University of California
Santa Barbara

Essays in Education and Labor Economics

A dissertation submitted in partial satisfaction
of the requirements for the degree

Doctor of Philosophy
in
Economics

by

Eunseo Kang

Committee in charge:

Professor Kelly Bedard, Chair
Professor Dick Startz
Professor Heather Royer

June 2024

The Dissertation of Eunseo Kang is approved.

Professor Dick Startz

Professor Heather Royer

Professor Kelly Bedard, Committee Chair

May 2024

Essays in Education and Labor Economics

I dedicate this to my mother, whose unconditional love and unwavering support have always been my strength.

Acknowledgements

I am deeply grateful to Professor Kelly Bedard for her invaluable support and advice throughout my research and career. When COVID struck and my research process was stalled during the lockdown, her crucial guidance helped me stay focused, navigate the uncertainties, and persevere through the challenges. Her unwavering support has been instrumental in reaching this stage of my academic journey, and I cannot imagine achieving this accomplishment without her help.

I am also immensely thankful to Professor Dick Startz, who opened the door to my journey in economic research. After the qualitative exam in my first year of the PhD program, his suggestion to work as a research assistant during the summer provided me with invaluable hands-on experience and bolstered my confidence. This opportunity significantly influenced my progress throughout the program and sparked my interest in the Economics of Education. Additionally, I would like to express my gratitude to Professor Heather Royer for her crucial advice on effectively conveying the results of my study and for her technical and emotional support.

My deepest thanks go to my family for their unwavering support. My mother's sacrifice and devotion, along with her love of learning and pioneering a new field in her 60s, have given me the courage to explore and pursue greater opportunities in the wider world. Her belief in me has strengthened my confidence in myself, even in moments of doubt, and inspired me to dream big and aim high. To my husband, Hyunseok, thank you for always standing by my side and being a pillar of strength. Your unwavering support has been a source of immense comfort and motivation throughout my journey and will continue to be as we move forward together.

Curriculum Vitæ
Eunseo Kang

Education

2024	Ph.D. in Economics (Expected), University of California, Santa Barbara.
2024	M.A. in Statistics (Expected) , University of California, Santa Barbara.
2019	M.A. in Economics , University of California, Santa Barbara.
2014	B.A. in Economics , Yonsei University, Seoul, Republic of Korea.

Research Interest
Labor Economics, Economics of Education, Demographic Economics

Research in Progress

- Relative Age Effects on Educational Outcomes: International Evidence from the Developing World
  Presentation: All-California Labor Economics 2022 Conference (Poster), Western Economic Association International 2023 Conference, UCSB Applied Micro Lunch Seminar, Calvin University
- The Impact of International Students on Undergraduate Minorities in US Higher Education
- Trends in Academic Achievement in OECD countries: The Interaction of Gender and Socioeconomic Status (*Joint work with Kelly Bedard*)

Research and Professional Experience

Summer 2023	Economist Intern, Core AI, Amazon.com.
Summer 2020	Research Assistant to Prof. Dick Startz, University of California, Santa Barbara.
2015 - 2018	Junior Fund Manager, NH Life Co., Seoul, Republic of Korea.

Teaching Experience

2021 & 2022	Math Camp Instructor (PhD level), *Ratings:1.55 (2021), 1.53 (2022)*
2018 - 2024	Teaching Assistant, University of California, Santa Barbara.

Awards & Honors

Spring 2023	Research Quarter Fellowship, University of California, Santa Barbara.
2021	Best Second Year Paper Award (Robert T. Deacon Graduate Fellowship), University of California, Santa Barbara.
AY 2020-2021	Outstanding Graduate Course TA Award, University of California, Santa Barbara.

AY 2019-2020 Outstanding Undergraduate Course TA Award, University of California, Santa Barbara.

Skills & Additional

- **Programming/Skills**: Python, R, Stata, Spark, SQL, MATLAB, LaTeX, AWS, Google Cloud
- **Language**: Korean(native), English(fluent)
- **Citizenship**: Republic of Korea

References

Kelly Bedard (Advisor)
Professor, Department of Economics
University of California, Santa Barbara
kelly@econ.ucsb.edu

Dick Startz
Professor, Department of Economics
University of California, Santa Barbara
startz@ucsb.edu

Heather Royer
Professor, Department of Economics
University of California, Santa Barbara
royer@econ.ucsb.edu

Vikram Pathania
Principal Economist, Amazon
vikrpath@amazon.com

Abstract

Essays in Education and Labor Economics

by

Eunseo Kang

This dissertation contains three chapters in education and labor economics. In Chapter 1, I study whether relative age effects among fourth graders on math and science test scores exist in developing countries and investigate whether they are similar across all countries with different levels of development. Students with different birthdays who are subject to the same school-entry cutoff date have different ages at school entry. This difference in maturity may affect a child's outcomes in school because we might expect that students who are more mature relative to their peers will perform better; a phenomenon called 'relative age effects'. While the focus of previous studies has been limited to developed countries, this study aims to provide evidence of relative age effects in the context of developing countries. Using Trends in International Mathematics and Science Study data and assigned relative age as an instrumental variable that is formed exogenously by this cutoff, I find that positive relative age effects on test scores exist in developing countries, but they are smaller than those in developed countries. I also explore the educational factors correlated to the magnitude of relative age effects using cross-country data.

In Chapter 2, I examine the impact of the inflow of international students on the first-time, full-time enrollment of domestic minority students in US Higher Education using data from IPEDS. Since foreign enrollment is an endogenous variable, I employ the instrumental variables approach, using the institution's historical share of international students and the year's non-immigrant visa issuance. I find that there is no significant effect of the influx of international students on the new enrollment of domestic minorities as a whole. However, when I divide the institutions by the level of state funding per student, I find that an additional influx of international students increases domestic minority FTFT enrollment by 0.65. I suggest that this is because institutions with relatively little reliance on govern-

ment funding are more sensitive to the financial resources that international students bring in terms of determining the supply and demand of domestic minority enrollment.

In Chapter 3–joint work with Kelly Bedard, we examine how the academic achievement gap between different genders and socioeconomic groups within OECD countries has evolved over the years. Using Trends in International Mathematics and Science Study (TIMSS) data for eighth graders from eighteen OECD countries from 1995 to 2019, we first confirm that trends in academic achievement have progressed towards gender equality, particularly in science. Conversely, we find widening socioeconomic gaps, with high socioeconomic status (SES) groups showing greater improvements than low SES groups in both math and science test scores. When we examine the interactions between gender and socioeconomic groups to identify patterns driving these trends, we find that the SES gaps worsened for both males and females, and gender gaps similarly improved for both high and low SES students in most countries. Some countries show patterns that the worsening SES gaps are driven more by boys, and the improving gender gaps are driven more by the low SES students.

Contents

Chapter 1

Relative Age Effects on Educational Outcomes: International Evidence from the Developing World

Students with different birthdays who are subject to the same school-entry cutoff date have different ages at school entry. This difference in maturity may affect a child's outcomes in school because we might expect that students who are more mature relative to their peers will perform better; a phenomenon called 'relative age effects'. While the focus of previous studies has been limited to developed countries, this study aims to provide evidence of relative age effects in the context of developing countries. Using Trends in International Mathematics and Science Study data and assigned relative age as an instrumental variable that is formed exogenously by this cutoff, I find that positive relative age effects on test scores exist in developing countries, but they are smaller than those in developed countries. I also explore the educational factors correlated to the magnitude of relative age effects using cross-country data.

1.1 Introduction

When a country has a single school entry cutoff date, there is a range of ages in grade which is approximately one year between the youngest and the oldest child in the same cohort. This age difference in class is called *relative age*, and whether this age difference at school entry, generated by random birth months, makes a big difference in student outcomes has been an important empirical question for both individuals and society. First, if parents believe that a child would benefit from being older relative to other students in the cohort, they may strategically delay school entry[1]. According to (Deming and Dynarski, 2008), this delay accounts for two-thirds of the increase in school entry age over the past few decades. Knowing whether this behavior is based on scientific fact is important because delaying school entry also delays entry into the labor market and imposes additional childcare costs (Elder and Lubotsky (2009)).

The question is also of interest to society, as relative age effects can increase inequality. If relative age plays a role in generating differences in academic performance, this may lead to differences in other outcomes that ultimately determine an individual's long-term social status. The advantaged groups may seek to exploit this factor for the benefit of their children, thereby exacerbating social inequality. Therefore, recognizing relative age effects and developing intervention strategies to mitigate them are essential for educational practitioners or policymakers working to reduce educational inequalities in any country.

Previous studies generally find that the oldest kids in the grade perform better in school.[2] However, the focus of studies on relative age effects has been limited to developed countries largely due to the greater availability and quality of data in developed countries. Given the significant differences in educational factors, including government funding, teacher quality, grade repetition rates, and preschool systems, between developed and developing countries, it is not straightforward to generalize research from developed countries to developing country contexts, and relative age effects may not exist in developing countries.

[1] This delay is sometimes called *red-shirting*, a term that originated in college sports

[2] See Fredriksson and Ockert (2005),Bedard and Dhuey (2006), Datar (2006), Puhani and Weber (2008), McEwan and Shapiro (2008), Elder and Lubotsky (2009), Dobkin and Ferreira (2010), Black et al. (2011), Altwicker-Hámori and Köllő (2012), and Kawaguchi (2011).

Fortunately, comparable test score data for a sufficient number of developing countries have been recently available. Using these data and collecting information on school entry cutoffs from each developing country by hand, this study contributes to the literature by extending discussions on relative age effects, traditionally focused on developed countries, to encompass developing countries. This study addresses the following three questions: *Do relative age effects exist in developing countries? Is the effect of being relatively older than other students on test scores similar across all countries with different development statuses? If not, what factors can possibly explain the cross-country differences in relative age effects?*

To answer the questions, I use data on fourth-grade math and science test scores from fifteen developing countries and twenty-five developed countries/regions with different education systems from the Trends in International Mathematics and Science Study (TIMSS). Since identifying the causal relationship between gaps in observed age and test scores using OLS faces endogenous problems as students with worse grades are more likely to repeat a grade[3], I use school-starting age rules to construct an instrumental variable to estimate country-specific relative age effects and then compare these estimates across developing and developed countries. The exogenous variations in the cutoffs and birth months across countries are used to construct *assigned relative age* among students, which is an instrumental variable for *observed age*. Students born in the month shortly before the cutoff have the smallest value of assigned relative age and the month shortly after the cutoff has the largest value for assigned relative age. In this setting, assigned relative age is correlated with observed age, and at the same time, has exogeneity based on the assumption that the timing of birth is exogenous and randomly distributed.[4]

The study shows that in developing countries, with the exception of a few outliers, being just one month older in the cohort can have a relative age effect ranging from 0.015 to 0.025 standard deviations in mathematics and 0.011 to 0.031 standard deviations in science.

[3] Early promotion for gifted students results in endogeneity problem in a symmetric way.

[4] Numerous empirical studies try to deal with the endogeneity problem in the context of school entry age by exploiting an instrumental variable (Bedard and Dhuey (2006); Datar (2006); Puhani and Weber (2008); Elder and Lubotsky (2009); Black et al. (2011)) while others use regression discontinuity designs around the cutoff date for school eligibility (Fredriksson and Ockert (2005); McEwan and Shapiro (2008); Dobkin and Ferreira (2010); Crawford et al. (2014)). RDD requires small windows around the cutoff, while the TIMSS has information only about birth months for students. Therefore, the IV method is adopted in this study.

Assuming a linear relationship for the effect, the weighted average estimates suggest that being the oldest within a cohort relative to being the youngest corresponds to a relative age effect of about 0.19 standard deviation in mathematics and 0.21 standard deviation in science. This average relative age effect in developing countries, taking into account the maximum age difference within the cohort, accounts for about one-fourth of the free and reduced-price lunch (FRPL) gap and the white-black achievement gap among fourth graders in the United States.[5]

For comparison, the study also analyzes relative age effects in developed countries, which range from 0.015 to 0.036 standard deviations for math and 0.017 to 0.038 standard deviations for science test scores. The weighted averages of the relative age effects for developed countries are 0.027 standard deviations for both mathematics and science. Interestingly, the average relative age effect in developing countries is statistically significantly lower than in developed countries. On average, the relative age effect in developing countries is about 0.009 standard deviations lower for mathematics and 0.007 standard deviations lower for science. These differences imply that the age premium for the oldest students relative to the youngest students is about 0.1 standard deviations higher in developed countries than in developing countries, simply because of the difference in the level of development of the countries in which these students are born.

The differences in relative age effects across countries with different levels of development naturally raise the question of whether there are any explanatory characteristics that might explain these differences. Therefore, this study uses a cross-country analysis to investigate whether certain country determinants are correlated with relative age effects. Using a combination of school- and teacher-level data from TIMSS and country-level data from the World Bank, I show that a country's relative age effects are positively associated with the establishment of a preschool system and negatively associated with the repetition rate in primary education.

The contribution of this study is twofold. First, this is the first study to analyze rela-

[5]According to Hansen et al. (2018)'s calculations based on National Assessment of Educational Progress (NAEP) data, the free and reduced-price lunch (FRPL) gaps and the White-Black achievement gap in both math and reading among fourth graders in 2017 are 0.75 and 0.74 standard deviations, respectively.

tive age effects across multiple developing countries using internationally comparable data. While previous studies have examined relative age effects in developing contexts (Peña (2017); Peña (2020); Morales (2020); Ryu et al. (2020); Chen and Park (2021)), most have focused on the effect of changes in school-entry cutoffs within a single country. By collecting data on school entry cutoffs and test scores from multiple developing countries, this study goes beyond any country-specific settings and allows for the generalizability of positive relative age effects across developing countries. Moreover, by including multiple countries, this study is also able to capture variation that only exists across countries, such as institutional characteristics of school systems, which are often much more significant.(Hanushek and Woessmann (2011)) Second, this study provides a rigorous examination of the validity of using birth month as an instrumental variable for actual age, addressing concerns related to seasonality, birth-month targeting, and monotonicity. While existing studies have addressed the validity of instruments in this context to some extent, this study stands out as the first to comprehensively assess the full set of assumptions necessary to estimate the local average treatment effect within a consistent contextual framework.

This paper is organized in the following order. Section 2 reviews the relevant literature. Section 3 describes the data used in the analysis, and Section 4 discusses the empirical strategy of the study. Section 5 reports the main results on relative age effects in developing countries and compares them with those in developed countries. Subsections in Section 5 examine the validity of an instrumental variable, robustness checks, and heterogeneity analysis. Section 6 explains the relationship between relative age effects and educational characteristics across countries. Section 7 concludes the paper.

1.2 Literature Review

Previous studies have attempted to disentangle different age effects. Because age at test is the sum of age at school entry and years of schooling, there is a perfect linear relationship between these two ages–test age and age at school entry–which makes it difficult to separate them. For this reason, a number of studies, including this study, identify the combined effect of school entry age effect and age-at-testing effect (Bedard and Dhuey (2006); Datar (2006); Crawford et al. (2010); Puhani and Weber (2008); McEwan and Shapiro (2008); Altwicker-Hámori and Köllő (2012)). On the other hand, other studies focus on separating school starting age effects from test age effects by breaking the perfect linear relationship between two ages. They either use test scores taken outside a school at the constant age (Black et al. (2011); Fredriksson and Öckert (2014); Dee and Sievertsen (2018)), the difference in pupil's test scores over time (Datar (2006)), regional or temporal variation in school admission policy (Crawford et al. (2010), Elder and Lubotsky (2009); Smith (2010), Bedard and Dhuey (2012)), or an experimental setting (Cascio and Schanzenbach (2016)).

The literature has also focused on long-run outcomes, but the results show mixed evidence of effects. While some studies show positive age effects on test scores for eighth graders and on university attendance in OECD countries (Bedard and Dhuey (2006)), or positive effects on prime-age earnings for individuals with low-educated parents (Fredriksson and Öckert (2014)), Black et al. (2011) show negative but disappearing effects on earnings as individuals age and Dobkin and Ferreira (2010) find no evidence that the age at school entry affects job market outcomes. Many studies show that age effects become less important as a child grows up, however, Kawaguchi (2011) first finds that the positive effects of older school starting age do not wash out and have life-long effects in the context of Japan.

In addition, the literature studying age effects is divided into two streams: one that examines individual decisions that take the starting age rules as given, and another that focuses on policy changes such as cutoff date changes. The first stream, which this study belongs to, focuses on the relative age difference generated by a uniform cutoff itself. Since relative age effects always exist whenever the cutoff month is, this literature focuses on possible performance differences between the older and the younger kids in the cohort regardless of when the cutoff is or whether the cutoff has moved. In such studies, the results generally show positive effects of entering school older on school performance (Bedard and Dhuey (2006); Datar (2006); Puhani and Weber (2008); McEwan and Shapiro (2008);

(Elder and Lubotsky (2009); Altwicker-Hámori and Köllő (2012)). The second stream of research focuses on policy changes involving school entry cutoff date changes. Many countries and regions have implemented changes in cutoff months to increase the age of eligibility, aiming to enhance students' ability to engage with the curriculum better with more maturity. Studies examining these policy changes have found that backing up the cutoff improved educational achievement in 4th and 8th grades (Fletcher and Kim (2016)), as well as increased wages in adulthood in the United States (Bedard and Dhuey (2012)).

