$$

where η_{ij} is a latent variable representing "overall" academic achievement for student *i* in school *j*, which is influenced by the student's family background characteristics (X_{kij}) and indicated by the student's test scores in math, science, and reading (Y_{cij}, c = 1, 2, 3); β_{kj} are the coefficients of background variables among students in school *j*; λ_{cj} are the student-level factor loadings for school *j*; and ε_{cij} and ξ_{ij} are error terms for student *i* in school *j* in the two within-school equations.

Level 2: Between-school Model

$$\eta_{j} = \Sigma\gamma_{p}Z_{pj} + \xi_{j} , \tag{3}$$

$$Y_{cj} = \lambda_{c0j} + \lambda_{c}\eta_{j} + \varepsilon_{cj} , \tag{4}$$

where the four dependent variables (η_j and Y_{cj}, c = 1, 2, 3) are all unobservable variables; η_j represents "overall" academic achievement for school *j*, which is influenced by school factors (Z_{pj}) and indicated by the average test scores in math, science, and reading for school *j* (Y_{cj}), which are actually the "latent" intercepts estimated in equation 2; γ_p are the coefficients of *p* school factors; λ_c are the school-level factor loadings, and λ_{c0} are the intercepts estimated in equation 4; and ξ_j and ε_{cj} are error terms for school *j* in the two between-school equations.

Note that the force of the MIMIC model is to place proportionality constraints on the reduced form coefficients estimated in a single population. That is to say, if the model is true, then in the population there are proportionality constraints on coefficients:

$$b_{kc}/b_{k'c} = \beta_{kc}/\beta_{k'c} \text{ (for } k \neq k'\text{)},$$

and

$$b_{kc}/b_{kc'} = \lambda_c/\lambda_{c'} \text{ (for } c \neq c'\text{)}.$$

Results not shown here indicate that the proportionality constraints in the MIMIC model are valid, as far as our analysis is concerned.

Several other notes are important here. First, at the between-school level, the coefficients of the student background variables are assumed not to vary across schools:

$$\beta_{kj} = \beta_k. \tag{5}$$

Second, the regression of the within-school factors on X variables is assumed to have a random intercept varying across schools. Accordingly, the factor loadings are constrained to be equal across the within and the between levels:

$$\lambda_{cj} = \lambda_c. \tag{6}$$

Third, within the MIMIC model, one outcome variable is set as a reference index to identify the relative factor loadings of other outcome variables. In this analysis, math test scores are set as the reference index at both student and school levels.

Fourth and more importantly, there are two alternative ways to specify error variances and covariances: one allows covariances among outcome variables, but no disturbances in latent factors; the other allows disturbances in the latent factors (η), but no covariances among outcome variables. The latter is stronger and more restrictive than the former. We have estimated both model specifications. To save space, in this article we report results obtained from the stronger MIMIC model, as depicted in Figure 1. Parallel results yielded by the less restrictive type of model specifications are available upon request.

(Figure 1 about here)

In Figure 1, the rectangles represent observed variables, including family background variables (X), school factors (Z), and outcome variables (Y) at the individual level. The circles represent latent variables. The arrows represent regression relationships between variables, whereas the curved, double-headed arrows represent covariances between variables. In the within-school part of the model, the filled small circles at the end of the arrows from the within factor η_{ij} to Y_{1ij}, Y_{2ij}, and Y_{3ij} represent random intercepts that are referred to as Y_{1j}, Y_{2j}, and Y_{3j} in the between-school part of the model. These intercepts, varying across schools, are indicators of the between-school factor η_j. Note that there are no curved, double-headed arrows on the right-hand side of the model depicted in Figure 1. The indicator disturbances (i.e., error terms in Y) are specified as mutually independent, because the latent variable (η) is able to capture the potential positive correlations between subjects, in the presence of a disturbance term ($\xi \neq 0$). The basic idea of this model specification is that the latent variable (η) is a single latent variable that accounts for the covariance of the Y variables; once the effects of the X and Z variables via η are removed, there should no longer remain any correlation among the Y variables. It is important to note that the reduced-form disturbances under this model are not independent of each other, since they all have the disturbance ξ in common.

With respect to data considerations, in this analysis we use only the family and school variables in PISA data that are available for all six countries. Further, we restrict the analysis sample to students with complete information on the family and school variables. Family factors (X) include the following: (1) father's years of schooling, (2) mother's years of schooling, (3) the highest occupational status measured by the International Socioeconomic Index (ISEI) of either parent, (4) number of books at home, with response categories: 1 = 0-10 books, 2 = 11-25 books, 3 = 26-100 books, 4 = 101-200 books, 5 = 201-500 books, and 6 = more than 500 books, and (5) home educational resources, with items indicating the total sum of whether or not the student has a desk to study at, a quiet place to study, a computer to use for school work, educational software, textbooks to help with school work, technical reference books, and a dictionary.

School factors (Z) include two dummy variables indicating whether the school attended is private or public and whether the school is located in a rural (village, small town, or town) or urban area (city or large city). The school location variable may also be interpreted as a community factor. Other dummy variables are included to indicate distinctive types of secondary schools in each country and grade level. We consider the following school types for the 10th grade: one type for the United States (comprehensive schools); two types for Japan and Korea (academic/general and vocational high schools); and three types for Taiwan (senior high schools, senior vocational schools, and comprehensive high schools, in which both academic and vocational tracks are available), the Czech Republic (Gymnasium, technical schools with the school-leaving exam, and vocational schools without the school-leaving exam, including special schools), and Germany (Gymnasium, comprehensive schools (Realschule), and vocational schools (Hauptschule), including basic schools). For 9th grade, we consider one school type for Taiwan (comprehensive schools), two types for the Czech Republic (Gymnasium and upper-level elementary schools), and three types for Germany (Gymnasium, Realschule, and Hauptschule). Note that while the term "comprehensive school" is officially used in the Taiwanese, German, and American systems, its meaning and content differs across these countries and, in the Taiwanese case, between 9th and 10th grades.

The two-level MIMIC model is estimated on the analysis sample weighted by the student and school sampling probabilities provided in the PISA data, for each case examined. Sample selection and sample size are reported in Appendix Table A-2, which also indicates that missing data is a serious problem in the Czech Republic and Germany. While there are similar gender differentials in the whole and analytic samples, there is still reason for caution in interpreting the findings for those two countries. We also note that in most subjects and grade levels, differences are larger in the entire sample than among those with complete data used in the analysis. Unfortunately, it is not possible to obtain results from the whole sample, as the model cannot be run with missing values. To ensure that our key results comparing the size of adjusted and unadjusted gender gaps are not biased by missing data, we use the same analytic sample for both, as well as use that sample for all other results in the study.

Results

Gender Gaps in Mean and Variance of Test Scores

A primary objective of this analysis is to describe the pattern of differential gender gaps by subject. It is therefore constructive to begin by describing gender and cross-national differences using internationally comparable test scores. Data for this study are derived from the 2012 PISA surveys, which are internationally standardized assessments of students who are aged 15-16 at the time of the assessment. PISA data enable us to analyze both 9th and 10th graders in Taiwan, the Czech Republic, and Germany, but only 10th graders in Japan, Korea, and USA, because almost all Japanese and Korean students and most American students participating in the PISA surveys are in the 10th grade. While the three East Asian countries share several commonalities in school systems, they differ in cutoff dates for school entry: in Taiwan school starts in the fall, but in the spring in Japan and Korea.

We first gauge the mean test scores for boys and girls as well as mean differences between girls and boys, by subject, grade, and country, based on the analytic sample weighted by the student and school sampling probabilities provided in the PISA data; see Appendix Table A-3. To describe the observed size of gender gap, many researchers (e.g., Hyde 2005; Xie and Shauman 2003) use Cohen's *d*-statistics, which is defined as the mean test score for girls minus the mean test scores for boys, divided by the pooled within-gender standard deviation). Thus, a positive value of *d* represents a female advantage and a negative value of *d* represents a male advantage in average performance. Likewise, we use the *d*-statistics to describe the size of the observed gender gap. We depict the value of *d*-statistics by subject, grade, and country in Figure 2. The significance sign in the figure indicates the test results for the null hypothesis that there is no mean difference in test scores between girls and boys.

(Figure 2 about here)

Figure 2 confirms three regularities established in the international testing results. First, girls do better than boys in reading in every country, irrespective of grade. Second, boys outperform girls in math in all countries except Taiwan, as well as in 10^{th} grade science in all countries except Taiwan and Germany. The data for Taiwan is notable in that its variance in math achievement is particularly large, especially in the 9^{th} grade, but that this variance is not explained by gender. Overall, Figure 2 reveals that the size of the gender gap in math, science, or reading varies widely across countries, as well as between grades within countries.

For the 10^{th} grade, we find that the female disadvantage in math is larger than the female disadvantage in science in all cases, and that the female advantage in reading is relatively larger in Germany, Taiwan, and the Czech Republic than in the other three countries. Comparing 9^{th} and 10^{th} grades, we find that the pattern of differential gender gaps by subject varies with grade and country. In contrast to Germany and the Czech Republic, where boys outperform girls in math in 9th and 10th grades, in Taiwan there is no sign of male advantage in math or science in either grade, while the size of the female advantage in reading slightly declines across grades.

Next, we gauge the variance in test scores in each subject for boys and girls, separately, which is decomposed into the within-school variance and the between-school variance. The lower the share of between-school variance (% between), the lower the degree of inequality in educational outcomes across schools. Table 1 presents the total amount of variance and the percentage of between-school variance by subject, grade, and country for each gender, based on the analytic sample and weighted by the student and school sampling probabilities.

(Table 1 about here)

In Table 1, total variance in nearly all subjects and countries is larger for boys than for girls; boys' between-school variance also tends to be similar or slightly larger than for girls. A striking finding is that Taiwan has a very small percentage of between-school variance in the 9th grade, irrespective of subject and gender, while it also has a very large variance in student test scores in that grade, especially in math and among boys. Large variances in math performance in Taiwan have also been reported in other studies (e.g. Huang 2009; Montt 2011) that have found that the dispersion in math scores in Taiwan is among the largest in the world. In our data, we find that variance in each subject is largely reduced in the 10th grade, while the share of between-school variance increases from 9th to 10th grade. We suggest that this change is due to the Taiwanese school system, which sorts students into different types of senior high schools (or different tracks within schools). The use of examinations in student selection is equivalent to sorting students based on their previous academic ability. As a consequence of the positive relationship between students' academic performance and their socioeconomic background, high SES students are more likely than low SES students to enter good schools in the 10th grade. By contrast, in the 9th grade the allocation of students into different junior high schools is independent of student's academic ability, and thus most variance among students occurs within schools, indicating that inequality in educational outcomes within schools is much more profound than inequality between schools.

Similar to Taiwan, total variance in test scores in the Czech Republic declines from the 9th to 10th grades, while the between-school variance increases across those grades. We suggest that this is also due to the role of sorting mechanisms in the transition to upper secondary schools in the 10th grade. While Czech pupils in the 9th grade are already more sorted than in Taiwan (due to the presence of 6-year and 8-year gymnasia, which are elite, academically oriented schools teaching approximately 10% of most talented pupils based on competitive entrance exams), in the transition to upper secondary schools sorting mechanisms are even more profound, as the remaining 90% of pupils are sorted into 4-year gymnasia (which, along with 6-year and 8-year gymnasia, constitute "academic schools"), technical schools, and vocational schools without a school-leaving exam. This sorting increases between-school variance as a share of total variance. In Germany by contrast, which has an even more diversified school system at the lower secondary level, there is little change in total variance and the share of between-school variance between 9th and 10th grades. At the opposite extreme, between-school variance is the smallest in the United States among the analyzed countries, possibly due to relative scarcity of selective secondary schools there. For further details on the distribution of boys and girls by type of school and by other conditions, please consult Appendix Table A-4 for the 9th grade and A-5 for the 10th grade.

Model fit of Multilevel-MIMIC Models

To test for gender similarities and differences in the determination of academic achievement inequality, we first use the gender-specific data to estimate the effects of the observables for boys and girls separately, and then use the pooled data to test for similarities and differences in these effects between the two gender groups. That is to say, we start from estimating the model depicted in Figure 1 separately for boys and girls (Model A1 and A2). We then adapt the model to an equivalent one with a two-group comparative design, and estimate a series of models ranging from a gender differences model (Model A3) in which all structural parameters estimated (λ_{c0}, λ_c, β_k, and γ_p) vary across the two gender groups, to a gender similarities model in which all the above parameters are invariant across the two gender groups (Model F). To evaluate model fit, we use the sample-size adjusted BIC (Bayesian Information Criterion) statistic, given in Mplus 7. The BIC statistic combines negative two times the log likelihood of the postulated model, the number of parameters, and the sample size. As a general rule, the model with a lower BIC provides a better fit to the data, implying a more plausible representation of reality.

Table 2 reports a summary of model selection for comparisons between boys and girls for the 9th grade, and Table 3 for the 10th grade. Inspection of both tables reveals that Model E yields the lowest BIC statistics in all cases examined, except Germany (10th grade). Model E fixes the school-level factor loadings (λ_c) and the effects of the family and school characteristics (β_k and γ_p) across genders, while continuing to allow the between-school intercepts (λ_{c0}) for academic achievement to vary across genders. In other words, Model E says that boys and girls have different mean levels of achievement, and that their academic performance is influenced by the observed background and school factors in a similar way. The better fit of Model E also highlights the importance of unobserved differences between schools (that are not based on the family and school characteristics described in the data) in accounting for variance in academic achievement by gender, whereas the better fit of the "gender differences" model for German 10th graders suggests that the observed school and family characteristics, along with the unobservables, all contribute to the explanatory power of that model. The fact that Model E has the best fit with the data on German 9th graders but not 10th graders suggests that the role of observables declines across those two grades, though we do not have an explanation for this.

Based on the above results, we report estimates of Model E for factor loadings, family effects, and school effects for all cases except for German 10th graders, for whom we report results of Model A3.

(Table 2 about here)

(Table 3 about here)

Model Estimates

Because Model E assumed that factor loadings of the three subjects on latent achievement does not vary between boys and girls, the degree to which family and school factors impact each of the three subjects follows the same pattern by gender. That is, in most countries, math performance (for both boys and girls) is more strongly impacted by family and school characteristics than is science and reading performance, though the precise order and degree of impact varies by country. In Taiwan, for example, family and school factors impact math performance substantially more than for reading and science, at both the 9th and 10th grade, compared to the impact of those factors in other countries. This might be due to the cultural importance of mathematics in Taiwanese families and schools, or possibly due to the importance of mathematics in Taiwanese exams for entrance to upper secondary and tertiary education. Note that the pattern with which family and school factors impact academic achievement varies by gender in Germany (this is precisely what A3 allows, but Model E does not), where those factors impact math performance more than science performance in the case of girls, and vice versa in the case of boys. We do not have a cultural or other explanation for this result, nor does the literature suggest reasons for such differing effects across subjects, genders, and grades.

(Table 4 about here)

Tables 5-1 and 5-2 report family and school effects for the 9^{th} and 10^{th} grades, respectively. Under Model E, vocational schools are associated with very poor academic achievement for both boys and girls in every country (except for girls in Japanese vocational schools); vocational schools are particularly detrimental for academic achievement for 10^{th} grade pupils in Taiwan and the Czech Republic. Rural schools also significantly underperform urban schools in Taiwan for both boys and girls, as well as for Korean boys. We also note that Taiwanese private schools outperform in the 9^{th} grade but underperform in the 10^{th} grade, which is due to differences in the nature and purpose of private schools at the lower secondary and upper secondary levels. Overall, we can conclude that school effects are generally similar for boys and girls in each country.

In terms of family effects, in both 9^{th} and 10^{th} grades in all countries, books in the home is the dominant factor of the background family predicting student achievement. In the case of Taiwan, books in the home have over twice the effect on 9^{th} grade girls' academic achievement than on boy's achievement, while books have nearly twice the importance for Korean boys as for Korean girls in the 10^{th} grade. Besides those exceptions, books in the home have a consistently strong and similar effect for both boys and girls. We also note that educational resources in the home seem to be relevant for student achievement in the East Asian countries, but not in Western ones, possibly due to the wide availability of such resources in Western homes. Other family effects are generally weak, except for the importance of parental occupational status in Taiwan and the United States. As with school effects, the impact of family effects on student achievement seem to be similar for both boys and girls. The similarity in the importance of family and school effects on boys' and girls' academic achievement suggests that those effects do not broadly account for gender differences in academic achievement in any of the six countries.