Several studies have examined relative age effects in a developing context, however, most of them have focused on policy changes to the school-entry cutoff within individual countries. For instance, studies by Peña (2017) and Peña (2020) utilize an unanticipated policy reform on school-entry cutoff in Mexico to demonstrate the advantage associated with relative age on achievement tests. Similarly, Morales (2020) examines relative age effects in Peru by analyzing a change in the school entry cutoff. Other studies have examined the effects of more country-specific school entry policies, but do not necessarily focus on relative age effects. For example, Ryu et al. (2020) examine a policy change in Brazil that lowers the school entry age while increasing the duration of primary education. In the case of China, Chen and Park (2021) study the effects of the first establishment of clear cutoff criteria, which were previously vague.

1.3 Data

The primary data in this study is the Trends in International Mathematics and Science Study (TIMSS).[6] The TIMSS assessment cycle is every four years and spans from 1995 to 2019. The data has both fourth and eighth graders, however, only the data for fourth graders are used. While long-term outcomes are interests of study about relative age effects, upper-division students are a subset of the original cohort, and oftentimes, they are not randomly selected due to dropping out. This selection bias tends to be more pronounced in developing countries, which decreases the number of compliers and complicates the identification.

[6]Students' birth month information is from restricted-use data.

In fact, data from the World Bank reveals a significant disparity in the average primary education repetition rates between developing and developed countries. The average repetition rate in primary education for the developing countries included in this study is 2.24%, which is more than twice the rate observed in developed countries (0.93%).[7] Therefore, this study focuses on whether there exist relative age effects at least up to lower division including fourth grade and other comparable grades.[8]

Table 1 lists the 77 countries included in the original TIMSS data along with information about school entry cutoffs and test years. Following World Bank Country Classification in 2019, relatively lower-income countries (countries with GNI (Gross National Income) per capita smaller or equl to \$12,535) are mapped to developing countries, and high-income countries with GNI per capita above \$12,535 are defined as developed countries.[9] 6 countries that changed their development status between 1995 and 2019 are categorized as Low-to-high-income countries and are excluded from the analysis due to their small number and the complication of the comparison between developing and developed countries.[10]

To be included in the analysis, each country's data from the original list had to satisfy certain conditions to properly construct the assigned relative age, which serves as an instrumental variable to identify a causal effect. The final sample, as described in Table 2, comprises 15 developing countries with 138,827 observations and 25 developed countries/regions with 592,845 observations.[11] The four conditions to determine the exclu-

[7] These statistics are calculated by the author using World Bank data from 1995 to 2019.

[8] Some countries include adjacent grades or have test scores for slightly different grades. For example, most countries have two adjacent grades in the data for 1995, and Norway has data both for grade 4 and grade 5 in 2015. For retaining the data points, all data points from the adjacent grades are included in the analysis and controlled by the grade-fixed effects. Grade 4 of the Netherlands in 1995 and Grade 4 of Slovenia in 2003 were dropped since the number of samples for that grade and year was too small.

[9] According to World Bank Country Classification in 2019, countries are divided into Low-income (L), Lower-middle-income (LM), Upper-middle-income (UM), and high-income countries. Among the entire countries in TIMSS data, Yemen is the only low-income country. L, LM, and UM countries are countries with GNI (Gross National Income) per capita smaller or equal to \$1,035, between \$1,036 and \$4,045, and between \$4,046 and \$12,535, respectively. High-income countries are those with GNI per capita above \$12,535.

[10] Chile, Czech Republic, Hungary, Latvia, Lithuania, and Slovenia are included in this category.

[11] UK's data are separately collected in England, Northern Ireland, and Scotland since each part has a different school entry cutoff. The TIMSS doesn't include the data for Wales. Only partial years of the data for Iran, Poland, Georgia, Armenia, Kuwait, and Latvia are included in the analysis. For more information about the reason for the exclusion of certain countries or years, see Table 1.

sion of specific countries and data are as follows. First, there must be reliable information about the school entry cutoff. Based on the documents provided by the IEA (International Association for the Evaluating of Educational Achievement), countries with no information about school entry cutoff are dropped. Second, if the empirical distribution of births does not align with the officially informed school entry cutoff of a country, that country is excluded from the analysis, even if information about the cutoff is available. Third, there must be a single school entry cutoff month in a country. If there are multiple cutoffs or cutoffs are different across regions and the data for those regions are not collected separately, those countries are dropped. Lastly, the school entry cutoff month must be well-settled. If a country is in the middle of cutoff month changes, and therefore, when the distribution seems to fluctuate in those transitional years, those years are dropped. As an exception, Iran's school-entry cutoff is in the middle of September and Azerbaijan allows students born in October or November to enroll earlier by law. In this case, it is challenging to distinguish whether a student was born before or after the cutoff month. Therefore, data points for students born in September for Iran and October/November for Azerbaijan are not assigned IV values and will be excluded from the 2SLS analysis.

TIMSS achievement scores are scaled based on the achievement distribution in the TIMSS 1995 across all participating countries, treating each country equally. Achievement data from the subsequent TIMSS assessment cycles were linked to these scales so that increases or decreases in average achievement might be monitored across assessments.[12] I standardized these scores with 0 as the mean and 1 as the standard deviation within the testbook-country-year level. In all analyses, sampling weights provided by the TIMSS are

[12]TIMSS assessments use item response theory (IRT)-based domain scoring method for providing a flexible framework for estimating proficiency scores from students' responses to test items. In mathematics as well as in science, this translates into several hundred achievement items, only a fraction of which can be administered to any one student given the available testing time. Therefore, these achievement items are arranged in blocks that are then assembled into student booklets that contain different (but systematically overlapping) sets of item blocks. IRT is particularly well suited to handle such data collection design in which not all students are tested with all items. TIMSS uses a population model to estimate distributions of proficiencies based on the likelihood function of an IRT model. Five plausible values are generated by imputing based on background characteristics. Although it is recommended to use all five plausible values, I only use the first plausible value because plausible values and multiple imputation techniques are not yet well established in IVs estimation and statistical tests applied routinely in IVs analysis (Pokropek (2016)). Please see Methods and Procedures in the TIMSS and PIRLS 2019 on the TIMSS and PIRLS website for further details (http://timssandpirls.bc.edu).

used. All estimation in this study includes basic individual-level socioeconomic controls collected from the TIMSS. These controls include sex, the index for the number of books in the household, parental education level, and whether the student has a desk.[13] The index for books in the household is 0 for 0-10 books, 1 for 11-25 books, 2 for 26-100 books, 3 for 101-200 books, and 4 for more than 200 books. The index for parental education level is 0 for some primary, lower secondary, or no school, 1 for lower secondary, 2 for upper secondary, 3 with post-secondary but not the university, and 4 with university or higher. Dummy variables indicating missing data are included in the covariate matrix in the analysis. For further analysis, students with high socioeconomic status (high SES), are defined as those with either parental education is at least a bachelor's degree holder or the number of books in the household is more or equal to 100.[14] Low SES is defined as non-high SES. In addition, a categorical variable about the age when a child started primary school is collected from the TIMSS and it has categories of '5 years old or younger', '6 years old', '7 years old', and '8 years old or older'. Since more than half of the observations are missing this information, this information is only used as an additional control in Section 5.

1.3.1 Descriptive Statistics

Table 2 reports descriptive statistics for the final list of developing and developed countries for analysis. On average, developing countries have significantly lower math and science scores compared to developed countries. Additionally, the average age of students in developing countries is 10.08 while that of developed countries is 10.18. This suggests that developing countries tend to have younger students. Interestingly, the percentage of female students in developing countries is 47.97%, which is lower than that of developed countries (49.35%). Moreover, the proportion of students who have a desk, the index for books, and the index for parental education all show that developing countries have lower educational resources and parental education levels. Consequently, the proportion of low

[13] The information about the parental education level has the data points only after the year 2011.

[14] The reason for synthesizing the standards to define the high SES is to keep as many observations as possible when some countries don't have data points for parental education. I could derive similar results in the same exercise with the high SES definition using the number of books only.

SES students in developing countries is higher than that of developed countries. Although there are some missing data points for the entrance age variable, the available data points indicate that the average entrance age in developing countries is slightly higher than that of developed countries. Overall, the descriptive statistics suggest that there are notable differences between developing and developed countries in terms of educational outcomes and resources.

1.4 Empirical Strategy

Equation (1) is used to estimate the effect of relative age on test scores,

$$Y_{it} = \beta_0 + \beta_1 A_{it} + X_{it}\beta_2 + \epsilon_{it} \tag{1.1}$$

where Y_{it} is student i's standardized test score at time t, A_{it} denotes *observed age* in months, X_{it} is a matrix for control variables including sex, the index for the number of books in the household, parental education level, whether the student has a desk, dummies for each missing variable, and year/grade fixed effects,[15] and ϵ_{it} is an error term. The parameter of interest is β_1. However, simple OLS estimates are insufficient to explain the causal effect of relative age on test scores due to the presence of unobservable variables such as parent's preferences, family background, and student's ability, all of which influence both the observed age A_{it} and test scores Y_{it} simultaneously. For instance, students with lower academic performance are more likely to be held back, leading to a negative relationship between their scores and age. Symmetrically, early promotion of students with exceptional academic performance can introduce a similar bias. Ignoring this factor could result in underestimating the true effect of relative age on test scores β_1. Similarly, parental decisions regarding accelerating or redshirting of their children can also lead to biased estimates of β_1. Parents with higher-ability children may choose to enroll them early, while those with lower-ability children may delay their enrollment, creating a spurious relationship between

[15] Since some countries have data for multiple grades depending on their educational system, the grade fixed effects are included to estimate the within-grade effect.

age and test scores. On the other hand, lower-income parents may be more likely to enroll their children on time to save the cost of childcare services, potentially leading to a positive relationship between relative age and test scores and upward biased estimates for β_1.

To address the endogeneity problem, *assigned relative age* R_{it} is used as an instrumental variable. This variable is constructed based on the birth month relative to each country's school entry cutoff. The youngest student in the cohort is assigned $R = 0$, while the oldest is assigned $R = 11$. For example, if the cutoff date is January 1, students born in January will be the oldest in their cohort, assigned $R = 11$, while those born in December will be the youngest, assigned $R = 0$.[16] It is worth noting that assigned relative age is determined based on birth months rather than observed ages. This implies that even if a student born in December is held back a year, they would still be assigned $R = 0$, rather than $R = 12$. Therefore, assigned relative age can be considered a "predicted" age, in accordance with the terminology used in Bedard and Dhuey (2006), and is uncorrelated with confounding factors that may be associated with the outcomes under the assumption that the decision of birth month is exogenous. Figure 1 shows the distribution of birth months and assigned relative age by birth months for each developing country. As explained earlier, assigned relative age jumps at the cutoff month since it depends entirely on birth months. While there are some variations in the percentage of births by each month, the distribution of birth months appears evenly distributed, with the highest percentage for a particular month being around 10%. These results support the argument that assigned relative age is almost randomly determined.

Therefore, this study employs assigned relative age as an instrumental variable and estimates the two-stage least squares (2SLS) of Equation (1) with the first-stage (FS) equation presented as:

$$A_{it} = \alpha_0 + \alpha_1 R_{it} + X_{it}\alpha_2 + e_{it} \tag{1.2}$$

[16] As previously mentioned, this study excludes data points for students born in September for Iran and October/November for Azerbaijan when running 2SLS regression as R is not assigned for these months for each country.

where R_{it} refers to the assigned relative age, X_{it} is the same matrix for control variables as in Equation (1), and e_{it} represents the error term. Robust standard errors are clustered at the school level.

Since 2SLS estimates capture the Local Average Treatment Effect (LATE) (Imbens and Angrist (1994)), it is important to note that the interpretation of the IV estimates is limited to the effects of students who follow the age regulations and entry cutoff when entering school.[17] To interpret the IV estimand as the average causal effect for compliers when the effects are heterogeneous, key assumptions, including the Stable Unit Treatment Value Assumption (SUTVA), a strong first stage, the independence assumption, the exclusion restriction, and monotonicity, must be satisfied (Imbens and Angrist (1994); Angrist and Imbens (1995); Angrist et al. (1996)). In this section, I examine the validity of the first two assumptions and defer the discussion of the remaining assumptions to Section 5, as it is preferable to present the primary results before delving into the discussion of these assumptions. First, SUTVA is reasonable to assume since there is no compelling reason to believe that a student's potential test scores are influenced by other students' birth months or their assigned relative age based on the school-entry cutoff month.

Another critical requirement for employing an instrumental variable is a strong correlation between the instrument and the endogenous regressor, which, in this case, is the observed age. This assumption is testable by checking the significance of the estimates for α_1 in Equation (2). However, the relationship between assigned relative age and observed age may not always be robust in some countries, as parents may not follow school-entry regulations to secure a more favorable age rank for their child, or students may repeat grades, causing the weak connection between the two variables. Figure 1 and Figure A1 illustrate that the assigned relative age is highly predictive of observed age across both developing and developed countries, as student's observed and assigned relative ages tend to be the highest in the months following the school-entry cutoff and lowest in the months preceding it. Although the level of consistency between the two variables is lower in developing

[17] Elder and Lubotsky (2009) caution that this LATE may disproportionately reflect the experience of low SES students whose parents are more likely to comply with school entry policies.

countries compared to developed ones, it is still noteworthy that these variables exhibit very similar trends in developing countries. Additionally, the strong first-stage assumption can be directly tested, as shown in Table 3, which presents first-stage results from Equation (2) for each country. The second column for each developing and developed country indicates significant positive first-stage estimates, confirming that the strong first-stage assumption is satisfied. Furthermore, the quality of the IV estimates is assessed using F-statistics in the subsequent columns. As per convention, a weak instrument test is deemed satisfactory if the F-statistic is greater than or equal to 10 for a single endogenous regressor (Staiger and Stock (1994); Stock and Yogo (2002)).[18] The results demonstrate F-statistics that exceed 100, thereby passing the weak instrument test at the 0.01 level.

1.5 Results

1.5.1 Relative Age Effects in Developing Countries

The results for all students in developing countries are presented in Table 4A. The first columns for each subject display the OLS estimates, while the second columns for each subject show the 2SLS estimates, which correspond to the estimates for β_1 in equation (1) for each country. The estimates in columns (1) and (3) suggest that there are considerable non-compliers with age regulations in developing countries, as indicated by the downward OLS biases compared to the 2SLS estimates. The 2SLS results in columns (2) and (4) reveal that developing countries have significant positive relative age effects. Specifically, an additional month of relative age increases the average math test score by 0.015-0.025 standard deviation and the average science test score by 0.011-0.031 standard deviation, with a few exceptions where either no effects or extremely high effects are observed in Azerbaijan and Ukraine.

Although Ukraine has notably large relative age effects for both subjects, the majority of developing countries exhibit relative age effects lower than 0.03σ. To derive the average

[18] The critical value increases to 16.38 in Stock and Yogo (2002) for a single endogenous regressor.

relative age effects among developing countries, an alternative equation (3) is adopted to use the entire data across all developing countries with country indicator matrix C_c.

$$Y_{ict} = \gamma_0 + \gamma_1 A_{ict} + X_{ict} d_2 + C_c \gamma_2 + v_{ict} \tag{1.3}$$

In the last row of Table 4A, the average relative age effects are 0.017σ for math and 0.019σ for science. Assuming linearity of the effect, the average estimates could be interpreted as the relative age effects of being oldest ($R = 11$) within a cohort compared to being relatively youngest ($R = 0$) as 0.19σ for math and 0.21σ for science. When assuming a normal distribution of test scores, this advantage of being 11 months older within a cohort in developing countries equates to a 6.5 percentile premium for math and a 7.2 percentile premium for science around the mean of test score ranking, which is a significant impact.[19]

1.5.2 Comparison with Developed Countries

Table 4B presents the results for developed countries. Similar to the findings in developing countries, the OLS estimates in columns (1) and (3) are downward-biased, indicating the presence of non-compliers with age regulations in developed countries, with the exception of a few countries such as Chinese Taipei, England, Japan, Northern Ireland, and Norway.[20] The results demonstrate that developed countries exhibit significantly positive relative age effects, consistent with previous research. Specifically, as shown in columns (2) and (4), an additional month of relative age increases the average math test score by 0.015-0.036 standard deviation and the average science test score by 0.017-0.038 standard deviation, with the exception of Israel. The majority of developed countries show relative age effects ranging from 0.02 to 0.03σ, with some countries exceeding 0.03σ.