(Tables 5-2 and 5-2 about here)

Table 6 reports the R^2 coefficients under Model E (but Model A3 for German 10th graders) for the six countries examined in terms of both the within-school and between-school explained variance. Most importantly, the table reveals a significant increase from the 9th to 10th grades in the explained variance of latent student achievement accounted for by between-school factors, reflecting the changes in school systems between lower and upper secondary levels in Taiwan as well as the Czech Republic. By contrast, the change in between-school explained variance is much smaller in Germany between 9th and 10th grades, precisely because the presence of a high degree of school differentiation already in the 9th grade. Also note that the between-school explained variances in Korea in the 10th grade are smaller than for the three previously mentioned countries, and are smaller still in Japan and the United States. These patterns seem to reflect the smaller degree of differentiation in school types in the latter set of countries, particularly in the United States. These results also bring into question the commonly held belief that school-level SES is strongly related to student achievement in that country. Lastly, it is worth noting that, because Model E constrains the structural coefficients to be equal for boys and girls within each country, the differences in explained variance by gender in Table 6 are entirely attributable to gender differences in variances.

(Table 6 about here)

Adjusted Gender Gaps

A note has to be made here. As mentioned earlier, we first run the model depicted in Figure 1 separately for boys and girls (Model A1 and A2), and then adapt the model to an equivalent one with a two-group comparative design (Model A3). The term "equivalent" is used because in two-group two-level MIMIC model, the between-level intercept (λ_{c0}) for the index subject (math) is automatically set by Mplus to be equal between the two groups (even when it is specified to be different), and the between-group contrast in this intercept is captured by a new parameter automatically yielded, that is, the intercept for the latent variable at the between level. This intercept can represent the gender gap in math estimated by the model. We can change the index subject to science or reading and thus obtain gender gaps in these subjects. These gender gaps are conceptually similar to net gender differences that can be obtained from regression models, except that our estimates ae based on a MIMIC model predicting latent academic achievement for all three subjects, and adjusted by parameters at the individual and school levels and by corresponding error variances.

Table 7 reports the magnitude of this type of gender gap in three subjects estimated by Model E for all cases, along with the "raw" gender mean gap, as observed in the analysis sample (also see Appendix A-3). We do not report the estimates for Model A3 for German 10th graders, because intercept differences in the presence of gender interaction are variable, depending on where they are located. By contrast, in Model E there are no interactions, so the intercepts are invariant to the place in the distributions where they are evaluated.

(Table 7 about here)

Table 7 yields a number of important empirical findings, particularly in light of our previous results. First, it should be noted that the magnitude of the intercepts should be interpreted as the gender gap in the PISA score "adjusted" for the role of family and school effects. These intercepts can be compared to the "unadjusted" mean gender differences. In the case of Taiwan, we can observe that after accounting for model parameters, the gender gap in mathematics, science and reading in the 9th grade remain similar to the unadjusted gender gap, according to which the gender gaps in math and science are insignificant while the gender gap in reading remains large, favoring girls. By contrast, it should be noted that the adjusted gender gap in the 9th grade for the Czech Republic and Germany widens in the direction favoring boys, compared to the unadjusted gender gap. For both countries, the gender gap in mathematics widens, and narrows in the case of reading. The Czech Republic also has a statistically significant gender gap in science favoring boys.

The same patterns hold for those three countries also in the 10th grade: the Taiwanese adjusted gender gap is generally the same as the unadjusted gap, while in the case of the Czech Republic and Germany the adjusted gender gap widens in the direction favoring boys for all three subjects. 10th grade science achievement in Germany also exhibits a gender gap favoring boys. As for Japan, Korea, and the United States, the results are quite different. The gender gap in Japan remains similar in the case of reading (favoring girls), while the gender gap in math and science performance, believed to favor boys, actually disappears. The gender gap in math and science in Korea likewise disappears once model parameters are taken into account, while the gender gap in reading widens, further favoring girls. In the United States, gender gaps in math and reading remain similar to the unadjusted gaps (favoring boys and girls, respectively), while there is no longer a gender gap in science.

From these results we can conclude that there is no gender gap in math and science achievement in any of the East Asian countries, while in all three countries a large gender gap favoring girls persists in reading performance. In the Western countries, gender gaps are generally large and subject specific. We also note that, in the countries with more differentiated school systems (Germany and the Czech Republic), the adjusted measures in all subjects shift in the direction favoring boys, confirming our hypothesis, such that boys outperform girls in both math and science, while girls continue to outperform in reading, though to a lesser degree. Above all, we cannot infer a systematic underachievement of either boys or girls across the six industrialized countries examined here.

Discussion and Conclusion

We can draw a number of conclusions from our findings. First, we find gender gaps in reading (favoring girls) in all countries examined, even after taking into account a large range of factors relating to family background and schooling conditions. This leaves open the possibility that biological factors, or some other common factor unobserved in our analysis, may play a role in gender gaps in reading achievement. Biological factors could mean any form of unobservable individual heterogeneity rooted in sex differences, such as differences in brain development or neurological maturity, that systematically advantage school-age girls in reading comprehension and other linguistic tasks, but not in mathematical or scientific knowledge. However, the gender gap in reading could also be due to other unobservables, such as the role of girls' social skills or educational expectations from parents and schools, that are common in all six countries and provide a social-psychological environment conducive to reading ability. From our data, it is impossible to discern which unobserved factors are at work.

Second, and relatedly, we also found that students' reading ability is influenced by family background to a lesser extent than students' math (or science) ability in all countries examined (see Table 4), for both girls and boys. This suggests that the gender gap in mathematics and science, often favoring boys in the Western countries, is largely due to environmental or structural factors – either observed in our data or unobserved, while those factors play a weaker role in the persistence of the gender gap in reading.

Third, we explored the heterogeneity in effects of family background and school factors across genders, because some have suggested that girls' academic achievements are influenced by family socioeconomic status to a larger extent than are boys' (e.g., Tansel 2002), and that boys from disadvantaged family backgrounds might have more behavioral problems in school and lower performance than girls from similar backgrounds. Our findings, however, indicate gender homogeneity in effects of family background and school factors on latent academic achievement in all cases except German 10th graders. That is, families and schools do not seem to provide educational resources or opportunities to boys and girls in ways that would systematically lead to strong effects of those conditions on overall academic performance by gender. Rather, the gender homogeneity of family and school effects on latent academic achievement could reflect the modernization or industrialization of the six countries examined, because traditional obstacles to girls' educational advancement have long been removed (Shavit, Arum, and Gamoran 2007).

Fourth, the intercepts in Table 7, especially when compared to the raw gender gaps in the analytic sample reported in the same table, indicate that family and school factors, along with unobservables, do not increase or decrease gender gaps in academic performance in a systematic way across countries. These factors do not, for example, systematically reduce gender gaps in a particular subject, thus contributing to underachievement of boys or girls in that subject. The explanation for such changes likely depends on structural factors in national school systems, particularly the way boys and girls are sorted into different types of school, into schools of different vocational orientations, or by other conditions unobserved in our data. When those differences in sorting (or via other, unobserved mechanisms) are taken into account, the gender gaps estimated by the models are smaller than compared to the OECD reported gender gaps, particularly for the countries with the most intensive sorting.

Lastly, our results highlight the advantages of using the MIMIC model approach in examining inequality in academic performance across groups. By applying a two-level MIMIC model of family background and school effects on latent academic performance across six nations participating in the PISA survey, we confirmed that there are country-specific and subject-specific patterns in the relative importance of these effects – patterns that would be masked had we run the analysis separately for each subject. That is, the MIMIC model enables us to confirm that family and school effects matter for some subjects substantively more than others – i.e. that there are substantive differences in performance inequality depending on the subject examined, and thus that performance inequality in one subject cannot be taken as a proxy for overall inequalities without an analysis that actually warrants such an assumption. Compared to the common approach in which scholars run separate regressions for each subject, the MIMIC model approach enables testing equality constraints imposed on the factor loadings for each subject, thereby establishing whether those loadings are statistically different from each other, which is certainly more robust than the common practice of eyeballing coefficients across different models.

This study has a number of limitations. First, the analysis used a few observable explanatory variables that are mostly comparable across countries. It would be advantageous to examine more variables, but in such a case we might have to abandon the comparative design of this paper. On the other hand, we restricted the analytic sample to students with complete information on observables (as is common practice in studies of this kind), and the inclusion of more observables would augment the missing data problem we already have. Third, this analysis obviously focused on performance in international PISA tests, and thus we did not examine other important dimensions of gender gaps in academic achievement, such as in teachers' grades, in educational and occupational expectations, or in the attainment or completion of secondary or tertiary education. This caveat is important in light of the fact that the sociological literature on gender gaps in education often focuses on those other dimensions of achievement, where boys' relative underachievement is more apparent. For example, in the United States, there has been a high degree of stability in the gender gap in academic performance over many decades, with girls outperforming boys in course grades (DiPrete and Buchmann 2013). Similarly, at the same level of PISA performance in mathematics, Czech boys get substantially worse math grades than girls (Mateju and Smith 2015). Overall, many studies have corroborated Buchmann and DiPrete's (2006) prediction that women will make educational choices similar to those of men, conditional on gender parity in educational resources and incentives. Because of girls' higher educational expectations, it is possible that women as a group will catch up to men or even outpace men in many dimensions of educational achievement.