When controlling for all covariates and country indicators in equation (3), the weighted average relative age effects for developed countries using pooled samples are 0.028σ both

[19] A histogram and Q-Q plot for test scores in each country show the data approximately follows a normal distribution. As a standard deviation around the mean roughly corresponds to a 34.13 percentile test score ranking premium, this percentile premium is calculated as $34.13 \times 0.19 = 6.5$.

[20] These countries are referred to as “clean countries” in Bedard and Dhuey (2006).

for math and science, as shown in the last row of Table 4B. Assuming linearity, this implies that the relative age effect of being the oldest ($R = 11$) within a cohort compared to the youngest ($R = 0$) is 0.31σ both for math and science. Thus, the advantage of being 11 months older within a cohort in developed countries corresponds to a 10.5 percentile premium for both subjects in test score ranking.

One noteworthy finding is that the weighted average of relative age effects in developing countries is statistically significantly smaller than in developed countries, with greater heterogeneity in these effects as described in Figure 2. To test whether the difference is statistically significant, an equation that includes an interaction term for developing countries is used as follows:

$$Y_{ict} = \phi_0 + \phi_1 A_{ict} + \phi_2 A_{ict} \times D_c + X_{ict}\phi_3 + C_c\phi_4 + \eta_{ict} \tag{1.4}$$

where D_c is a dummy variable with one when a country c is a developing country and zero otherwise. Adding an interaction term complicates the causal inference since $A_{ict} \times D_c$ is again going to be an endogenous variable. Therefore, I add $R_{ict} \times D_c$ as the second instrument to derive 2SLS estimates with two endogenous regressors and two instruments.

The third panel in columns (2) and (6) of Table 5 presents the difference in average relative age effects between developing and developed countries for math and science. On average, developing countries exhibit lower relative age effects of 0.009σ and 0.007σ for math and science, respectively. These differences translate to an additional premium of 0.1σ (equivalent to a 3.4 percentile premium) and 0.08σ (equivalent to a 2.7 percentile premium) in math and science scores, respectively, for the oldest students when compared to the youngest students, assuming they were born in developed countries instead of developing countries.

This disparity is further highlighted in Figure 3, which shows that the IV estimates for developed countries are consistently higher than those for developing countries. Specifically, while the two groups display a similar spread of first-stage estimates, the reduced form estimates for developed countries (represented by navy circles) are generally located above those for developing countries (represented by orange hollow diamonds).[21] The fitted lines indicate the average IV estimates for developed and developing countries, respectively. Again, the line for developed countries is above the line for developing countries, meaning that the average relative age effect for developed countries is higher than that for developing countries.

1.5.3 Validity of Instrument

Independence Assumption

In the context of this study, one concern about the instrumental variable estimation is that assigned relative age may be influenced by unobserved factors that could impact test scores, thus violating the independence assumption. If parents intentionally target a child's birth month for specific reasons, then the independence assumption is violated, resulting in biased IV estimates. One possible reason for birth month targeting is when high socioeconomic status parents try to benefit by having a child born in the quarter immediately following the school-entry cutoff to make them one of the oldest children.[22] For this reason, the literature has warned about the links between birth month and family income, suggesting that parents with different socioeconomic backgrounds may engage in birth month targeting (Bound et al. (1995); Bound and Jaeger (2000); Buckles and Hungerman (2013)).

To check for evidence of birth month targeting, the proportion of births for each quarter and differences in proportions between high and low SES children are examined in Table A1-1 and Table A1-2. Q1 is the first three months after the school entry cutoff, and Q4 is the three months before the cutoff, with Q1 representing the oldest group and Q4 the youngest in each country. All proportions are population-weighted. In both developing and developed countries, the results support the independence assumption of the instrumental

[21] The reduced form (RF) equation is presented as $Y_{it} = \gamma_0 + \gamma_1 R_{it} + X_{it}\gamma_2 + u_{it}$ where γ_1 denotes the "net" impact or intent-to-treatment effect of assigned relative age on test scores, which includes the effect from grade repetition or early/late entry. The slope of the line that passes through the origin and the point is interpreted as the IV estimate, as IV estimate=Reduced form estimate/First stage estimate.

[22] Some countries may have the opposite pattern, with high SES parents trying to place their children among the youngest.

variable, as each quarter's proportion of births does not significantly differ from others in most cases. While there are some relatively small proportions in Q4 for some countries, they are not at extreme levels staying above 21% of total samples, with the highest proportion never going above 30%.

In addition, differences in the proportions of births in each quarter between high and low SES groups are investigated to check for potential endogeneity due to parental birth-month targeting. As described in Bedard and Dhuey (2006), targeting patterns by high SES groups could result in significant differences in birth proportions between Q1 and Q4, where high SES groups aim to avoid belonging to the youngest group and target the oldest. In such a case, significant figures with opposite directions in Q1(+) and Q4(-) can be detected. The starred figures in the tables indicate that the fraction of births to high SES children differs from the fraction of births to low SES at the 5 percent level or better. In Table A1-1, the results indicate that in developing countries, there is no significant pattern of birth months across different SES groups, except for a slight negative selection of age by high SES in some countries such as Argentina, Macedonia, Montenegro, Tunisia, and Ukraine. However, this may bias the estimates downward, which makes the results more conservative since the statistically significant estimates of positive relative age effects in developing countries would simply be a lower bound of the true relative age effects. In Table A1-2, most developed countries do not show a significant pattern of birth months across different socioeconomic statuses, except for Belgium and Israel.

Another way to ensure that the independence assumptions are met is to include observable variables representing the socioeconomic status in the model. If there is no significant difference in the estimation results with and without these controls, it suggests that the independence assumption is likely to be satisfied. To test this, different specifications of the regression model, with and without home background controls such as parental education, the number of books, and whether a student has a desk, were compared using pooled samples. The results showed consistency in both cases in Table 5. Specifically, comparing columns (1) and (2) for math and columns (5) and (6) for science reveals no differences in the results with and without the home background controls. This provides additional evidence for the independence of the instrument.

Exclusion Restriction

Another assumption that could potentially be violated is the exclusion restriction, which assumes that the assigned relative age based on birth month affects test scores only through observed age. This assumption may be violated if there is a direct connection between birth month and potential outcome determinants such as ability. However, it is unlikely that birth month directly causes differences in test scores, which arguably satisfies the exclusion restriction. Nonetheless, some studies suggest that seasonality in the birth timing and relative age can be related to factors that affect the outcome variable, resulting in a direct effect of the instrument on the dependent variable (Bound et al. (1995), Bound and Jaeger (2000), Buckles and Hungerman (2013)). According to the literature, such seasonality can arise from differences in health, regional patterns, family income, or even personality. Therefore, to separate the effects of relative age and season of birth, an alternative specification that adds month of birth indicators to equation (3) is estimated. This approach is feasible when using data pooled across countries with different cutoff dates, as children born in the same calendar month with the same season of birth can have different relative ages if they live in different countries (Bedard and Dhuey (2006)). The alternative equation is given as follows:

$$Y_{ict} = \theta_0 + \theta_1 A_{ict} + X_{ict}\theta_2 + C_c\theta_3 + M\theta_4 + \xi_{ict} \tag{1.5}$$

where country indicators are represented by matrix C_c, and month of birth indicators are denoted by matrix M.

The results of this specification are presented in columns (3) and (7) of Table 5. When including all control variables and seasonality controls, the weighted average relative age effects using pooled samples of all developing countries are 0.019 standard deviation for math and 0.021 standard deviation for science. Hence, these estimates, which are generally consistent but slightly larger, suggest that the influence of seasonality is minimal and help alleviate concerns regarding the violation of the exclusion restriction.

Monotonocity

In order to interpret the 2SLS estimand as Local Average Treatment Effects (LATE) for compliers under heterogeneous treatment effects, it is essential to satisfy the monotonicity assumption. This assumption states that there should be no individual who, if born in the month after the cutoff, would not have enrolled in school in the eligible year, but would have if born in the month before the cutoff. Barua and Lang (2016) argue that standard instruments such as a quarter of birth and legal entry age provide inconsistent estimates of LATE because they violate monotonicity in the context of the US.[23] Hence, testing for violations of monotonicity is critical to achieving a causal interpretation of the relative age effects in developing countries.

Direct testing for monotonicity is not feasible, but there is an indirect method to assess it. The literature suggests that the first-stage estimates should be nonnegative for subsamples based on interactions of covariates (Angrist et al. (1996)). However, this study employs a multivalued treatment and instrument, whereas the implication in most research is based on the standard framework with binary treatment and instrument. Even though Angrist and Imbens (1995) suggests testing for instrument monotonicity in the multivalued treatment case using CDF, this approach also depends on binary instruments. Therefore, to examine monotonicity, I simplify the treatment and instrumental variables to binary cases and limit the samples to students born in three-month windows before and after the cutoff, as the primary concern for monotonicity arises among individuals born around the cutoff month.[24]

In this analysis, the treatment and instrument variables are redefined as binary terms: $D = 1$ if the observed age is above the cohort average and $D = 0$ otherwise, and $Z = 1$ if

[23] They find strategic behavior of parents in the US such that children born in May before the cutoff enter kindergarten in September following their fifth birthday while some children born in October after the cutoff enter before their fifth birthday.

[24] Note that alternative methods to test monotonicity in multivalued treatment and instrument cases exist, such as extension proposed by Kitagawa (2015)) for multivalued discrete instruments and the relaxation method under "compliers-defiers" (CD) condition presented by De Chaisemartin (2017)) in the supplement to the paper.

the student's birth month is within three months after the cutoff month and $Z = 0$ if it is within three months before the cutoff month. Then I test whether the first stage is positive in all subsamples, as is common in the literature using IV approach (Angrist et al. (1996); Dobbie et al. (2018); Bhuller et al. (2020); Agan et al. (2021)).

The data is divided into four groups based on the interaction of covariates: Male × High SES, Female × High SES, Male × Low SES, and Male × low SES. Table A2-1 and Table A2-2 present the results for developing and developed countries, respectively. Among developing countries, most countries show a strong positive first-stage relationship in each covariate group except for the Female × High SES group in Algeria. In developed countries, the first stage for each group exhibits a stronger positive relationship, with no exception observed. Hence, we can infer that the violation of monotonicity is a minor concern for both developing and developed countries included in this study.

1.5.4 Robustness Checks

Nonlinear Specification

The baseline model assumes a linear relationship between assigned relative age and test scores, as R is assigned from 0 to 11. To relax this assumption, a nonlinear specification is adopted. This specification uses three instrumental variables: Q1, Q2, and Q3, where Q1 represents the oldest relative quarter, and Q3 represents the second youngest relative quarter. The youngest quarter, Q4, is omitted. The results of this specification for developing countries are presented in Table A3. In the first columns for each subject, the point estimates and significance levels are largely similar to the original results presented in Table 4A.

In addition, using a single endogenous regressor A with multiple instrument variables facilitates the over-identification test that assesses instruments' exogeneity by measuring the correlation between instrument and error (Sargan (1958); Angrist and Krueger (1992)). The J-statistics from the Sargan-Hansen test of over-identifying restriction are provided in the second columns for each subject. The null hypothesis in this test is that the instruments are uncorrelated with the error term. In the third columns for each subject, except for a few countries such as Argentina, Montenegro, and Tunisia, the null hypothesis is not rejected with large p-values, implying that additional instruments satisfy exogeneity.

Restricted sample around the cutoff

The baseline model assumes the linearity of relative age effects, meaning that the impact of a one-month age gap is consistent across all months. However, it is possible that certain adjacent months exhibit stronger age effects, which could drive the overall results. In such cases, interpreting significant differences between the oldest and youngest child in a cohort by simply multiplying monthly effects can lead to incorrect conclusions. Therefore, to directly compare the youngest and oldest child and assess whether the results hold around the cutoff, samples are restricted to students born in the 1st, 2nd, and 3rd months on either side of the school entry cutoff.

The results, shown in Table A4 with varying windows, incorporate basic controls for gender, socioeconomic background, year, grade, and country fixed effects. Columns (1) and (4) show results with a window of one month, directly comparing children born shortly before and after the cutoff. Although this narrow window decreases sample size, average effects remain statistically significant, hovering around 0.02σ for developing countries. Average relative age effects for developed countries remain significantly positive but slightly decrease to approximately 0.025σ, and the statistical significance of the difference between the two groups disappears. Columns (2) and (5) show results with a two-month window, while columns (3) and (6) show results with a three-month window. As the window widens, average relative age effects for developed countries increase, and the difference between developing and developed countries regains statistical significance. Therefore, it suggests that relative age effects exist when directly comparing the oldest and youngest children, but the difference in effects between developing and developed countries is more of an on-average phenomenon across all birth months.

1.5.5 Heterogeneity

Socioeconomic Status

I further conducted investigations to examine whether relative age effects vary across different socioeconomic backgrounds within each country when taking into account their different development statuses. Previous research has suggested that the variation in relative age effects may be linked to disparities in exposure to alternative preschool systems among different socioeconomic groups.[25] If the polarization of alternative systems across socioeconomic statuses is more severe in developing countries compared to developed countries, it is possible that different patterns in the socioeconomic gap regarding relative age effects may emerge among countries with different development statuses.

Figure A2 visually depicts the disparities in relative age effects on math and science scores by socioeconomic status within each country. The estimates for developing and developed countries are presented in Table A5-1 and Table A5-2, respectively. While some developing countries, such as Greece, Armenia, Azerbaijan, and Ukraine, show larger relative age effects among high SES students compared to their low SES counterparts, Argentina demonstrates much larger relative age effects among low SES students. As for developed countries, Austria, Italy, Korea, and Norway show larger relative age effects among high SES students, while Japan exhibits larger relative age effects among low SES students. In general, there are no clear patterns in relative age effects across different socioeconomic backgrounds within a country, regardless of whether the country is developing or developed. Instead, the variance observed across countries with different development statuses appears to be more significant than the variation observed across different socioeconomic backgrounds within a country. These results align with a recent study by Dhuey et al. (2019), which also found a lack of heterogeneity in the effect of relative age on test scores across various demographic and socioeconomic groups within a country.

[25] Elder and Lubotsky (2009) found larger relative age effects among higher socioeconomic status individuals in the US. On the other hand, Datar (2006) discovered a more significant effect among at-risk or low socioeconomic status children who may not have access to high-quality preschool or daycare options if not enrolled in school.

Gender

Similarly, examining the heterogeneity of relative age effects by gender within a country is an important aspect to consider. Previous research has yielded mixed conclusions regarding this heterogeneity. For instance, Cascio and Schanzenbach (2016) found that relative age effects tend to be more pronounced among boys than girls based on data from Tennessee's Project STAR. Conversely, Datar (2006) showed that girls seem to benefit more in math, while boys benefit significantly more in reading when delaying school entrance in the US context. If differences in physical and cognitive development, social interactions, or learning styles between genders influence relative age effects, gender-specific patterns may be detected in the analysis across countries with different development statuses.

The results of the gender analysis are presented in Figure A3, Table A6-1, and Table A6-2. For developing countries, it is interesting to observe that several countries, such as Algeria and Macedonia for both subjects and Greece specifically for science, exhibit significantly larger relative age effects among female students compared to their male counterparts. In the cases of Algeria and Macedonia, it is particularly noteworthy that the relative age effects are primarily driven by girls, indicating delaying school entry and being the older group in the cohort tends to benefit girls in these countries. On the other hand, for developed countries, most countries do not demonstrate substantial differences in relative age effects between genders, except for a few instances with minor variations, such as Belgium, Japan, and Norway. Only Israel shows significantly large relative age effects among boys in math. These mixed findings or lack of heterogeneity in relative age effects by gender is consistent with Bedard and Dhuey (2006) in the context of developed countries.

1.6 Exploring the relationship with educational characteristics

Previous studies have suggested several mechanisms for how relative age effects might happen. Datar (2006) suggested that a curriculum that is geared to the average developmental level of the student, or how the child spends her time during the extra year she is out of school, could generate relative age effects. The tricky part is that these potential mechanisms tend to be universal or have very little variation within a country, making it difficult to observe evidence for them. From this perspective, cross-country analysis has the advantage of being able to use variations in institutional characteristics across countries to gain some insight into how these relative age effects arise. In particular, given that relative age effects are observed to be smaller in developing countries than in developed countries, they may be systematically related to country-specific educational characteristics. Therefore, in this section, I estimate the associations between several characteristics of the educational system across countries and relative age effects using a multivariate regression model as shown in equation (6). As it is impossible to include a full set of country-specific measures, omitted variable bias is a clear issue. Consequently, this exercise should be viewed as exploratory rather than causal.

$$\hat{\beta_c}^{IV^*} = \pi_0 + \sum_j \pi_j \Theta_{cj} + u_c \tag{1.6}$$

The star on the dependent variable is there to remind readers that $\hat{\beta_c}^{IV}$ is multiplied by 11 so it can be interpreted as the relative effect of being the oldest child compared to being the youngest child within a cohort.[26] Θ_{cj} includes an index for primary school investment, the repetition rate in primary education, an index for teacher quality, and an index for preschool system establishment of a country c.