References

Andrew, Megan and Robert M. Hauser. 2012. "Adoption? Adaptation? Evaluating the Formation of Educational Expectations." *Social Forces* 90 (2): 497-520.

Baker, David P. and Deborah Perkins Jones. 1993. "Creating Gender Equality: Cross-National Gender Stratification and Mathematical Performance." *Sociology of Education* 66 (2): 91-103.

Bandura, Albert. 1986. "The Explanatory and Predictive Scope of Self-Efficacy Theory." J*ournal of Social and Clinical Psychology* 4: 359-373.

Bedard, Kelly and Insook Cho. 2010. "Early Gender Test Score Gaps across OECD Countries." *Economics of Education Review* 29 (3): 348-363.

Buchmann, Claudia and Thomas A. DiPrete. 2006. "The Growing Female Advantage in College Completion: The Role of Parental Resources and Academic Achievement." *American Sociological Review* 71 (4): 515-541.

Ceei, Stephen J., Wendy M. Williams, and Susan M. Barnett. 2009. "Women's Underrepresentation in Science: Sociocultural and Biological Considerations." *Psychological Bulletin* 135 (2): 218-261.

Correll, Shelley J. 2004. "Constraints into Preferences: Gender, Status, and Emerging Career Aspirations." *American Sociological Review* 69 (1): 93-113.

Dickerson, Andy, Steven McIntosh and Christine Valente. 2015. "Do the Math: An Analysis of the Gender Gap in Mathematics in Africa." *Economics of Education Review* 46: 1-22.

DiPrete, Thomas A. and Claudia Buchmann. 2013. *The Rise of Women: The Growing Gender Gap in Education and What It Means for American Schools.* New York: Russell Sage.

Eccles, Jacquelynne S. 2011a. "Understanding Women's Achievement Choices: Looking Back and Looking Forward." *Psychology of Women Quarterly* 35 (3): 510-516.

Eccles, Jacquelynne S. 2011b. "Gendered Educational and Occupational Choices: Applying the Eccles et al. Model of Achievement-Related Choices." *International Journal of Behavioral Development* 35 (3): 195-201.

Else-Quest, Nicole M., Janet S. Hyde and Marcia C. Linn. 2010. "Cross-National Patterns of Gender Differences in Mathematics: A Meta-Analysis." *Psychological Bulletin* 136 (1): 103-127.

England, Paula. 2016. "Sometimes the Social Becomes Personal: Gender, Class, and Sexualities." *American Sociology Review* 81 (1): 4-28.

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

Guiso, Luigi, Ferdinando Mounte, Paola Sapienza and Luigi Zingales. 2008. "Culture, Gender, and Math." *Science* 320 (5880): 1164-1165.

Hauser, Robert M. 1972. "Disaggregating a Social-Psychological Model of Educational Attainment." *Social Science Research* 1: 159-188.

Hauser, Robert M. and Arthur S. Goldberger. 1971. "The Treatment of Unobservable Variables in Path Analysis." *Sociological Methodology* 3: 81-117.

Huang, Min-Hsiung. 2009. "Classroom Homogeneity and the Distribution of Student Math Performance: A Country-level Fixed-effects Analysis." *Social Science Research* 38 (4): 781-91.

Hyde, Janet S. 2005. "The Gender Similarities Hypothesis." *American Psychologist* 60 (6): 581-592.

Hyde, Janet S., Sara Lindberg, Marcia Linn, Amy Ellis and Caroline Williams. 2008. "Gender Similarities Characterize Math Performance." *Science* 321 (5888): 494-495.

Hyde, Janet S. and Marcia C. Linn. 2006. "Gender Similarities in Mathematics and Science." *Science* 314 (5799): 599-600.

Legewie, Joscha and Thomas A. DiPrete. 2012. "School Context and the Gender Gap in

Educational Achievement." *American Sociological Review* 77 (3): 463-485.

Legewie, Joscha and Thomas A. DiPrete. 2014. "Pathway to Science and Engineering Bachelor's Degrees for Men and Women." *Sociological Science* 1: 41-48.

Lindberg, Sara M., Janet S. Hyde, Jennifer L. Petersen and Maecia C. Linn. 2010. "New Trends in Gender and Mathematics Performance: A Meta-Analysis." *Psychological Bulletin* 136 (6): 1123-1135.

Lovaglia, Michael, Jeffrey Lucas, Jeffrey Houser, Shane Thye and Barry Markovsky. 1998. "Status Processes and Mental Ability Test Scores." *American Journal of Sociology* 104 (1): 195-228.

Machin, Stephen and Tomas Pekkarinen. 2008. "Global Sex Differences in Test Score Variability." *Science* 322 (5906): 1331-1332.

Marks, Gary N. 2008 "Accounting for the Gender Gaps in Student Performance in Reading and Mathematics: Evidence from 31 Countries." *Oxford Review of Education* 34 (1): 89-109.

Mateju, Petr and Michael L. Smith. 2015. "Are Boys that Bad? Gender Gaps in Measured Skills, Grades and Aspirations in Czech elementary schools." *British Journal of the Sociology of Education* 36 (6): 871-895.

Montt, Guillermo. 2011. "Cross-national Differences in Educational Achievement Inequality." *Sociology of Education* 84 (1): 49-68.

Morgan, Stephen L., Dafna Gelbgister, and Kim A. Weeden. 2013. "Feeding the Pipeline: Gender, Occupational Plans, and College Major Selection." *Social Science Research* 42 (4): 989-1005.

OECD (Organization for Economic Cooperation and Development). 2015. *The ABC of Gender Equality in Education: Aptitude, Behavior, Confidence*. Paris: OECD.

OECD (Organization for Economic Cooperation and Development). 2012. *Education at a*

Glance. Paris: OECD.

Park, Hyunjoon. 2010. ‘‘Japanese and Korean High Schools and Students in Comparative Perspective.’’ Pp. 255–73 in *Quality and Inequality of Education: Cross National Perspectives*, edited by Jopp Dronkers. New York: Springer Press.

Park, Hyunjoon, and Gary D. Sandefur. 2006. ‘Families, Schools, and Reading in Asia and Latin America.’’ *Research in Sociology of Education* 15:133–62.

Penner, Andrew M. 2008. “Gender Differences in Extreme Mathematical Achievement: An International Perspective on Biological and Social Factors.” *American Journal of Sociology* 114 (S1): S138-S170.

Riegle-Crumb, Catherine. 2005. “The Cross-National Context of the Gender Gap in Math and Science.” Pp. 227-243 in *The Social Organization of Schooling*, edited by Larry V. Hedges and Barbara Schneider. New York: Russell Sage Foundation.

Shavit, Yossi, Richard Arum and Adam Gamoran. 2007. *Stratification in Higher Education: A Comparative Study*. Stanford: Stanford University Press.

Stoet, Gijsbert and David C. Geary. 2013. “Sex Differences in Mathematics and Reading Achievement Are Inversely Related: Within- and Across-Nation Assessment of 10 Years of PISA Data.” *PLoS ONE* 8 (3): e57988.