The educational characteristics variables come from the World Bank country-level data and the TIMSS school- and teacher-level data. Each educational characteristic index for Θ_{cj}, except the repetition rate, is constructed using the standardized inverse covariance weight[27] of various variables in these datasets. The country's education expenditure index

[26]There are two reasons for multiplying relative age effects by 11. First, the relative age effect of being the oldest one compared to being the youngest one can be more of interest to parents and policymakers since people typically make year-based decisions about the school entry age. Second, it is necessary to scale up the effect size to interpret the association more conveniently.

[27]I use the swindex command in STATA, which implements the generalized least squares method of index construction proposed by Anderson (2008, Journal of the American Statistical Association 103: 14811495). The procedure increases efficiency by ensuring that highly correlated outcomes receive less weight than un-

is a weighted average of GDP per capita (PPP adjusted), government expenditure per pupil in primary education as a percentage of GDP per capita, the average number of computers in a school, the status of shortage of instructional materials, and the number of pupils in a class. The teacher quality index is the weighted average of the average length of teaching experience and the level of formal education completed. Finally, the index of the degree of the preschool system establishment is the weighted average of the preschool enrollment rate and the duration of preschool attendance. More details on each variable are described in Table A7.

The estimates of π_j are presented in Table 6, categorized by pooled, developing, and developed country data. The analysis using pooled data is presented to show the relationship with larger variations in country-specific characteristics globally and to determine whether the association is driven solely by the developing or developed world or is pervasive worldwide. The results of the estimation, also described in Figure 4A-Figure 4D, suggest that relative age effects are associated with a repetition rate in primary school and the establishment of the preschool system, while not having strong associations with the level of educational investment in primary education or the quality of teachers in a country when controlling for other factors.[28] While educational resources and teacher quality are crucial determinants of academic achievement (Hanushek et al. (2019); Singh (2020)), they may not significantly contribute to the *variations* in academic achievements among individuals due to relative age differences within a class, as their impact can be consistent across different age groups. Conversely, the correlations between relative age effects and the repetition rate in primary schools, as well as the establishment of preschool systems, appear to be robust when controlling for other characteristics, particularly in mathematics and are more pronounced in developing countries. These two findings provide insights into how the grade repetition system may affect the achievement gaps within a class and the role of the preschool system in the accumulation of academic skills gaps caused by age differences.

correlated outcomes, and uses all available data by assigning less weight to missing values. See Schwab et al. (2020).

[28]These statistically insignificant results may be attributed to the limited number of countries included in the analysis, potentially leading to low statistical power.

First, the estimates presented in the second panel of Table 6 reveal that a country with a 1 percentage point higher repetition rate tends to have a smaller relative age effect by 0.012 standard deviations. The magnitude of the point estimates is similar between developing and developed countries for math, and the association is stronger in developing countries. This is also well illustrated by the negative slope of the predictive lines of relative age effects with the marginal change in repetition rates in Figure 4B. How can this finding be interpreted in terms of the school retention system? Previous studies about retention have shown mixed results (Eide and Showalter (2001)). For example, some studies show the negative effects of retention on outcomes, such as increased dropout rates, low self-esteem, and poor academic achievements (Rumberger (1987); Wilson (1990)), while others show that retention results in positive effects on academic outcomes (Hauser (2005); Pierson and Connell (1992)). In this analysis with cross-country data, the LATE of the relative age within a class does not include those who are the most vulnerable, such as low-achieving younger students since they become the oldest ones in the class when they repeat the grade. Therefore, the negative association between the repetition rate and the relative age effects suggests that a retention system may function as a positive system by decreasing the achievement gap due to relative age differences observed in the class, especially in a developing context.

Second, the substantial positive relationship between a country's preschool system index and relative age effects, presented in the fourth panel of Table 6 and most pronounced in developing countries and math, suggests that the mechanism that generates a gap between older and younger students in their academic achievement can begin before formal primary school system. This relationship is well demonstrated in Figure 4D, where many developing countries have lower indexes for preschool systems due to a lack of resources to establish a robust preschool system compared to the developed world.[29] The result can be interpreted as preschool serving as an accelerated educational system that plays a significant role in accumulating relative age gaps in academic achievements during primary

[29] As shown in Table A7, the percentage of students attending preschool in developing countries averages 79%, while in developed countries, it is 96%. Additionally, the average number of years attending preschool in developing countries is lower than in developed countries (1.78 years vs. 2.53 years).

school. This finding is consistent with previous research on the existence of relative age effects already present in the kindergarten system (Elder and Lubotsky (2009); Dobkin and Ferreira (2010); Cascio and Schanzenbach (2016)).

1.7 Conclusion

Students who enter school at different ages due to varying birth dates may experience differences in maturity compared to their peers in the same cohort, potentially influencing their academic outcomes. This phenomenon, known as "relative age effects," has been extensively studied in developed countries but has received limited attention in the context of developing countries. This study addresses this research gap by investigating the presence of relative age effects in developing countries. I utilize data from the Trends in International Mathematics and Science Study (TIMSS) and employ assigned relative age as an instrumental variable, which allows an investigation of a causal relationship between relative age and test scores in a developing context. The results reveal the presence and prevalence of positive relative age effects on test scores within developing countries, albeit of lesser magnitude when compared to their counterparts in developed nations. Furthermore, the study explores the relationship between relative age effects and educational characteristics across various countries.

Considering the presence of relative age effects in developing countries, proactive intervention strategies might be necessary to reduce the achievement gap within the cohort and support those who may struggle solely because of their birth month, which makes them the youngest in their class. Several studies have suggested various educational practices to reduce the disparity in academic achievement due to the age gap (Urruticoechea et al. (2021)). These practices include changing the grouping system at the time of entry, such as having multiple cutoffs within a year or allowing for mobility between the two groupings (Cordero (1985)), assessing students when they are exactly the same relative age, or standardizing test scores by relative age (Crawford et al. (2010)), and implementing educational strategies to avoid diminishing the self-esteem of students with low academic performance due to relative age (Ando et al. (2019)).

In developing countries, it may be challenging to establish a robust support system for students lagging behind due to the relative age gap since their education systems are often less established and resources are constrained. Nevertheless, it is crucial to acknowledge that addressing achievement disparities resulting from relative age gaps becomes even more important in developing contexts where countries experience higher dropout and school failure rates. Therefore, resolving these issues that may have been caused by relative age effects should be their primary and foremost educational goal as they pursue national growth.

1.8 Figures and Tables

Figure 1: Distribution of birth months and assigned relative age/observed age by month of birth in developing countries

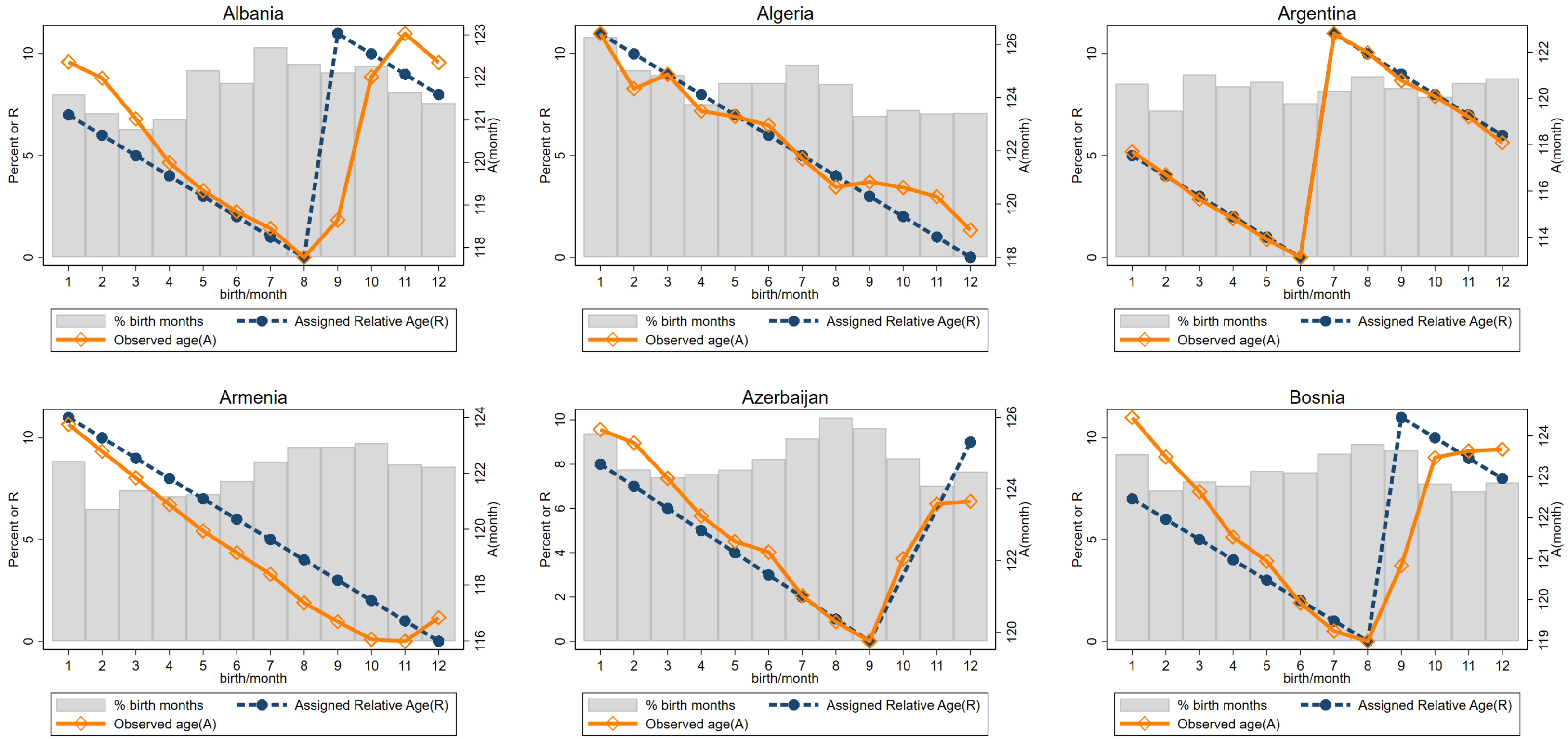


FIGURE 1: DISTRIBUTION OF BIRTH MONTHS AND ASSIGNED RELATIVE AGE BY MONTH OF BIRTH IN DEVELOPING COUNTRIES(CONTINUED)

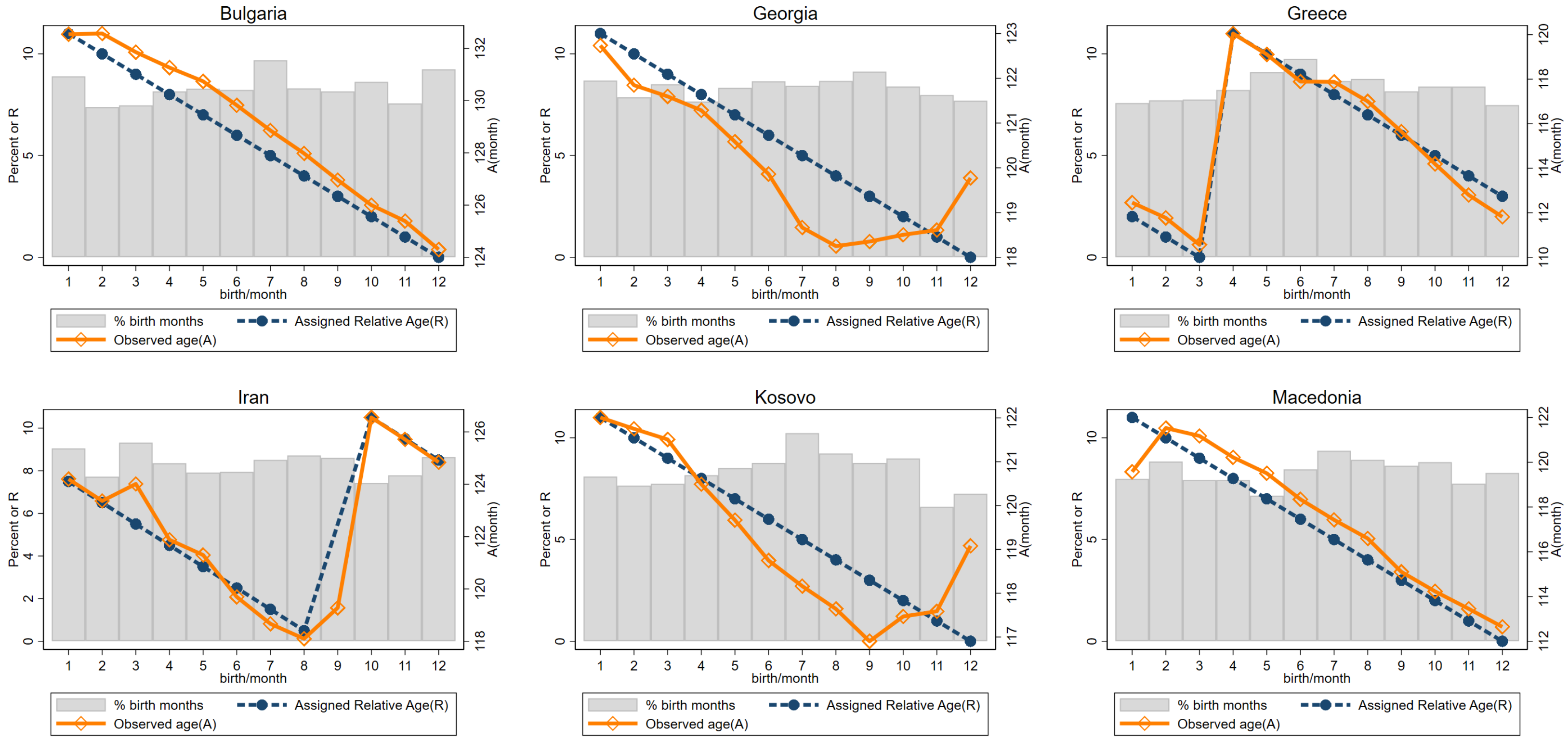
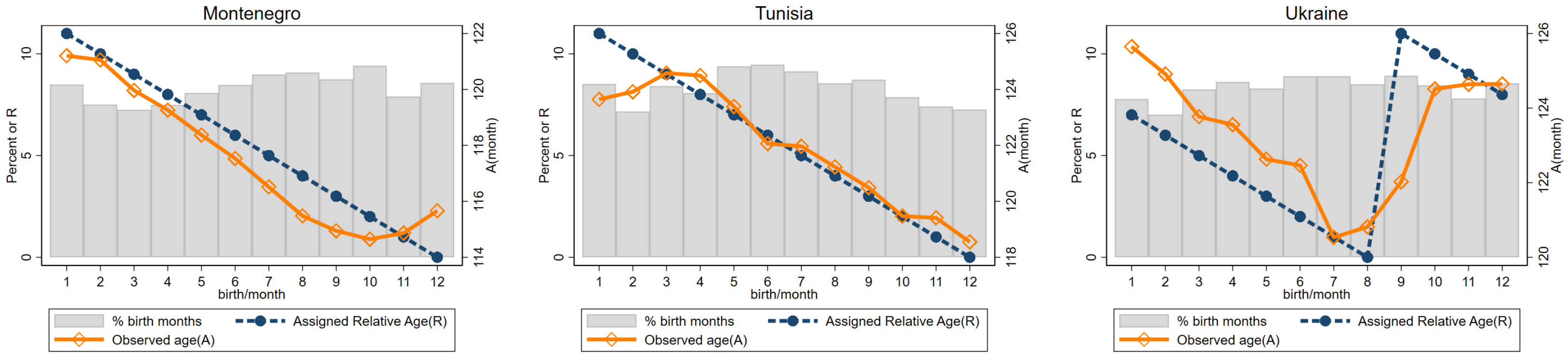
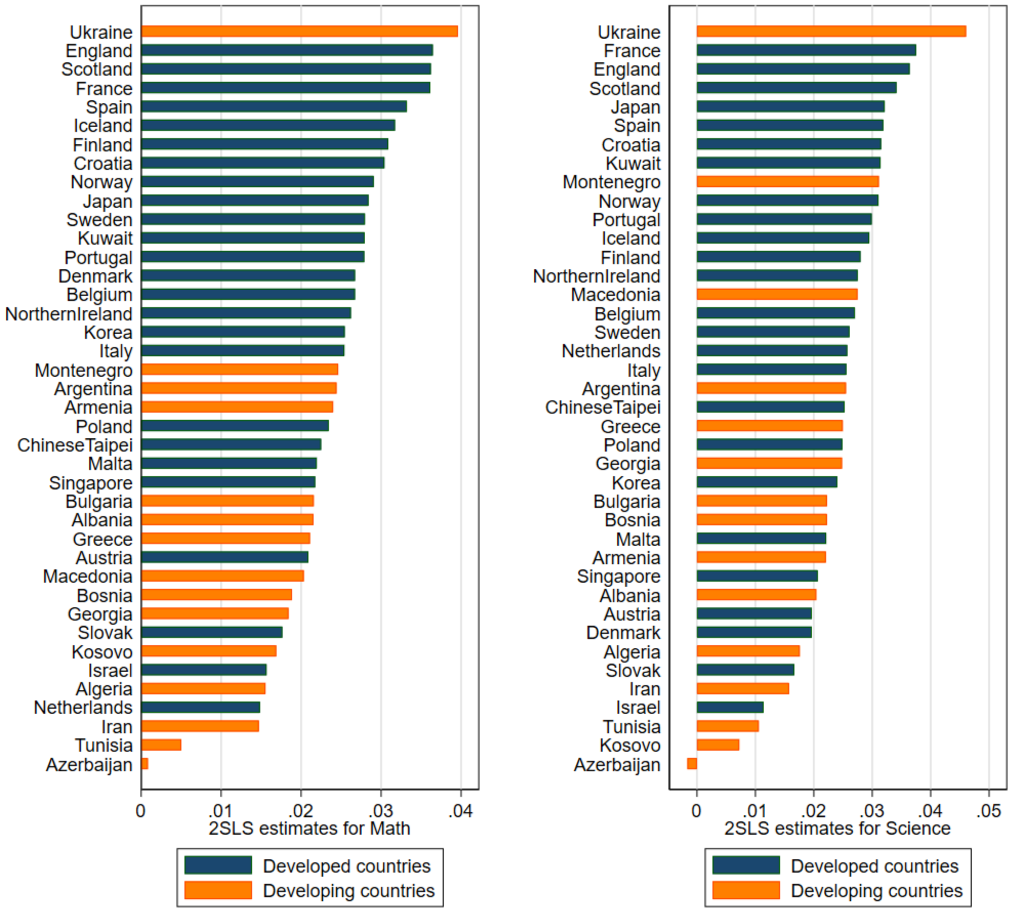


FIGURE 1: DISTRIBUTION OF BIRTH MONTHS AND ASSIGNED RELATIVE AGE BY MONTH OF BIRTH IN DEVELOPING COUNTRIES(CONTINUED)



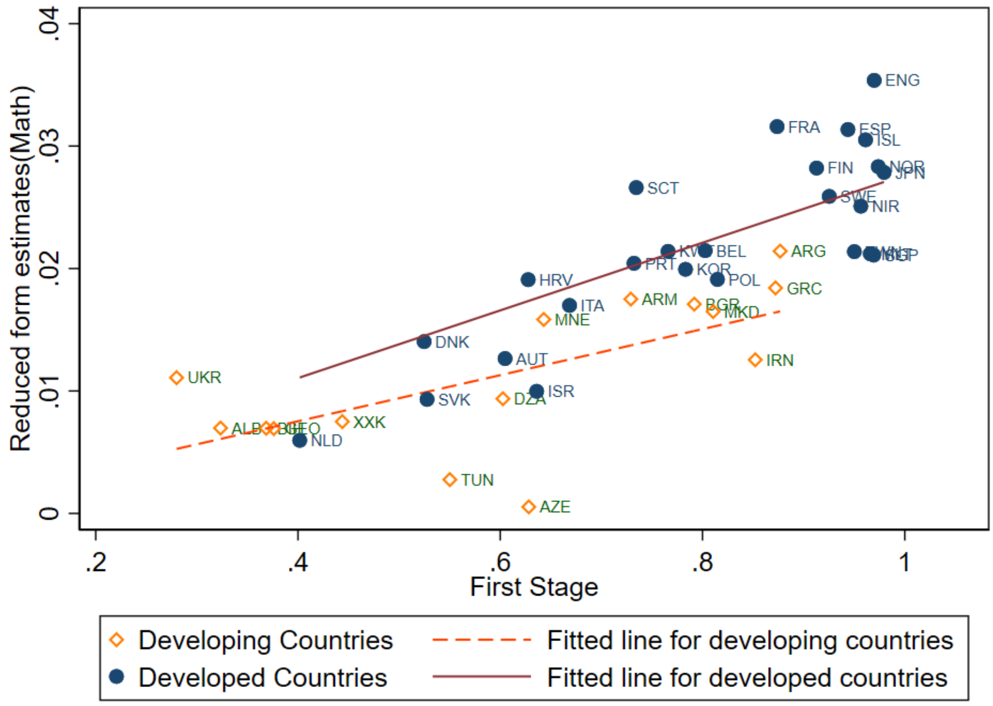
Notes: Source: TIMSS © IEA 1995-2019. Assigned relative age R_{it} is decided by each country's cutoff. A student who is the youngest kid in the cohort is assigned with $R = 0$, and who is the oldest kid is assigned with $R = 11$. Only grade 4 is included. Since Iran's school-entry cutoff is in the middle of September and Azerbaijan allows students born in October or November to enroll earlier by law, R is not assigned for these months for each country.

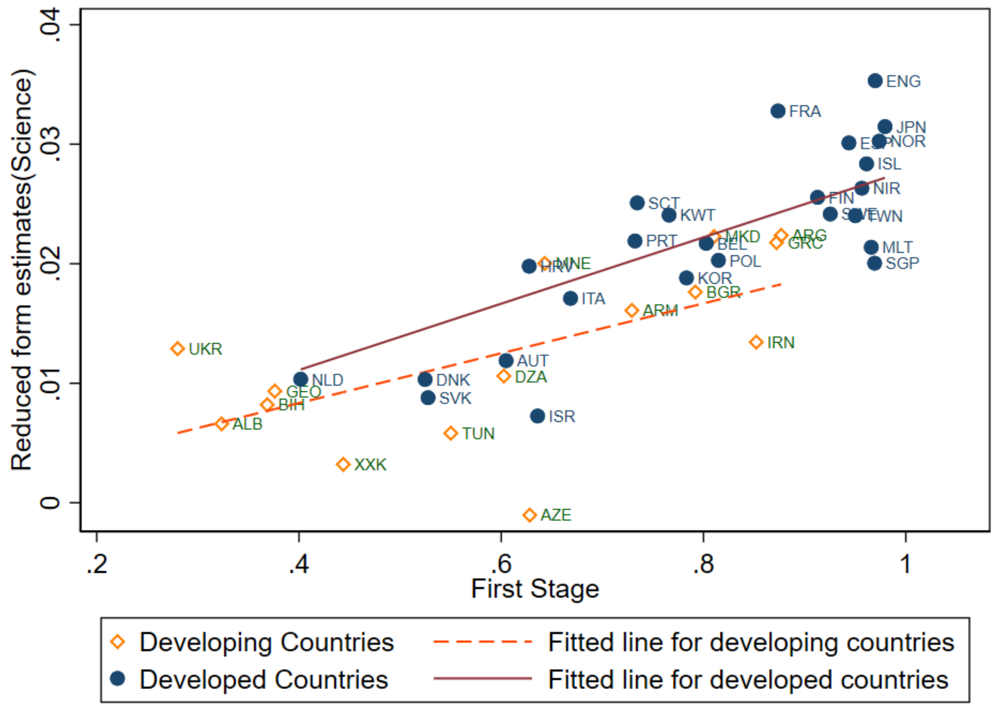
FIGURE 2: RELATIVE AGE EFFECTS FOR EACH COUNTRY



Notes: Source: TIMSS © IEA 1995-2019. The X-axis shows the 2SLS estimates for relative age effects. All regressions are population-weighted and include controls for students' gender, the index for books, the index for desks, the index for parental education, and year and grade fixed effects. Robust standard errors are clustered at the school level. Dashed bars denote the average relative age effects for developed and developing countries, respectively.

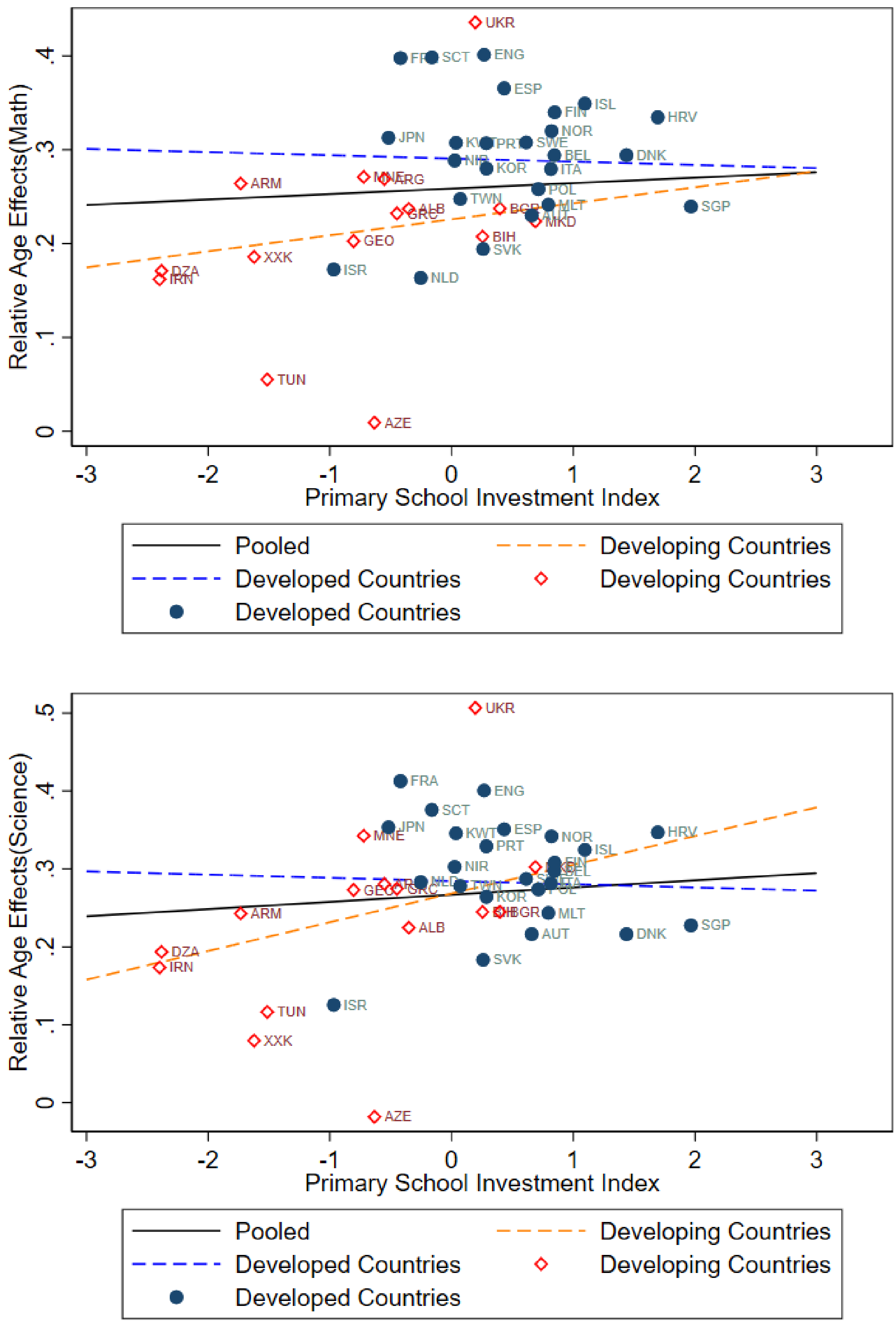
FIGURE 3: FIRST STAGE AND REDUCED FORM ESTIMATES FOR DEVELOPING AND DEVELOPED COUNTRIES





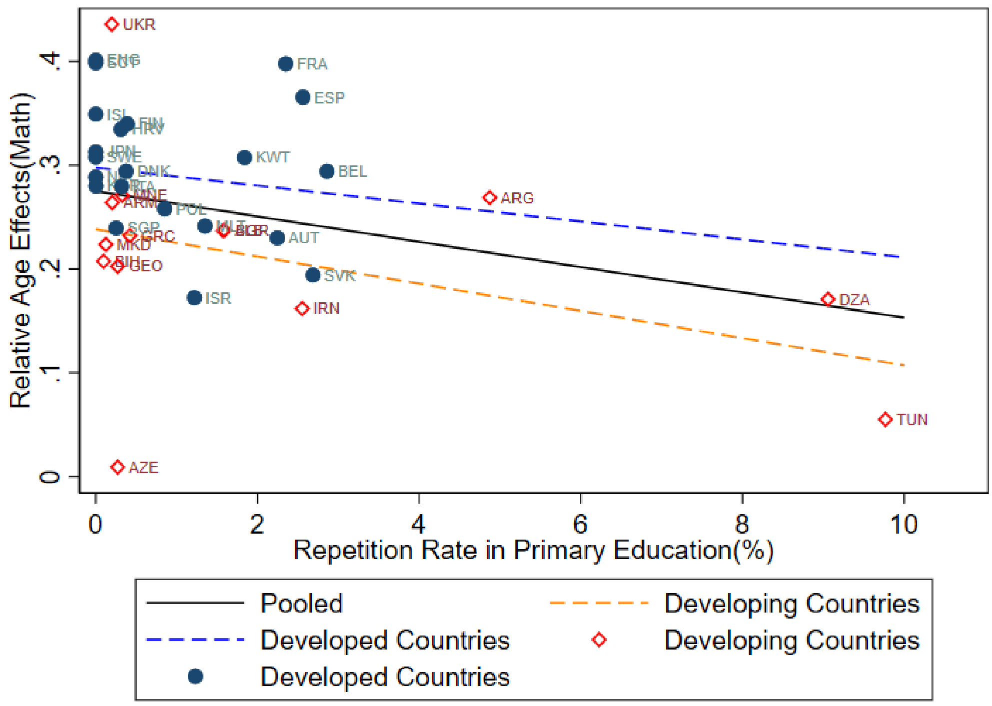
Notes: Source: TIMSS © IEA 1995-2019. The country abbreviation is displayed in Table 1. The slope of the line connecting the point from the origin shows the IV estimates. Fitted lines are drawn so that they pass through the origin, and the slope of each line shows the average relative age effects for developing and developed countries, respectively. IV estimates for each country are shown in Table 4A and Table 4B.

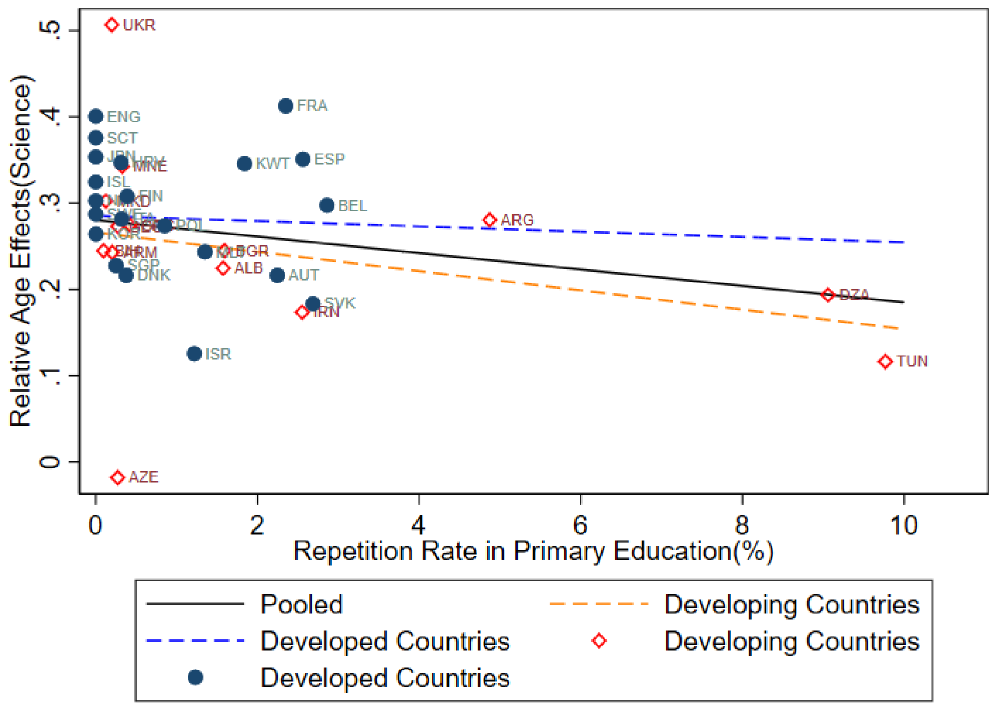
Figure 4A: Relative Age Effects and Primary School Investment Index



Notes: Source: TIMSS © IEA 1995-2019; World Bank EdStats. The country abbreviation is displayed in Table 1. Each line in the graph shows the predictive relative age effects by the marginal change in the primary school investment index when controlling other variables for the pooled, developing, and developed countries, respectively. Regression results are presented in Table 6.