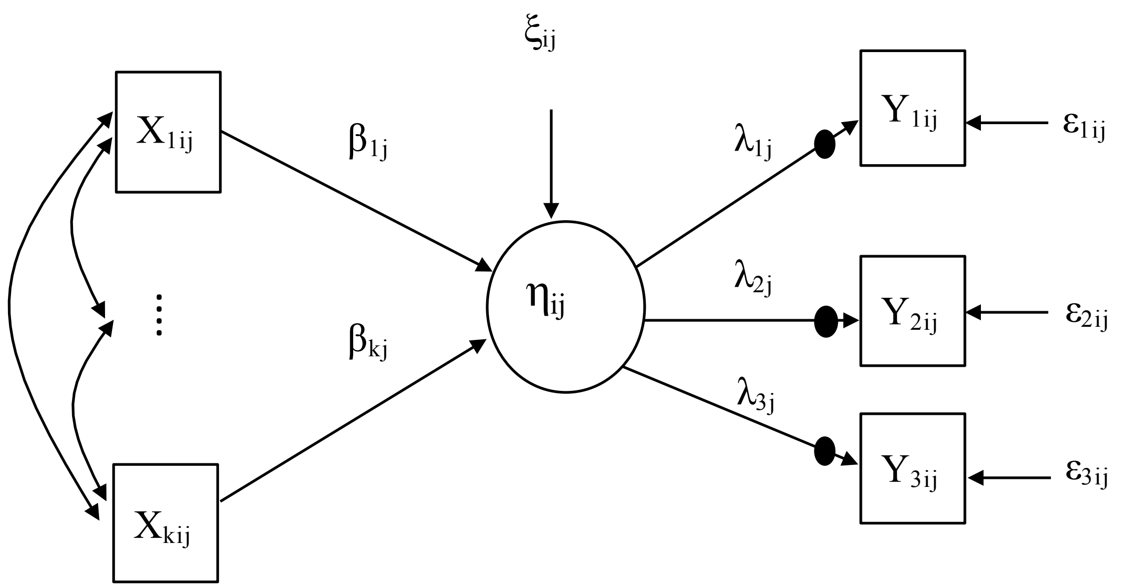
Tansel, Aysit. 2002. “Determinants of School attainment of Boys and Girls in Turkey: Individual, Household and Community Factors.” *Economics of Education Review* 21 (5): 455-470.

Tsai, Shu-Ling, Michael L. Smith, and Robert M. Hauser. 2017. “Families, Schools, and Student Achievement Inequality: A Multilevel MIMIC Model Approach.” *Sociology of Education* 90 (1): 64-88.

Xie, Yu and Kimberlee A. Shauman. 2003. *Women in Science: Career Processes and Outcomes*. Cambridge: Harvard University Press.

Figure 1. Diagrams for a Two-Level MIMIC Model of Student Academic Achievement

Within-School



Between-School

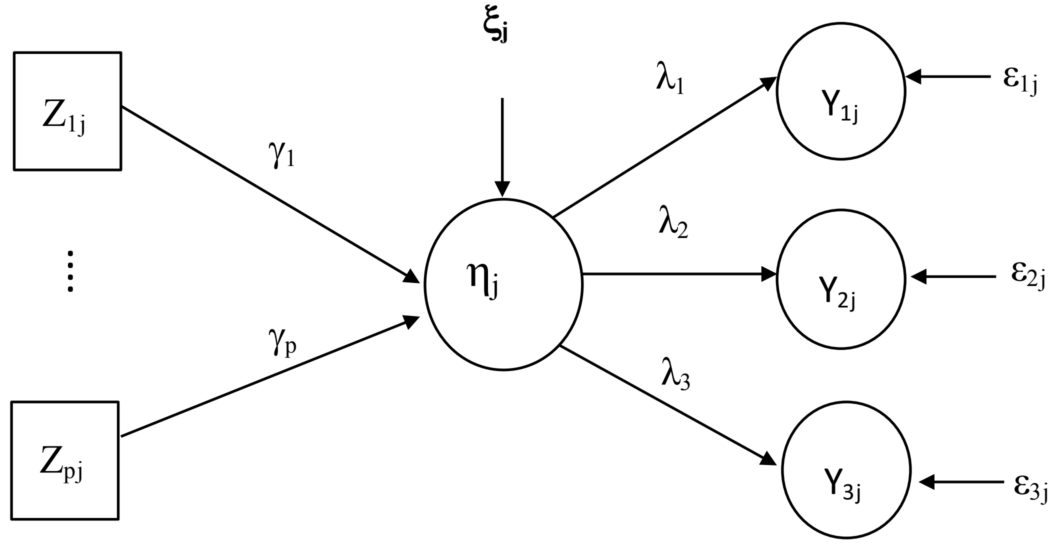
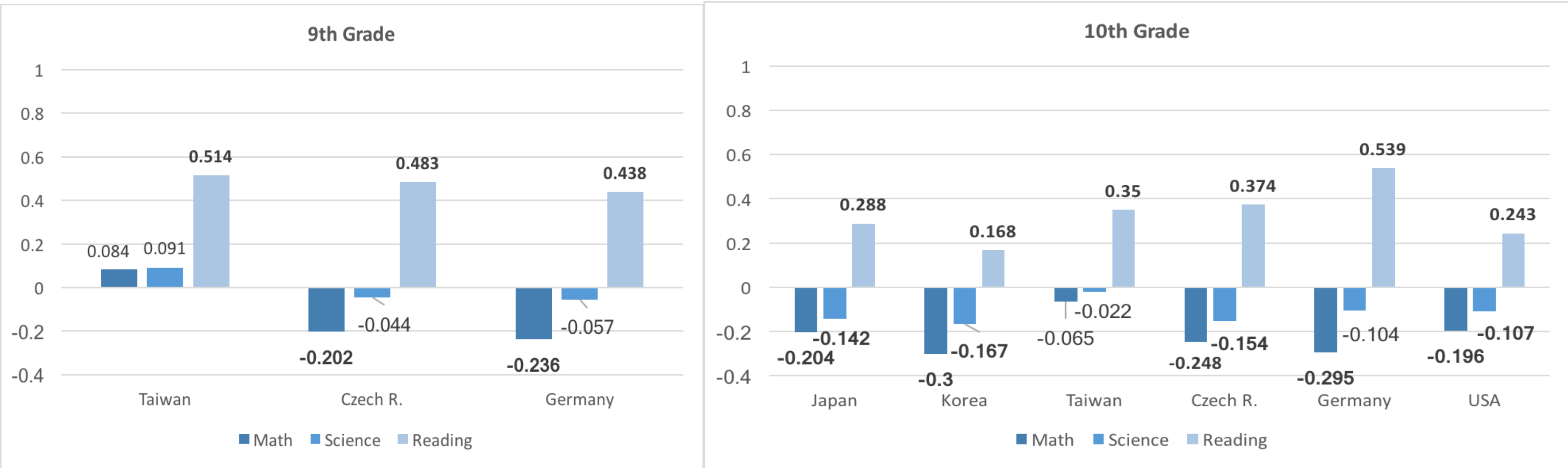


Figure 2. Standardized Mean Gender Gap (*d*-statistics) by Subject, Grade, and Country: Analytic Sample



Note: Coefficients in bold are significant at the p < .001 level.

Table 1. Total Variance and Share of Between-school Variance among Male and Female Students, by Subject, Grade, and Country: Analytic Sample

		9th Grade		10th Grade	
Country	Subject	Girls	Boys	Girls	Boys
Total Student Variance					
Japan	Math			3153.2	3937.7
	Science			3870.4	4693.6
	Reading			3582.7	4807.5
Korea	Math			4628.7	5517.9
	Science			3267.0	3737.1
	Reading			3357.8	4390.2
Taiwan	Math	8820.7	10165.6	4927.8	5163.2
	Science	4400.3	5138.4	2398.3	2510.9
	Reading	4838.6	6263.9	2923.9	3013.0
Czech Republic	Math	4272.7	4739.8	3045.7	2822.0
	Science	3779.6	4268.5	2834.0	2794.9
	Reading	3258.0	3515.3	2356.3	2554.5
Germany	Math	3088.2	3200.1	3069.1	3494.5
	Science	3276.4	3328.0	2926.1	3376.2
	Reading	2561.0	2551.5	2277.0	2973.6
USA	Math			4688.1	5328.3
	Science			5108.2	5908.1
	Reading			4610.5	5313.5
Between-school Variance (%)					
Japan	Math			56.2	56.1
	Science			46.4	47.6
	Reading			50.2	51.3
Korea	Math			41.6	42.6
	Science			41.3	40.6
	Reading			37.1	38.3
Taiwan	Math	3.2	13.9	52.8	59.0
	Science	2.5	10.4	55.6	61.2
	Reading	5.6	10.5	51.8	58.6
Czech Republic	Math	28.2	30.5	57.4	61.0
	Science	25.4	25.3	51.7	52.9
	Reading	27.1	34.6	61.5	59.5
Germany	Math	54.5	57.1	41.2	40.6
	Science	52.5	55.7	38.6	41.8
	Reading	57.4	59.8	42.1	50.2
USA	Math			12.5	10.7
	Science			12.2	11.8
	Reading			13.9	18.3

Table 2. Fit Statistics for a Series of Models with Gender Comparison, by Country: 9th Grade

Model and Contrast	Taiwan	Czech Republic	Germany
I. Model			
A. Gender differences model			
1. Girls' data			
Adjusted BIC	46942.478	50113.903	38911.531
(# parameters) [df]	(40) [16]	(41) [18]	(42) [20]
2. Boys' data			
Adjusted BIC	46307.525	59477.069	40661.458
(# parameters) [df]	(40) [16]	(41) [18]	(42) [20]
3. Pooled data			
Adjusted BIC	93305.376	109647.973	79631.111
(# parameters) [df]	(80) [32]	(82) [36]	(84) [40]
B = A3 + λ_c invariant			
Adjusted BIC	93309.809	109640.862	79623.967
(# parameters) [df]	(78) [34]	(80) [38]	(82) [42]
C = A3 + λ_c , β_k invariant			
Adjusted BIC	93309.749	109631.269	79605.798
(# parameters) [df]	(73) [39]	(75) [43]	(77) [47]
D = A3 + λ_c , γ_p invariant			
Adjusted BIC	93305.150	109629.920	79608.722
(# parameters) [df]	(76) [36]	(77) [41]	(78) [46]
E = A3 + λ_c , β_k , γ_p invariant			
Adjusted BIC	93303.271$^{@}$	109620.388$^{@}$	79590.383$^{@}$
(# parameters) [df]	(71) [41]	(72) [46]	(73) [51]
F. Gender similarities model = E + λ_{c0} invariant			
Adjusted BIC	93435.380	109820.904	79910.442
(# parameters) [df]	(69) [43]	(70) [48]	(71) [53]
II. Contrast in BIC			
Model B vs. Model A3	4.433	-7.111	-7.144
Model C vs. Model A3	4.373	-16.704	-25.313
Model D vs. Model A3	-0.226	-18.053	-22.389
Model E vs. Model A3	-2.105	-27.585	-40.728
Model F vs. Model A3	130.004	172.931	279.331

Note: $^{@}$ Best-fitting model

Table 3. Fit Statistics for a Series of Models with Gender Comparison, by Country: 10th Grade

Model and Contrast	Japan	Korea	Taiwan	Czech Republic	Germany	USA
I. Model						
A. Gender differences model						
1. Girls' data						
Adjusted BIC	134777.881	105891.154	98939.676	55394.344	33110.692	83042.055
(# parameters) [df]	(41) [18]	(41) [18]	(42) [20]	(42) [20]	(42) [20]	(40) [16]
2. Boys' data						
Adjusted BIC	142477.381	123572.845	88171.109	42844.541	26950.653	80801.845
(# parameters) [df]	(41) [18]	(41) [18]	(42) [20]	(42) [20]	(42) [20]	(40) [16]
3. Pooled data						
Adjusted BIC	277312.081	229521.009	187169.125	98297.782	60119.922@	163899.314
(# parameters) [df]	(82) [36]	(82) [36]	(84) [40]	(84) [40]	(84) [40]	(80) [32]
B = A3 + λ_c invariant						
Adjusted BIC	277321.050	229536.991	187159.325	98296.450	60136.687	163891.501
(# parameters) [df]	(80) [38]	(80) [38]	(82) [42]	(82) [42]	(82) [42]	(78) [34]
C = A3 + λ_c , β_k invariant						
Adjusted BIC	277301.971	229523.257	187141.355	98280.473	60138.007	163870.963
(# parameters) [df]	(75) [43]	(75) [43]	(77) [47]	(77) [47]	(77) [47]	(73) [39]
D = A3 + λ_c , γ_p invariant						
Adjusted BIC	277308.019	229523.880	187145.387	98283.838	60124.664	163884.002
(# parameters) [df]	(77) [41]	(77) [41]	(78) [46]	(78) [46]	(78) [46]	(76) [36]
E = A3 + λ_c , β_k , γ_p invariant						
Adjusted BIC	277288.586@	229510.477@	187126.422@	98267.867@	60127.412	163864.144@
(# parameters) [df]	(72) [46]	(72) [46]	(73) [51]	(73) [51]	(73) [51]	(71) [41]
F. Gender similarities model						
= E + λ_{c0} invariant						
Adjusted BIC	277546.305	229678.734	187296.857	98427.436	60423.001	164005.191
(# parameters) [df]	(70) [48]	(70) [48]	(71) [53]	(71) [53]	(71) [53]	(69) [43]

Table 3. Continued

Model and Contrast	Japan	Korea	Taiwan	Czech Republic	Germany	USA
II. Contrast in BIC						
Model B vs. Model A3	8.969	15.982	-9.8	-1.332	16.765	-7.813
Model C vs. Model A3	-10.11	2.248	-27.77	-17.309	18.085	-28.351
Model D vs. Model A3	-4.062	2.871	-23.738	-13.944	4.742	-15.312
Model E vs. Model A3	-23.495	-10.532	-42.703	-29.915	7.49	-35.17
Model F vs. Model A3	234.224	157.725	127.732	129.654	303.079	105.877

Note: [@] Best-fitting model

Table 4. Estimates of Factor Loading (λ_c), by Grade and Country

	9th Grade		10th Grade	
Country	Girls	Boys	Girls	Boys
Japan			Science ≈ Reading > Math (1.100) (1.093) (1.000)	Science ≈ Reading > Math (1.100) (1.093) (1.000)
Korea			Math > Reading > Science (1.000) (0.869) (0.831)	Math > Reading > Science (1.000) (0.869) (0.831)
Taiwan	Math > Reading > Science (1.000) (0.765) (0.725)	Math > Reading > Science (1.000) (0.765) (0.725)	Math > Reading > Science (1.000) (0.771) (0.726)	Math > Reading > Science (1.000) (0.771) (0.726)
Czech R.	Math > Science > Reading (1.000) (0.951) (0.862)	Math > Science > Reading (1.000) (0.951) (0.862)	Math > Science > Reading (1.000) (0.971) (0.899)	Math > Science > Reading (1.000) (0.971) (0.899)
Germany	Science > Math > Reading (1.025) (1.000) (0.911)	Science > Math > Reading (1.025) (1.000) (0.911)	Math > Science > Reading (1.000) (0.944) (0.842)	Science > Math > Reading (1.013) (1.000) (0.983)
USA			Science > Reading ≈ Math (1.074) (1.004) (1.000)	Science > Reading ≈ Math (1.074) (1.004) (1.000)

Note: Estimates are provided for the best-fitting model (Model E for all cases except for German 10th graders, where estimates of Model A3 are provided). Numbers in parentheses are metric coefficients estimated by the model. The > sign indicates statistically significant differences at the $p < .05$ level, while ≈ indicates insignificant differences.

Table 5-1. Pooled Estimates of Family and School Effects for Girls and Boys, by Country: 9th Grade (Model E)

Variables	Taiwan	Czech Republic	Germany
Family Effects			
Father's schooling (β_k)	1.812	0.618	-0.359
	(1.435)	(0.520)	(-0.505)
Mother's schooling	0.654	-1.329	0.416
	(0.590)	(-1.044)	(0.545)
Parents' highest ISEI	0.455^{**}	0.530^{***}	0.270^{*}
	(2.726)	(4.601)	(2.550)
Number of books in parents' home	18.229^{***}	15.730^{***}	11.657^{***}
	(8.704)	(9.534)	(7.827)
Educational resources in home	11.483^{***}	4.788^{*}	-1.248
	(6.561)	(2.338)	(-0.540)
School Effects (γ_p)			
School Sector	43.915^{***}	4.775	-1.805
Private (ref: Public)	(5.403)	(0.310)	(-0.184)
School Location	-18.263^{*}	-1.409	6.059
Rural (ref: else)	(-2.504)	(-0.208)	(0.842)
School Type (ref: Academic)			
Vocational schools	—	-77.964^{***}	-93.786^{***}
		(-11.483)	(-15.714)
Comprehensive schools	—	—	-156.837^{***}
			(-13.372)

Note: Estimates are provided for the best-fitting model (Model E for all cases except for German 10th graders, where estimates of Model A3 are provided). Numbers in parentheses are the values of the ratio of the metric coefficient to its standard error. $^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