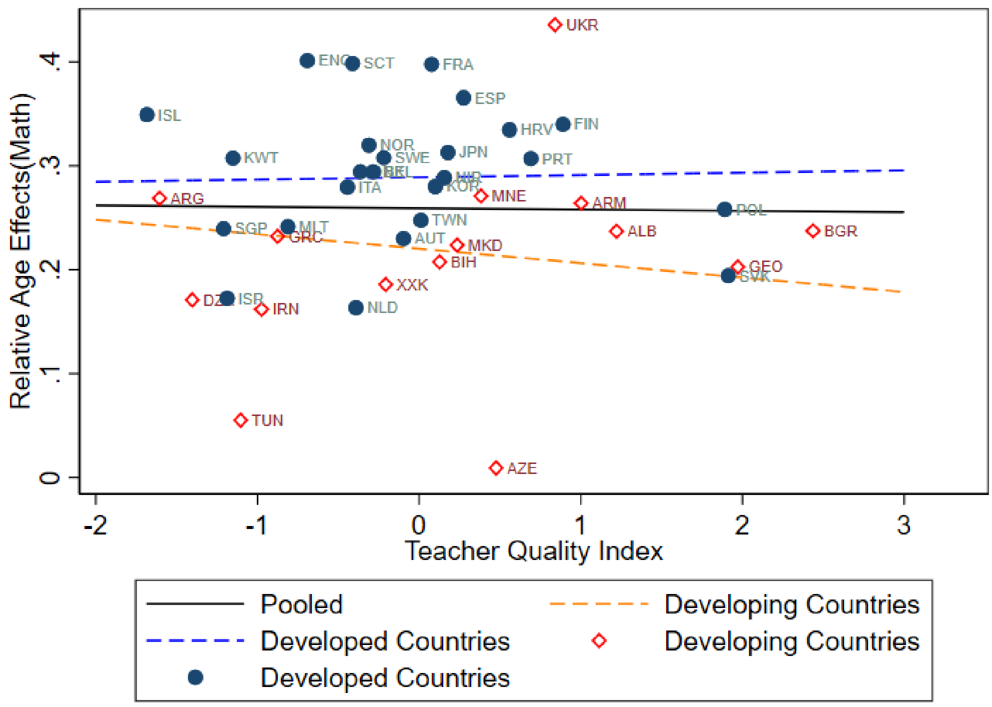
FIGURE 4B: RELATIVE AGE EFFECTS AND REPETITION RATE IN PRIMARY EDUCATION

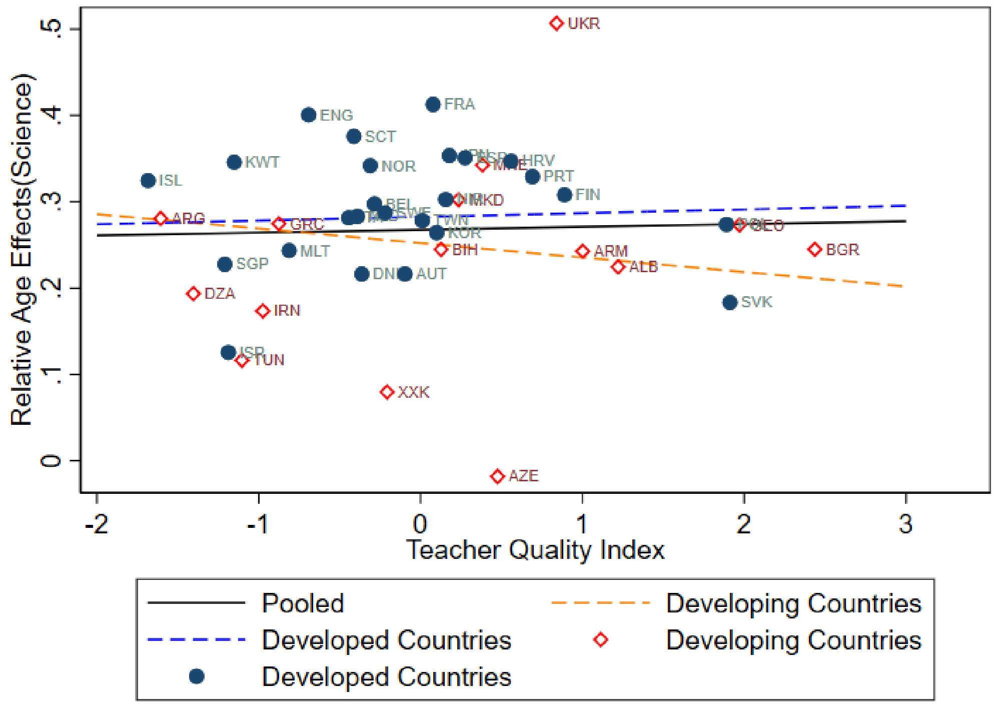




Notes: Source: TIMSS © IEA 1995-2019; World Bank EdStats. The country abbreviation is displayed in Table 1. Each line in the graph shows the predictive relative age effects by the marginal change in the repetition rate in primary education when controlling other variables for the pooled, developing, and developed countries, respectively. Regression results are presented in Table 6.

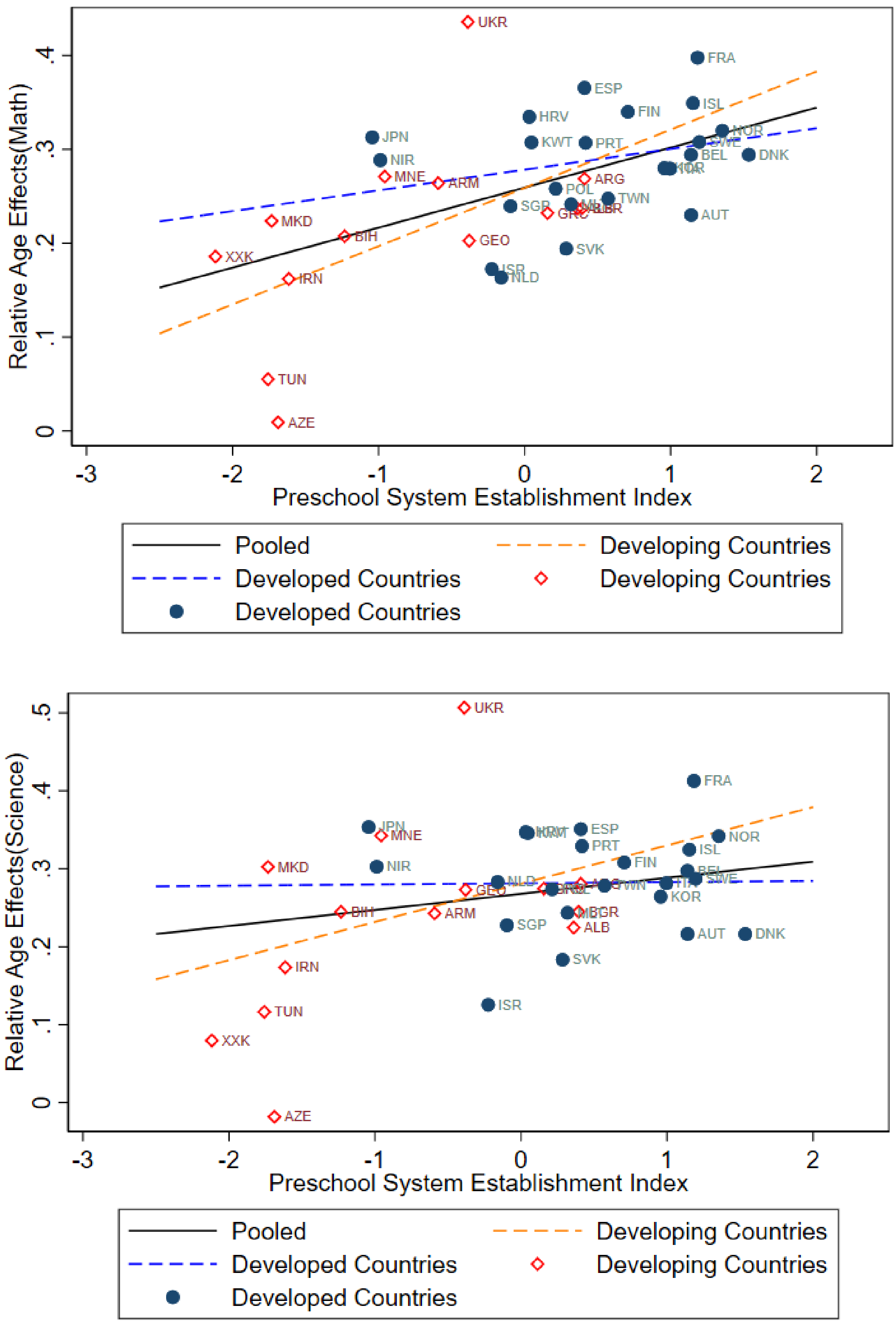
FIGURE 4C: RELATIVE AGE EFFECTS AND TEACHER QUALITY INDEX





Notes: Source: TIMSS © IEA 1995-2019; World Bank EdStats. The country abbreviation is displayed in Table 1. Each line in the graph shows the predictive relative age effects by the marginal change in the teacher quality index when controlling other variables for the pooled, developing, and developed countries, respectively. Regression results are presented in Table 6.

FIGURE 4D: RELATIVE AGE EFFECTS AND PRESCHOOL SYSTEM ESTABLISHMENT INDEX



Notes: Source: TIMSS © IEA 1995-2019; World Bank EdStats. The country abbreviation is displayed in Table 1. Each line in the graph shows the predictive relative age effects by the marginal change in the preschool system establishment index when controlling other variables for the pooled, developing, and developed countries, respectively. Regression results are presented in Table 6.

TABLE 1: LIST OF THE COMPLETE SAMPLE OF COUNTRIES IN THE TIMSS

	Country	**Abbrv.**	**Entry Cutoff**	**Test years**	**Reason for Exclusion**
Low Income (including Low(L), Lower-Middle(LM), Upper-Middle(UM))	Albania	ALB	September 1	2019	
	Algeria	DZA	January 1	2007	
	Argentina	ARG	July 1	2015	
	Armenia	ARM	January 1	2003-2019	exclude years before 2015 with empirical ≠ official
	Azerbaijan	AZE	October 1	2011, 2019	
	Bosnia and Herzegovina	BIH	September 1	2019	
	Botswana	BWA	July 1	2011	(2) empirical ≠ official
	Bulgaria	BGR	January 1	2015, 2019	
	Colombia	COL		2007	(1) cut off not reported
	El Salvador	SLV	June 1	2007	(2) empirical ≠ official
	Georgia	GEO	January 1	2007-2019	exclude 2019 with (4) cutoff changing
	Greece	GRC	April 1	1995	
	Honduras	HND	February 1	2011	(2) empirical ≠ official
	Indonesia	IDN	mid July	2015	(2) empirical ≠ official
	Iran	IRN	September 22	1995-2019	exclude 1995 with (2)
	Kazakhstan	KAZ		2007-2019	(1) cut off not reported
	Kosovo	XXK	January 1	2019	

Continued on next page

Table 1 – *Continued from previous page*

	Country	Abbrv.	Entry Cutoff	Test years	Reason for Exclusion
Low Income (including Low(L), Lower-Middle(LM), Upper-Middle(UM))	Moldova	MDA		2003	(1)cut off not reported
	Mongolia	MNG		2007	(1) cut off not reported
	Montenegro	MNE	January 1	2019	
	Morocco	MAR	September 1	2003-2019	(2) empirical ≠ official
	Northern Macedonia	MKD	January 1	2019	
	Pakistan	PAK	April 1	2019	(2) empirical ≠ official
	Philippines	PHL	September 1	2003, 2019	(2) empirical ≠ official
	Romania	ROU		2011	(2) empirical ≠ official
	Russia	RUS	September 1	2003-2019	(2) empirical ≠ official
	Serbia	SRB	September 1	2011-2019	(2) empirical ≠ official
	South Africa	ZAF	July 1	2019	(2) empirical ≠ official
	Thailand	THA	mid May	1995, 2011	(2) empirical ≠ official
	Tunisia	TUN	January 1	2003-2011	
	Ukraine	UKR	September 1	2007	
	Türkiye	TUR	January 1	2003-2011	(2) empirical ≠ official
	Yemen	YEM	October 1	2003-2011	(2) empirical ≠ official
High Income(H)	Australia	AUS		1995-2019	(3) multiple cutoffs
	Austria	AUT	September 1	1995, 2007, 2011, 2019	
	Bahrain	BHR	January 1	2011-2019	(2) empirical ≠ official
	Belgium	BEL	January 1	2003, 2011-2019	
	Canada	CAN		1995-2019	(3) multiple cutoffs
	Croatia	HRV	April 1	2011-2019	
	Denmark	DNK	January 1	2007-2019	
	England	ENG	September 1	1995-2019	
	Northern Ireland	NIR	July 1	2011-2019	

Continued on next page

Table 1 – *Continued from previous page*

	Country	Abbrv.	Entry Cutoff	Test years	Reason for Exclusion
	Scotland	SCT	March 1	1995-2007	
	Finland	FIN	January 1	2011-2019	
	France	FRA	January 1	2015, 2019	
	Germany	DEU		2007-2019	(3) multiple cutoffs
	Hong Kong	HKG	September 1	1995-2019	(2) empirical ≠ official
	Iceland	ISL	January 1	1995	
	Ireland	IRL		1995, 2011-2019	(1) cut off not reported
	Israel	ISR	January 1	1995	
	Italy	ITA	January 1	2003-2019	
	Japan	JPN	April 1	1995-2019	
	Korea	KOR	March 1(until 2011), January 1(since 2015)	1995, 2011-2019	
High Income(H)	Kuwait	KWT	March 15	1995, 2007-2019	exclude 1995 with (2)
	Malta	MLT	January 1	2011, 2019	
	Netherlands	NLD	October 1	1995-2019	
	New Zealand	NZL		1995-2019	(3) multiple cutoffs
	Norway	NOR	January 1	1995-2019	
	Oman	OMN	September 1	2011-2019	(2) empirical ≠ official
	Poland	POL	January 1	2011-2019	(4) exclude 2019 with cutoff changing
	Portugal	PRT	January 1	1995, 2011-2019	
	Qatar	QAT		2007-2019	(1) cut off not reported

Continued on next page

Table 1 – *Continued from previous page*

	Country	Abbrv.	Entry Cutoff	Test years	Reason for Exclusion
	Saudi Arabia	SAU		2011-2019	(1) cut off not reported
	Slovak Republic	SVK	September 1	2007-2019	
	Singapore	SGP	January 1	1995-2019	
	Spain	ESP	January 1	2011-2019	
	Sweden	SWE	January 1	2003-2019	
	Chinese Taipei	TWN	September 1	2003-2019	
	United Arab Emirates	ARE	January 1	2007-2019	(2) empirical ≠ official
	United States	USA		1995-2019	(3) multiple cutoffs
Low to High	Chile	CHL	April 1	2011-2019	
	Czech Republic	CZE	September 1	1995, 2007-2019	
	Hungary	HUN	June 1(until 2015), September 1(since 2019)	1995-2019	
	Latvia	LVA	January 1(since 2003)	1995-2007, 2019	exclude 1995 with (1)
	Lithuania	LTU	January 1(since 2003)	2003-2019	
	Slovenia	SVN	January 1	1995-2019	

Notes: Source: TIMSS © IEA 1995-2019. The first column shows the development classification in 2019 following World Bank Analytical Classifications. The fourth column shows the classification in 1995 and it is highlighted when the development classification has been changed. L means low income countries. In the seventh column, the reasons for exclusion are (1) when there is no or inconsistent cutoff information, (2) when the empirical cutoff does not match with the informed cutoff, (3) when the cutoff varies depending on state or territory, and (4) when the country was in the middle of entry cutoff change. UK's data are separately collected in England, Northern Ireland, and Scotland since each part has a different school entry cutoff. The TIMSS doesn't include the data for Wales.

TABLE 2: SUMMARY STATISTICS

	Math	Science	Age	Female(%)	Low SES(%)	Desk(%)	Book	Parental Education	Entrance Age	Sample Size
Developing Countries										
Albania	499.55	495.77	10.03	48.70	73.67	81.28	1.15	2.05	6.28	4630
	(85.10)	(84.16)	(0.44)	(49.99)	(44.05)	(39.01)	(1.13)	(1.35)	(0.50)	
Algeria	378.10	354.98	10.22	50.15	94.31	59.20	0.72			4211
	(86.65)	(97.87)	(0.90)	(50.01)	(23.17)	(49.15)	(0.96)			
Argentina	421.25	423.12	9.83	49.48	76.72	61.50	1.55	3.44	5.72	3104
	(79.57)	(87.01)	(0.46)	(50.01)	(42.27)	(48.67)	(1.20)	(0.91)	(0.51)	
Armenia	490.68	456.98	9.91	48.03	58.23	62.50	1.73	2.90	6.03	15167
	(73.78)	(81.77)	(0.37)	(49.96)	(49.32)	(48.41)	(1.25)	(1.03)	(0.43)	
Azerbaijan	488.69	434.89	10.22	46.38	65.08	68.71	1.05	2.73	6.25	13155
	(96.15)	(95.57)	(0.50)	(49.87)	(47.67)	(46.37)	(1.03)	(1.12)	(0.53)	
Bosnia	452.98	461.44	10.15	48.89	71.40	86.31	1.16	2.46	6.12	5615
	(74.77)	(77.14)	(0.36)	(49.99)	(45.19)	(34.37)	(1.10)	(0.96)	(0.43)	
Bulgaria	528.19	540.75	10.75	48.89	53.66	88.20	1.55	2.75	6.84	9612
	(81.34)	(94.56)	(0.41)	(49.99)	(49.87)	(32.27)	(1.23)	(1.26)	(0.41)	
Georgia	449.17	441.61	10.00	48.13	53.82	80.20	2.06	3.16	5.90	19506
	(85.50)	(84.16)	(0.47)	(49.97)	(49.86)	(39.85)	(1.31)	(0.89)	(0.51)	
Greece	432.69	450.64	9.12	50.29	71.64	86.14	1.94			5848

Continued on next page

Table 2 – *Continued from previous page*

	Math	**Science**	**Age**	**Female(%)**	**Low SES(%)**	**Desk(%)**	**Book**	**Parental Education**	**Entrance Age**	**Sample Size**
	(101.46)	(90.36)	(0.62)	(50.00)	(45.08)	(34.55)	(1.18)			
Iran	431.10	447.62	10.19	46.99	77.51	59.97	1.14	1.91	6.69	23759
	(92.72)	(97.90)	(0.58)	(49.91)	(41.75)	(49.00)	(1.23)	(1.37)	(0.54)	
Kosovo	448.41	420.24	9.93	49.04	73.72	72.84	1.19	2.24	6.03	4496
	(80.41)	(87.03)	(0.41)	(50.00)	(44.02)	(44.49)	(1.06)	(1.24)	(0.49)	
Macedonia	472.85	428.90	9.79	48.12	64.51	79.44	1.30	2.36	5.92	3396
	(98.71)	(102.67)	(0.38)	(49.97)	(47.86)	(40.42)	(1.10)	(1.42)	(0.44)	
Montenegro	452.12	454.47	9.78	46.56	67.36	80.78	1.49	2.91	5.93	5075
	(85.84)	(88.88)	(0.37)	(49.89)	(46.90)	(39.40)	(1.19)	(0.90)	(0.43)	
Tunisia	349.13	336.51	10.17	47.57	87.89	74.02	1.10			16961
	(104.69)	(128.72)	(0.78)	(49.94)	(32.62)	(43.85)	(1.14)			
Ukraine	477.33	482.44	10.27	48.95	76.53	88.76	1.85			4292
	(82.42)	(80.46)	(0.47)	(49.99)	(42.39)	(31.59)	(1.08)			
Average&Total N	447.50	438.38	10.08	47.97	69.17	73.10	1.42	2.58	6.27	138,827
	(101.55)	(107.85)	(0.62)	(49.96)	(46.18)	(44.34)	(1.24)	(1.23)	(0.59)	
Developed Countries										
Austria	510.88	521.00	10.23	48.72	63.63	92.07	1.99	2.70	6.17	20245
	(74.20)	(80.37)	(0.56)	(49.98)	(48.11)	(27.02)	(1.21)	(0.95)	(0.45)	
Belgium	545.22	511.65	10.04	50.37	59.08	88.48	1.92	3.23	5.82	26667
	(62.07)	(59.69)	(0.52)	(50.00)	(49.17)	(31.93)	(1.09)	(1.05)	(0.51)	
Chinese Taipei	587.90	557.56	10.23	48.04	59.30	83.97	1.88	2.88	6.92	41782

Continued on next page

Table 2 – *Continued from previous page*

	Math	Science	Age	Female(%)	Low SES(%)	Desk(%)	Book	Parental Education	Entrance Age	Sample Size
	(69.05)	(70.06)	(0.31)	(49.96)	(49.13)	(36.69)	(1.27)	(1.01)	(0.73)	
Croatia	501.93	527.16	10.60	49.54	65.96	93.39	1.54	2.82	6.66	12354
	(67.57)	(60.91)	(0.34)	(50.00)	(47.39)	(24.85)	(1.09)	(0.88)	(0.49)	
Denmark	531.75	524.17	10.94	50.46	55.17	90.75	1.98	3.53	5.77	22019
	(72.96)	(71.53)	(0.39)	(50.00)	(49.73)	(28.98)	(1.13)	(0.78)	(0.51)	
England	523.91	528.55	10.06	50.18	61.60	76.80	2.18			32031
	(97.11)	(83.81)	(0.47)	(50.00)	(48.64)	(42.21)	(1.23)			
Finland	536.02	558.20	10.79	48.71	38.39	90.00	2.22	3.22	6.72	17037
	(70.41)	(66.84)	(0.34)	(49.98)	(48.64)	(29.99)	(1.06)	(0.91)	(0.48)	
France	484.55	486.49	9.88	49.09	54.15	87.36	1.97	2.86	5.85	10989
	(77.83)	(76.42)	(0.39)	(49.99)	(49.83)	(33.23)	(1.19)	(1.04)	(0.43)	
Iceland	416.34	443.22	9.16	51.06	45.53	89.05	2.69			3825
	(85.16)	(98.81)	(0.58)	(50.00)	(49.81)	(31.24)	(1.11)			
Israel	512.83	492.45	10.05	50.74	62.21	95.24	2.26			2822
	(81.91)	(88.81)	(0.42)	(50.00)	(48.50)	(21.30)	(1.12)			
Italy	507.16	520.83	9.75	49.21	70.35	70.93	1.68	2.35	5.84	23656
	(74.45)	(74.95)	(0.36)	(49.99)	(45.67)	(45.41)	(1.18)	(1.08)	(0.44)	
Japan	575.49	553.09	10.33	49.53	65.22	87.35	1.74	3.15	6.02	42707
	(75.88)	(71.12)	(0.43)	(50.00)	(47.63)	(33.24)	(1.08)	(0.85)	(0.31)	
Korea	595.17	581.51	10.27	48.20	30.40	91.10	2.82	3.37	6.86	21231
	(71.19)	(68.02)	(0.52)	(49.97)	(46.00)	(28.48)	(1.19)	(0.82)	(0.53)	
Kuwait	345.64	356.42	9.82	52.15	61.96	69.98	1.43	3.20	5.83	32271

Continued on next page

Table 2 – *Continued from previous page*

	Math	Science	Age	Female(%)	Low SES(%)	Desk(%)	Book	Parental Education	Entrance Age	Sample Size
	(106.56)	(126.13)	(0.51)	(49.95)	(48.55)	(45.83)	(1.28)	(1.03)	(0.55)	
Malta	500.18	465.72	9.80	47.12	62.02	83.15	2.00	2.19	5.10	10520
	(77.90)	(96.26)	(0.40)	(49.92)	(48.54)	(37.43)	(1.12)	(1.25)	(0.34)	
Netherlands	532.30	521.43	10.05	49.86	68.26	92.50	2.02	3.26	5.83	22582
	(62.42)	(60.18)	(0.56)	(50.00)	(46.55)	(26.34)	(1.13)	(0.91)	(0.49)	
Northern Ireland	569.72	521.54	10.42	49.09	56.80	76.59	2.03	2.78	5.02	10333
	(85.21)	(69.31)	(0.32)	(49.99)	(49.54)	(42.35)	(1.18)	(1.24)	(0.19)	
Norway	497.48	501.47	10.04	49.25	50.65	88.29	2.19	3.48	5.78	37779
	(86.92)	(81.47)	(0.62)	(50.00)	(50.00)	(32.15)	(1.16)	(0.80)	(0.45)	
Poland	519.59	536.42	10.41	49.28	49.82	79.62	1.97	2.69	6.48	14466
	(76.72)	(73.22)	(0.57)	(50.00)	(50.00)	(40.28)	(1.14)	(1.25)	(0.58)	
Portugal	498.54	485.70	9.91	48.74	66.08	82.25	1.65	2.33	5.83	18521
	(95.93)	(88.29)	(0.73)	(49.99)	(47.34)	(38.21)	(1.16)	(1.38)	(0.49)	
Scotland	480.16	495.31	9.52	50.56	57.92	76.81	2.27			15820
	(88.48)	(88.79)	(0.53)	(50.00)	(49.37)	(42.20)	(1.28)			
Singapore	589.63	551.05	10.21	49.02	57.29	86.53	2.04	3.07	6.65	66827
	(92.52)	(104.72)	(0.57)	(49.99)	(49.47)	(34.14)	(1.18)	(1.08)	(0.54)	
Slovak	505.15	529.21	10.41	48.83	60.83	82.64	1.88	2.68	6.25	27823
	(79.56)	(82.12)	(0.54)	(49.99)	(48.81)	(37.88)	(1.14)	(1.03)	(0.48)	
Spain	513.51	524.32	9.90	48.65	44.62	88.02	2.07	2.87	5.60	34747
	(68.75)	(65.18)	(0.39)	(49.98)	(49.71)	(32.47)	(1.18)	(1.27)	(0.53)	
Sweden	511.11	533.99	10.79	49.21	49.85	95.96	2.15	3.19	6.68	23813

Continued on next page

Table 2 – *Continued from previous page*

	Math	**Science**	**Age**	**Female(%)**	**Low SES(%)**	**Desk(%)**	**Book**	**Parental Education**	**Entrance Age**	**Sample Size**
	(69.77)	(74.55)	(0.33)	(49.99)	(50.00)	(19.69)	(1.18)	(0.98)	(0.52)	
Average&Total N	525.88	519.99	10.18	49.36	56.93	85.12	1.99	2.94	6.19	592,845
	(98.67)	(94.28)	(0.59)	(50.00)	(49.52)	(35.59)	(1.20)	(1.11)	(0.72)	

Notes: Source: TIMSS © IEA 1995-2019. Standard errors in parentheses. Mean values in each country. Developing countries are mapped to Low Income countries in Table 1 after exclusion. Developed countries are mapped to High Income countries in Table 1 after exclusion. High SES is defined as those with either parental education is at least a bachelor's degree holder or the number of books in the household is more or equal to 100. Low SES is defined as non-high SES. The index for books in the household is 0 for 0-10 books, 1 for 11-25 books, 2 for 26-100 books, 3 for 101-200 books, and 4 for more than 200 books. The index for parental education level is 0 for some primary, lower secondary, or no school, 1 with lower secondary, 2 with upper secondary, 3 with post-secondary but not the university, and 4 with university or higher. The entrance age is the average primary school entrance age for existing data points. UK's data are separately collected in England, Northern Ireland, and Scotland since each part has a different school entry cutoff. The TIMSS doesn't include the data for Wales.

TABLE 3: FIRST STAGE

Developing	**FS**	**F-statistics**
Albania	0.323*	113
	(0.030)	
Algeria	0.603*	137
	(0.051)	
Argentina	0.877*	897
	(0.029)	
Armenia	0.729*	1780
	(0.017)	
Azerbaijan	0.628*	730
	(0.023)	
Bosnia	0.369*	218
	(0.025)	
Bulgaria	0.792*	2778
	(0.015)	
Georgia	0.376*	384
	(0.019)	
Greece	0.872*	2495
	(0.017)	
Iran	0.852*	3739
	(0.014)	
Kosovo	0.444*	256
	(0.028)	
Macedonia	0.810*	747
	(0.030)	
Montenegro	0.643*	875
	(0.022)	
Tunisia	0.550*	390
	(0.028)	
Ukraine	0.280*	106
	(0.027)	

Developing	**FS**	**F-statistics**
Austria	0.605*	1857
	(0.014)	
Belgium	0.803*	3061
	(0.015)	
ChineseTaipei	0.950*	45058
	(0.004)	
Croatia	0.628*	2157
	(0.014)	
Denmark	0.525*	1499
	(0.014)	
England	0.970*	74027
	(0.004)	
Finland	0.913*	11236
	(0.009)	
France	0.874*	4613
	(0.013)	
Iceland	0.961*	6934
	(0.012)	
Israel	0.636*	231
	(0.042)	
Italy	0.669*	2337
	(0.014)	
Japan	0.979*	156246
	(0.002)	
Korea	0.783*	3950
	(0.012)	
Kuwait	0.766*	970
	(0.025)	
Malta	0.966*	5491
	(0.013)	
Netherlands	0.402*	715
	(0.015)	
NorthernIreland	0.957*	25890
	(0.006)	
Norway	0.974*	105981
	(0.003)	
Poland	0.815*	2597
	(0.016)	
Portugal	0.732*	1604
	(0.018)	
Scotland	0.734*	2194
	(0.016)	
Singapore	0.969*	25973
	(0.006)	
Slovak	0.528*	1476
	(0.014)	
Spain	0.944*	13982
	(0.008)	
Sweden	0.925*	17238
	(0.007)	

Robust standard errors in parentheses.
* $p < 0.05$

Note: Source: TIMSS © IEA 1995-2019. The second column shows the first stage estimates for each country. All regressions are population-weighted and include controls for students' gender, the index for books, index for desk, the index for parental education, and year and grade fixed effects. Robust standard errors are clustered at school level.

TABLE 4A: RELATIVE AGE EFFECTS ON MATH/SCIENCE TEST SCORES IN DEVELOPING COUNTRIES

	Math		Science	
	(1)	(2)	(3)	(4)
Country	OLS	2SLS	OLS	2SLS
Albania	0.001	0.022*	0.002	0.020*
	(0.003)	(0.012)	(0.003)	(0.012)
Algeria	-0.012***	0.016**	-0.017***	0.018***
	(0.003)	(0.006)	(0.002)	(0.007)
Argentina	-0.008*	0.024***	-0.008*	0.026***
	(0.005)	(0.005)	(0.004)	(0.005)
Armenia	0.008***	0.024***	0.007**	0.022***
	(0.002)	(0.004)	(0.003)	(0.004)
Azerbaijan	-0.002	0.001	-0.001	-0.002
	(0.002)	(0.006)	(0.002)	(0.005)
Bosnia	0.009***	0.019**	0.011***	0.022**
	(0.003)	(0.008)	(0.003)	(0.009)
Bulgaria	0.007**	0.022***	0.009***	0.022***
	(0.003)	(0.004)	(0.003)	(0.003)
Georgia	0.003	0.018***	0.005**	0.025***
	(0.002)	(0.006)	(0.002)	(0.007)
Greece	0.007**	0.021***	0.007***	0.025***
	(0.003)	(0.004)	(0.003)	(0.004)
Iran	-0.012***	0.015***	-0.012***	0.016***
	(0.001)	(0.003)	(0.001)	(0.003)
Kosovo	0.002	0.017*	0.003	0.007
	(0.003)	(0.009)	(0.003)	(0.009)
Macedonia	0.001	0.020***	0.008**	0.027***
	(0.004)	(0.005)	(0.004)	(0.005)
Montenegro	0.000	0.025***	-0.002	0.031***
	(0.004)	(0.006)	(0.005)	(0.005)
Tunisia	-0.029***	0.005	-0.032***	0.011**
	(0.001)	(0.004)	(0.001)	(0.005)
Ukraine	0.002	0.040***	0.004	0.046***
	(0.003)	(0.015)	(0.003)	(0.014)
Average		0.017***		0.019***
		(0.001)		(0.001)
N		134,770		134,770

Robust standard errors in parentheses and clustered at the school level.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: Source: TIMSS © IEA 1995-2019. All regressions are population-weighted and include controls for students' gender, the index for books, the index for desks, the index for parental education, and year and grade fixed effects. The data points for students born in September in Iran and October/November in Azerbaijan are excluded from the 2SLS analysis as they are not assigned with IV values. Estimates in average are calculated from equation (3) using the entire data for developing countries with country indicators.