Table 5-2. Estimates of Family and School Effects by Country: 10^{th} Grade

Variables	Japan	Korea	Taiwan	Czech Rep.	Germany		USA
	Girls and Boys	Girls and Boys	Girls and Boys	Girls and Boys	Girls	Boys	Girls and Boys
Family Effects							
Father's schooling (β_k)	-0.270	0.465	1.205	-0.629	0.895	-0.414	0.972
	(-0.449)	(0.652)	(1.719)	(-0.860)	(1.141)	(-0.346)	(1.331)
Mother's schooling	0.355	-0.173	0.195	-0.032	-0.620	-3.324^{***}	-0.313
	(0.516)	(-0.193)	(0.271)	(-0.038)	(-0.710)	(-3.247)	(-0.361)
Parents' highest ISEI	0.051	0.095	0.349^{***}	0.124	-0.018	0.440^{**}	0.674^{***}
	(0.859)	(1.138)	(4.481)	(1.489)	(-0.116)	(2.716)	(7.247)
Number of books in parents' home	7.432^{***}	9.119^{***}	7.215^{***}	12.812^{***}	15.836^{***}	11.574^{***}	14.052^{***}
	(8.103)	(6.989)	(6.712)	(10.327)	(7.600)	(4.420)	(10.707)
Educational resources in home	2.622^{**}	7.541^{***}	5.816^{***}	0.970	-7.314^{*}	5.031	0.691
	(2.584)	(4.394)	(5.204)	(0.569)	(-2.557)	(1.361)	(0.526)
School Effects (γ_p)							
School Sector	-2.578	6.978	-67.509^{***}	-17.006^{*}	1.928	9.045	2.597
Private (ref: Public)	(-0.245)	(0.637)	(-8.999)	(-2.468)	(0.239)	(0.894)	(0.319)
School Location	-19.859^{*}	-26.973^{*}	-30.648^{***}	-4.703	-2.013	14.807	5.263
Rural (ref: else)	(-2.005)	(-2.150)	(-4.513)	(-0.959)	(-0.350)	(1.141)	(0.852)
School Type (ref: Academic)							
Vocational schools	-22.387^{**}	-63.741^{***}	-113.361^{***}	-154.405^{***}	-82.437^{***}	-85.165^{***}	—
	(-2.601)	(-6.511)	(-11.327)	(-25.364)	(-11.799)	(-11.826)	
Comprehensive schools	—	—	-64.163^{***}	—	-102.430^{***}	-101.221^{***}	—
			(-8.906)		(-6.422)	(-4.024)	
Technical schools with school-leaving exam	—	—	—	-77.429^{***} (-13.799)	—	—	—

Note: Pooled estimates from Model E are provided for all countries except for Germany, where coefficients from Model A3 are provided separately for boys and girls, which the results in Table 3 found to be statistically different from each other. Numbers in parentheses are the values of the ratio of the metric coefficient to its standard error. $^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

Table 6. Explained Variance (R^2) for Boys and Girls, by Grade and Country

	Japan		Korea		Taiwan		Czech Republic		Germany		USA	
	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys
9th Grade												
Within Level												
Math					0.911	0.940	0.899	0.872	0.862	0.883		
Science					0.976	0.980	0.921	0.930	0.918	0.914		
Reading					0.925	0.943	0.891	0.879	0.875	0.870		
Latent Factor					0.226	0.167	0.176	0.185	0.116	0.128		
Between Level												
Math					0.963	0.896	0.869	0.889	0.970	0.974		
Science					0.750	0.897	0.811	0.801	0.940	0.961		
Reading					0.834	0.828	0.720	0.796	0.910	0.937		
Latent Factor					0.575	0.321	0.500	0.388	0.721	0.656		
10th Grade												
Within Level												
Math	0.848	0.871	0.897	0.881	0.871	0.877	0.844	0.857	0.901	0.863	0.921	0.926
Science	0.904	0.919	0.882	0.912	0.956	0.952	0.906	0.888	0.882	0.921	0.969	0.976
Reading	0.898	0.904	0.902	0.896	0.889	0.905	0.832	0.810	0.878	0.900	0.934	0.937
Latent Factor	0.048	0.040	0.088	0.077	0.111	0.101	0.118	0.129	0.156	0.149	0.185	0.164
Between Level												
Math	0.928	0.947	0.976	0.948	0.995	0.991	0.966	0.959	0.974	0.924	0.839	0.948
Science	0.971	0.965	0.952	0.957	0.972	0.973	0.919	0.947	0.897	0.926	0.780	0.826
Reading	0.960	0.959	0.948	0.937	0.954	0.958	0.914	0.947	0.841	0.900	0.582	0.597
Latent Factor	0.078	0.059	0.269	0.264	0.789	0.820	0.838	0.896	0.799	0.787	0.010	0.009

Note: Estimates are provided for the best-fitting model (Model E for all cases except for German 10th graders, where estimates of Model A3 are provided).

Table 7. Adjusted and Unadjusted Gender Gaps, by Subject, Grade and Country: Analytic Sample

Grade/ Country	Subject	Gender Disparities in the Between-level Intercept (λ_{c0}) Model E	Gender Mean Gap
9th Grade			
Taiwan	Math	7.589	8.732
	Science	5.058	6.589
	Reading	39.284***	40.694***
Czech R.	Math	-25.176***	-16.562***
	Science	-13.143*	-3.321
	Reading	23.440***	34.777***
Germany	Math	-25.968***	-20.862***
	Science	-9.667	-5.046
	Reading	31.585***	36.138***
10th Grade			
Japan	Math	-12.744	-18.374***
	Science	-7.839	-12.825***
	Reading	30.270**	26.722***
Korea	Math	-18.428	-29.254***
	Science	-6.099	-13.227***
	Reading	22.956**	13.812***
Taiwan	Math	-6.797	-7.404
	Science	-2.066	-1.846
	Reading	30.728***	30.003***
Czech R.	Math	-33.934***	-21.428***
	Science	-23.824***	-12.285***
	Reading	14.441**	29.656***
Germany	Math	-34.084***	-23.595***
	Science	-18.334**	-7.980
	Reading	32.416***	39.611***
USA	Math	-17.610**	-15.129***
	Science	-10.697	-8.594**
	Reading	16.938*	18.994***

Note. *p < .05 **p < .01 ***p < .001

Appendix Table A-1. Key features of the educational systems of the six countries

	Japan	Korea	Taiwan	Czech Republic	Germany	USA
Grade of first tracking	10	10	10	6	5	-
Main types of school attended in 9th grade	3-year junior high school	3-year middle school	3-year junior high school	9-year elementary school; 6-year and 8-year gymnasium	8-year gymnasium; 6-year Real school; 5-year middle school ; 5-year+ comprehensive school	4-year high school
Main types of school attended in 10th grade	Senior high school	High school	3-year senior high school; 3-year senior vocational school; 5-year junior college	4-year, 6-year and 8-year gymnasia; 4-year technical high school; 2-year, 3-year, and 4-year vocational high school	8-year gymnasium; 6-year Real school; 5-year middle school ; 5-year+ comprehensive school	4-year high school
Dead-end secondary tracks	No	No	No	Yes	Yes	No
Centralized exit examinations	Yes	Yes	Yes	Yes	Yes	No
Upper secondary graduation rate	96%	94%	88%	79%	87%	77%

Note: Data from OECD (2012) and authors' own sources. Grade of first tracking refers to the grade in which pupils are allocated on the basis of a competitive exam. Dead-end secondary tracks refer to the completion of a secondary school program (usually ISCED 3C) that do not qualify pupils for university entry.

Appendix Table A-2. Sample Size and Sample Selection

	Whole Sample				Analytic Sample				% deleted due to missing data			
	9th Grade		10th Grade		9th Grade		10th Grade		9th Grade		10th Grade	
Country	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys
Number of Students												
Japan			3021	3330			2525	2630			16.4%	21.0%
Korea			2209	2519			1973	2277			10.7%	9.6%
Taiwan	982	1007	2127	1924	866	839	1845	1628	11.8%	16.7%	13.2%	15.4%
Czech	1267	1494	1346	1040	956	1115	1068	815	24.5%	25.4%	20.6%	21.6%
Germany	1243	1363	985	844	706	733	611	490	43.2%	46.2%	38.0%	41.9%
USA			1830	1803			1528	1470			16.5%	18.5%
Number of Schools												
Japan			177	179			177	178			0.0%	0.6%
Korea			108	109			108	109			0.0%	0.0%
Taiwan	76	75	116	110	73	72	110	104	3.9%	4.0%	5.2%	5.4%
Czech	190	192	148	149	156	158	123	124	17.9%	17.7%	16.9%	16.8%
Germany	220	219	198	195	179	179	159	150	18.6%	18.3%	19.7%	23.1%
USA			157	157			150	150			4.4%	4.4%

Appendix Table A-3. Gender Gap in Mean Test Scores, by Subject, Grade, and Country: Analytic Sample

Grade/ Country	Subject	9th Grade				10th Grade			
		Girls	Boys	Gap	*d*-statistics	Girls	Boys	Gap	*d*-statistics
Japan	Math					513.672	532.046	-18.374	-0.204***
	Science					531.107	543.932	-12.825	-0.142***
	Reading					541.791	515.069	26.722	0.288***
Korea	Math					542.486	571.740	-29.254	-0.300***
	Science					533.571	546.798	-13.227	-0.167***
	Reading					545.695	531.883	13.812	0.168***
Taiwan	Math	536.983	528.251	8.732	0.084	576.919	584.323	-7.404	-0.065
	Science	513.886	507.297	6.589	0.091	534.143	535.989	-1.846	-0.022
	Reading	528.317	487.623	40.694	0.514***	551.847	521.844	30.003	0.350***
Czech R.	Math	490.059	506.621	-16.562	-0.202***	512.783	534.211	-21.428	-0.248***
	Science	510.446	513.767	-3.321	-0.044	521.118	533.403	-12.285	-0.154***
	Reading	512.097	477.320	34.777	0.483***	528.996	499.340	29.656	0.374***
Germany	Math	483.375	504.237	-20.862	-0.236***	551.106	574.701	-23.595	-0.295***
	Science	504.818	509.864	-5.046	-0.057	564.023	572.003	-7.980	-0.104
	Reading	511.975	475.837	36.138	0.438***	567.643	528.032	39.611	0.539***
USA	Math					486.659	501.788	-15.129	-0.196***
	Science					508.981	517.575	-8.594	-0.107**
	Reading					521.401	502.407	18.994	0.243***

Note: **p < .01 ***p < .001 for the hypothesis that there is no mean difference between girls and boys.

Appendix Table A-4: Doubled-Weighted Descriptive Statistics, by Country and Gender: 9th grade

	Taiwan		Czech R.		Germany	
Variable	Girls	Boys	Girls	Boys	Girls	Boys
PISA 2012 test scores						
Math	536.983	528.251	490.059	506.621	483.375	504.237
	(9503.316)	(12056.539)	(6257.519)	(7121.210)	(7455.902)	(8083.155)
Science	513.886	507.297	510.446	513.767	504.818	509.864
	(4595.057)	(5833.923)	(5240.850)	(6075.185)	(7714.947)	(8045.969)
Reading	528.317	487.623	512.097	477.320	511.975	475.837
	(5405.680)	(7139.447)	(4766.540)	(5556.469)	(6684.138)	(6925.459)
Student-level variables						
Father's schooling	12.020	12.145	13.064	12.976	13.087	13.495
	(7.451)	(7.833)	(3.805)	(4.340)	(11.948)	(10.487)
Mother's schooling	12.046	12.128	13.117	13.150	12.298	12.718
	(7.520)	(8.721)	(3.637)	(4.029)	(12.707)	(11.559)
Highest ISEI of parents	44.361	44.712	48.314	46.732	48.239	48.773
	(350.378)	(366.530)	(369.193)	(364.341)	(392.115)	(381.529)
Number of books in home	3.152	3.047	3.641	3.175	3.517	3.333
	(1.976)	(1.848)	(1.752)	(1.817)	(1.757)	(2.101)
Educational resources in home	5.392	5.145	6.266	6.103	5.958	5.820
	(1.906)	(2.374)	(0.705)	(0.974)	(1.124)	(1.601)
School-level variables						
School Sector (%)						
Public	91.2	91.7	97.7	97.8	95.1	95.6
Private	8.8	8.3	2.3	2.2	4.9	4.4
School Location (%)						
Urban	32.8	31.7	23.2	22.7	21.1	21.7
Rural	67.2	68.3	76.8	77.3	78.9	78.3
School Type (%)						
Academic	—	—	12.0	11.8	25.3	25.3
Vocational	—	—	88.0	88.2	58.9	59.9
Comprehensive	100.0	100.0	—	—	15.8	14.8
Sample Size						
# Students	866	839	956	1115	706	733
School	73	72	156	158	179	179

Note: Numbers in parentheses are variances.

Appendix Table A-5 Doubled-Weighted Descriptive Statistics, by Country and Gender: 10th grade

	Japan		Korea		Taiwan	
Variable	Girls	Boys	Girls	Boys	Girls	Boys
PISA 2012 test scores						
Math	513.672	532.046	542.486	571.740	576.919	584.323
	(6801.161)	(9402.178)	(8111.542)	(10749.873)	(11884.295)	(14204.089)
Science	531.107	543.932	533.571	546.798	534.143	535.989
	(6920.411)	(9421.133)	(5592.176)	(6797.058)	(6207.375)	(7345.211)
Reading	541.791	515.069	545.695	531.883	551.847	521.844
	(6805.466)	(10282.461)	(5525.986)	(7749.500)	(6754.715)	(7979.204)
Student-level variables						
Father's schooling	13.339	13.506	13.345	13.687	12.483	12.456
	(4.940)	(4.675)	(6.973)	(6.895)	(7.525)	(7.980)
Mother's schooling	13.188	13.258	13.098	13.315	12.368	12.410
	(3.000)	(3.493)	(5.342)	(6.269)	(7.299)	(7.439)
Highest ISEI of parents	48.772	48.422	51.754	53.911	48.559	47.608
	(396.834)	(409.449)	(355.078)	(359.889)	(393.555)	(400.803)
Number of books in home	3.301	3.459	3.818	4.016	3.334	3.261
	(1.837)	(1.912)	(1.767)	(1.760)	(2.221)	(2.175)
Educational resources in home	5.111	4.931	5.584	5.668	5.465	5.384
	(1.392)	(1.643)	(1.597)	(1.701)	(1.898)	(2.132)
School-level variables						
School Sector (%)						
Public	74.7	81.5	51.4	53.5	56.0	58.2
Private	25.3	18.5	48.6	46.5	44.0	41.8
School Location (%)						
Urban	64.0	61.0	65.5	69.4	58.4	55.9
Rural	36.0	39.0	34.5	30.6	41.6	44.1
School Type (%)						
Academic	65.0	66.0	76.5	75.3	42.7	43.1
Vocational	35.0	34.0	23.5	24.7	28.3	29.1
Comprehensive	—	—	—	—	29.0	27.8
Sample Size						
# Students	2525	2630	1973	2277	1845	1628
School	177	178	108	109	110	104

Note: Numbers in parentheses are variances.

Appendix Table A-5 (continued) Doubled-Weighted Descriptive Statistics by Country and Gender: 10th grade

	Czech R.		Germany		USA	
Variable	Girls	Boys	Girls	Boys	Girls	Boys
PISA 2012 test scores						
Math	512.783	534.211	551.106	574.701	486.659	501.788
	(7593.017)	(7349.385)	(6054.554)	(6811.539)	(5618.464)	(6251.306)
Science	521.118	533.403	564.023	572.003	508.981	517.575
	(6385.311)	(6248.796)	(5416.476)	(6573.123)	(5958.457)	(6941.335)
Reading	528.996	499.340	567.643	528.032	521.401	502.407
	(6297.298)	(6302.605)	(4511.636)	(6510.343)	(5540.557)	(6654.758)
Student-level variables						
Father's schooling	12.930	12.944	13.817	13.822	12.950	13.124
	(3.674)	(4.156)	(12.688)	(12.854)	(8.346)	(7.689)
Mother's schooling	12.964	13.063	13.361	13.175	13.368	13.560
	(3.507)	(4.048)	(10.454)	(10.601)	(7.078)	(7.084)
Highest ISEI of parents	48.300	49.405	54.254	52.538	55.784	55.580
	(372.787)	(340.806)	(404.993)	(443.604)	(400.109)	(417.591)
Number of books in home	3.501	3.327	3.922	3.673	3.153	3.020
	(1.700)	(1.897)	(1.750)	(2.001)	(1.908)	(2.167)
Educational resources in home	6.350	6.310	6.217	6.071	5.610	5.516
	(0.666)	(0.859)	(0.788)	(1.010)	(2.210)	(2.612)
School-level variables						
School Sector (%)						
Public	75.3	79.6	93.3	93.9	84.5	80.2
Private	24.7	20.4	6.7	6.1	15.5	19.8
School Location (%)						
Urban	20.0	23.8	20.7	22.6	18.1	18.3
Rural	80.0	76.2	79.3	77.4	81.9	81.7
School Type (%)						
Academic	33.5	32.0	32.8	32.7	—	—
Vocational	20.0	23.3	60.3	60.3	—	—
Comprehensive	46.5	44.7	6.9	7.0	100.0	100.0
Sample Size						
# Students	1068	815	611	490	1528	1470
School	123	124	159	150	150	150

Note: Numbers in parentheses are variances.