TABLE 4B: RELATIVE AGE EFFECT ON MATH/SCIENCE TEST SCORES IN DEVELOPED COUNTRIES

	Math		Science	
	(1)	(2)	(3)	(4)
Country	OLS	2SLS	OLS	2SLS
Austria	-0.024***	0.021***	-0.021***	0.020***
	(0.001)	(0.004)	(0.001)	(0.004)
Belgium	-0.030***	0.027***	-0.019***	0.027***
	(0.002)	(0.003)	(0.002)	(0.003)
ChineseTaipei	0.019***	0.023***	0.021***	0.025***
	(0.002)	(0.002)	(0.002)	(0.002)
Croatia	0.005**	0.030***	0.005**	0.032***
	(0.002)	(0.004)	(0.002)	(0.004)
Denmark	-0.008***	0.027***	-0.003*	0.020***
	(0.002)	(0.005)	(0.002)	(0.005)
England	0.030***	0.036***	0.031***	0.036***
	(0.002)	(0.002)	(0.002)	(0.002)
Finland	-0.005**	0.031***	-0.004*	0.028***
	(0.003)	(0.003)	(0.003)	(0.003)
France	-0.011***	0.036***	-0.005**	0.038***
	(0.002)	(0.003)	(0.002)	(0.003)
Iceland	0.020***	0.032***	0.018***	0.030***
	(0.004)	(0.004)	(0.005)	(0.005)
Israel	-0.020***	0.016	-0.024***	0.011
	(0.005)	(0.013)	(0.004)	(0.013)
Italy	0.015***	0.025***	0.014***	0.026***
	(0.002)	(0.003)	(0.002)	(0.003)
Japan	0.025***	0.028***	0.029***	0.032***
	(0.002)	(0.002)	(0.002)	(0.002)
Korea	0.010***	0.025***	0.011***	0.024***
	(0.002)	(0.003)	(0.002)	(0.003)
Kuwait	0.001	0.028***	0.002	0.031***
	(0.002)	(0.004)	(0.002)	(0.004)
Malta	0.002	0.022***	0.005	0.022***
	(0.003)	(0.004)	(0.003)	(0.004)
Netherlands	-0.025***	0.015***	-0.015***	0.026***
	(0.002)	(0.005)	(0.002)	(0.005)
NorthernIreland	0.022***	0.026***	0.022***	0.028***
	(0.003)	(0.003)	(0.003)	(0.003)
Norway	0.023***	0.029***	0.025***	0.031***
	(0.002)	(0.002)	(0.002)	(0.002)
Poland	-0.001	0.023***	0.009***	0.025***
	(0.002)	(0.003)	(0.002)	(0.003)
Portugal	-0.017***	0.028***	-0.012***	0.030***

Continued on next page

Table 4B – *Continued from previous page*

	Math		Science	
	(1)	(2)	(3)	(4)
Country	OLS	2SLS	OLS	2SLS
	(0.001)	(0.003)	(0.001)	(0.003)
Scotland	0.017***	0.036***	0.016***	0.034***
	(0.003)	(0.004)	(0.003)	(0.004)
Singapore	0.014***	0.022***	0.007***	0.021***
	(0.001)	(0.001)	(0.001)	(0.001)
Slovak	-0.025***	0.018***	-0.027***	0.017***
	(0.003)	(0.004)	(0.003)	(0.003)
Spain	-0.009***	0.033***	-0.007***	0.032***
	(0.002)	(0.002)	(0.002)	(0.002)
Sweden	0.009***	0.028***	0.005**	0.026***
	(0.002)	(0.003)	(0.002)	(0.002)
Average		0.028***		0.028***
		(0.001)		(0.001)
N		590,197		590,197

Robust standard errors in parentheses and clustered at the school level.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: Source: TIMSS © IEA 1995-2019. All regressions are population-weighted and include controls for students' gender, the index for books, the index for desks, the index for parental education, and year and grade fixed effects. Estimates in average are calculated from equation (3) using the entire data for developed countries with country indicators.

TABLE 5: 2SLS RESULTS USING POOLED SAMPLES

	Math				**Science**			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Developing	0.017***	0.017***	0.019***	0.021***	0.019***	0.019***	0.021***	0.022***
	(0.001)	(0.001)	(0.002)	(0.003)	(0.001)	(0.001)	(0.002)	(0.003)
F-statistics	8659	8774	5881	2251	8659	8774	5881	2251
Observation	134770	134770	134770	81142	134770	134770	134770	81142
Developed	0.028***	0.027***	0.027***	0.031***	0.028***	0.027***	0.028***	0.031***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
F-statistics	89349	91272	58117	15570	89349	91272	58117	15570
Observation	590197	590197	590197	290135	590197	590197	590197	290135
Difference	-0.011***	-0.009***	-0.008***	-0.008***	-0.009***	-0.007***	-0.006***	-0.006**
	(0.002)	(0.001)	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)	(0.002)
F-statistics	53333	53911	33388	9682	53333	53911	33388	9682
Observation	724967	724967	724967	371277	724967	724967	724967	371277
Controls		✓	✓	✓		✓	✓	✓
Month of Birth			✓	✓			✓	✓
School Entry Age				✓				✓

Robust standard errors in parentheses and clustered at the school level.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: Source: TIMSS © IEA 1995-2019. The estimates show the results from 2SLS regression using assigned relative age as an instrument. All regressions are population-weighted and include year, grade, and country fixed effects. Columns (1) and (5) don't have any control over the child's gender and socioeconomic background, such as an index for books, parental education, and desk. Columns (2) and (6) have basic controls for gender and socioeconomic status. Columns (3) and (7) add controls for the month of birth to control the potential seasonality. Columns (4) and (8) add controls for the primary school entry age. The primary school entry age variables are dummy variables divided into '5-year-olds or younger', '6-year-olds', '7-year-olds', and '8-year-olds or older'. Since some countries do not have information about the school entry age range, the number of samples decreases in this specification. F-statistics are from the weak instrument test in the first stage.

TABLE 6: ASSOCIATION BETWEEN RELATIVE AGE EFFECTS AND EDUCATIONAL CHARACTERISTICS OF COUNTRIES IN MULTIVARIATE REGRESSION

	Math			**Science**		
	Pooled	Developing	Developed	Pooled	Developing	Developed
Index for educational investment in primary school	0.006	0.017	-0.003	0.009	0.037	-0.004
	(0.012)	(0.022)	(0.027)	(0.015)	(0.024)	(0.038)
Repetition rate in Primary School(%)	-0.012**	-0.013*	-0.009	-0.010	-0.011	-0.003
	(0.005)	(0.006)	(0.014)	(0.006)	(0.008)	(0.015)
Index for teacher quality	-0.001	-0.014	0.002	0.003	-0.017	0.004
	(0.009)	(0.016)	(0.016)	(0.011)	(0.017)	(0.021)
Index for preschool system establishment	0.043**	0.062*	0.022	0.021	0.049	0.002
	(0.017)	(0.030)	(0.019)	(0.023)	(0.041)	(0.025)
R-squared	0.391	0.439	0.076	0.146	0.307	0.005
N	32	13	19	32	13	19

Robust standard errors in parentheses and clustered at the school level.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: Source: TIMSS © IEA 1995-2019; World Bank EdStats. Each estimate shows the OLS regression results when projecting relative age effects with 11 months gap after assuming the linearity of the effect. Each index combines the related explanatory variables using the standardized inverse-covariance weight and `swindex` command in STATA.

Chapter 2

The Impact of International Students on Minority Enrollment in US Higher Education

Using data from IPEDS, this paper examines the impact of the inflow of international students on the first-time, full-time enrollment of domestic minority students. Since foreign enrollment is an endogenous variable, I employ the instrumental variables approach, using the institution's historical share of international students and the year's non-immigrant visa issuance. I find that there is no significant effect of the influx of international students on the new enrollment of domestic minorities as a whole. However, when I divide the institutions by the level of state funding per student, I find that an additional influx of international students increases domestic minority FTFT enrollment by 0.65. I suggest that this is because institutions with relatively little reliance on government funding are more sensitive to the financial resources that international students bring in terms of determining the supply and demand of domestic minority enrollment.

2.1 Introduction

Over the past two decades, the number of international undergraduate students at universities in the United States has increased dramatically. Data from the Digest of Education Statistics show that the number of international undergraduate students increased from 218,735 to 548,557, a 251% increase from 1990 to 2019. A natural question is how this dramatic change in the composition of the student population in US higher education affected domestic students. In particular, it is important to answer whether this effect arises in a different pattern toward minority groups since higher education has been viewed as a bridge for disadvantaged groups to gain opportunities to earn higher incomes and pave the way for higher social status (Baum et al. (2013)).

Patterns of the impact of international student inflows on domestic enrollments may differ between minority and non-minority groups for several reasons. First, on the demand side, the college enrollment of minority groups may increase due to the influx of international students if the financial resources brought in by international students are disproportionately allocated to support disadvantaged groups, thereby attracting more domestic minorities. This scenario is possible if low-income minorities receive more financial aid compared to other groups, based on the additional institutional funds contributed by international students. Similarly, from the supply side, the influx of international students may provide institutions with an opportunity to admit more domestic minority students, particularly if the school is financially constrained and requires additional financial inflows to expand the student body. On the other hand, domestic minority students may be crowded out by international students, similar to Borjas (2004)'s finding for white students in a graduate program. Alternatively, they may not be affected if there are few connections between the enrollments of domestic minorities and international students, or if there are cancelled-out effects within domestic minority groups.

This paper is the first to study how the inflow of international students has affected the first-time enrollment of domestic minority groups in 4-year universities in the United States. I utilize data from the Integrated Postsecondary Education Data System (IPEDS) spanning from 2000 to 2019, which encompasses every college and university in the US participating in federal student financial aid programs. This dataset includes institution-level data for the enrollment of each racial group and non-resident undergraduate students, who are considered international students in this study, as well as various institutional characteristics. Since most international students are concentrated at 4-year universities, this study focuses on 4-year universities.

Identifying the effect of international students on the enrollment of domestic minority students is challenging because of omitted variable bias and the potential for reverse causation. The flow of international students to the U.S. is not random, and if certain time-varying, unobservable within-institution factors, such as the attractiveness of the institution to both foreign and minority students, play a role, this will result in positively biased OLS estimates even after controlling for time and institution fixed effects. On the other hand, if institutions experiencing a decline in domestic students, including minority students, try to fill the slots by accepting more international students, this will lead to negatively biased OLS estimates.

This study addresses these concerns by using an instrumental variable approach to estimate the causal effect of the influx of international students on the enrollment of domestic minority students. To construct an instrumental variable, I exploit two sources of variation. The first arises from plausibly exogenous variation in the issuance of B visas, tourist visas that share many administrative features with F-1 visas, which are student visas that international students are required to have to legally enroll in U.S. higher education. I assume that variation in B visa issuance is primarily driven by the restrictiveness of visa policies, which is arguably independent of domestic minority demand for higher education or institutions' decisions to admit certain groups of students. Second, I use the historical share of international students for each institution out of the total international student population for that year. This is based on the assumption that schools with larger initial shares attract more international students in the following years and have more exposure to any exogenous shocks in the inflow of international students. By interacting F-visa approval rates and the initial share of international students at each institution, I construct an instrument for the first-time full-time (FTFT) enrollment of international students for each institution in a given year. Conditional on controls, my empirical strategy assumes that the instrumental variable, which is the interaction term in the first stage, is arguably exogenous. This strategy follows the same logic as that of the difference-in-differences estimator, comparing the difference in new domestic minority enrollments in years of high visa restrictiveness to years of generous visa issuance at institutions that have historically had more international students relative to schools with a low share of international students.

In all regressions, I include the institution and year-fixed effects to control for any time-invariant institutional characteristics and common time trends across institutions. However, regarding the instrumental variable restriction assumption, while it is unlikely that overall visa restrictions are directly related to domestic minority first-time full-time enrollment, the initial distribution of international students across institutions may be correlated with the economic conditions that each institution faces in the region to which it belongs, which may affect domestic minority first-time enrollment through channels other than the inflow of international students. To address this, I also include rich time-varying regional characteristics such as unemployment rates, the college-age population, the percentage of males, the percentage of blacks, the percentage of Hispanics, the median income, and the poverty rate at both the state and county levels to control for any time-varying determinants of domestic and international student enrollment.

The 2SLS regression results suggest a near-zero effect of international student inflows on new domestic minority enrollments, with statistically insignificant estimates. Given that the average international student inflow is 26.94, the point estimates and confidence interval suggest that the average effect of international student inflows cannot be greater than an 8.9 increase or a 2.7 decrease in new domestic minority enrollments, with the maximum possible change being only 4.38 percent of the average first-time full-time minority enrollment. In addition, the analysis of heterogeneity within minority groups reveals mixed effects, with positive point estimates for Hispanic FTFT enrollment contrasting with negative estimates for Black, American Indian, and Alaska Native students, resulting in a net zero effect on total minority enrollment.

However, when I further examine the heterogeneous effects by the level of government funding per student, I find that universities with low government funding per student have a statistically significant positive effect of international student inflows on domestic minority FTFT enrollment. This suggests that one additional international student increases domestic minority enrollment by 0.65. Given that the average number of international student inflows at institutions with low government funding is 16.85, this effect size implies an average increase of 10.95 domestic minority students due to the influx of international students, which accounts for 8.7 percent of total domestic minority FTFT enrollment. In contrast, the estimates for universities with high government funding per student show negative but statistically insignificant point estimates. This suggests that the overall insignificant result for domestic minority students is due to the offsetting effects between institutions with low and high government funding, while clear and strong positive effects of international student inflows on domestic minority FTFT enrollments are observed at institutions with relatively low government funding.

The results of this study contribute to the literature in that this study focuses specifically on domestic minority groups, which include Hispanic, Black, American Indian, and Alaska Native students, in contrast to other existing studies that mostly focus on the impact of international students on the overall domestic student population in higher education. While Hoxby (1998) is one study that examined the impact of foreign-born students on native-born disadvantaged students and found that immigrants may displace them at selective schools, the focus was primarily on students who immigrated to the United States at a younger age and from affluent Caribbean and Latin American families in the 1980s and 1990s, whereas the composition of international students in the current context has changed significantly since then. More recent studies have focused directly on the impact of international students who are clearly defined as nonresidents in higher education. However, these studies show conflicting results. Borjas (2004) examined how the growth in the number of international students enrolled in graduate programs affected domestic enrollments. Using a fixed effects regression, he showed that there is a crowding out effect on native white males, especially in private universities. However, using only fixed effects has a limitation in that it can bias the results if there are factors within universities that change over time. To address this issue, Shih (2017) used instrumental variables to control for the possible endogeneity problem and showed that international students appear to increase domestic enrollment in U.S. graduate programs. Although both studies looked only at graduate school data, the opposite results tell us that the direction in which the inflow of international students affects the outcomes of domestic students is ambiguous. Indeed, Shen (2016) examines data at the undergraduate level and finds no significant impact on domestic enrollment, except for a significant crowding-out effect at high-ranking research schools. However, these studies examine the average impact on the outcomes of domestic students as a whole, and their heterogeneity analysis focuses on the type of university, such as private/public or selective/non-selective schools. This study, on the other hand, focuses on heterogeneity in the demand side and delves more deeply into the effects on minority students, which is closely related to improving social inequality, one of the important goals of U.S. higher education.

The structure of this paper is as follows. section 2 describes the data and descriptive statistics, Section 3 outlines the identification strategy, and Section 4 presents the main empirical results and heterogeneous effects. Section 5 conducts robustness tests. Section 6 discusses the results and Section 7 concludes.

2.2 Data and Descriptive Statistics

The primary data come from the Integrated Postsecondary Education Data System (IPEDS). From this dataset, I use the information on institutional-level first-time, full-time (FTFT) enrollment by race, and the main outcome of interest is the sum of FTFT enrollment for Hispanic, Black, and American Indian and Alaska Native students. I also examine FTFT enrollment for all domestic students and other racial groups, such as white and Asian students. This dataset also collects information on the FTFT enrollment of nonresident students separately from domestic students, who can be considered international students, the main treatment variable of the study. One thing to note is that the enrollment data are for the new entry of groups of students, which is already a flow variable rather than a stock such as the enrollment of continuous students. Therefore, the effect that this study focuses on is how changes in *new* international student enrollments affect the number of *new* domestic minority student enrollments in U.S. higher education.

Since most international students are concentrated in 4-year institutions[1], I only examine schools that grant at least a bachelor's degree. The IPEDS includes rich information on whether the school is public or private, the highest level of degree offered, the degree of urbanization, whether the school has a hospital or grants a medical degree, the Carnegie Classification, which includes information on whether the school is a research-focused school[2], and other financial information such as average full-time tuition, grants and aids, and school expenditures. The data also includes county code information. Therefore, I merged it with regional information such as unemployment rates, number of college-age population (age between 19 to 23), percentage of males, percentage of blacks, percentage of Hispanics, median income, and poverty rate at both the state and county level from the Census, BLS, and Bridged-Race Population Estimates to construct control variables.

I use data from 2000 to 2019 because the years before 2000 have incomplete data, and the years after 2020 likely overlap with the COVID-19 pandemic. I have made the data a balanced panel, including only those universities with data points from 2000 to 2019. Figure 1 shows that international students' FTFT enrollment in 2000 was concentrated in several states, including California, New York, Pennsylvania, Texas, Illinois, Ohio, and Massachusetts. International students appear to be attracted to the states that have historically had more international students, and these states show a greater increase in international student enrollment in 2019. Table 1 describes the final dataset, which includes 1,627 universities, of which 33 percent are public universities. Research and selective schools

[1] In 2019, 85 percent of all international students were pursuing a bachelor's degree. Source: Opendoors

[2] Research universities include extensive and intensive doctoral or research universities.

make up slightly more than one-tenth of the sample.[3] The average 12-month unduplicated enrollment, which is the total enrollment of a school including new and continuing students, is approximately 5,741 students. The average first-time full-time enrollment is much smaller because it is part of the total enrollment. The average FTFT international student was about 27 students, and the average FTFT domestic minority student was about 203 students, with Hispanics slightly more than blacks. In Figure 2, the trends in FTFT enrollments and proportions for each student group are described. The FTFT enrollments of domestic and international students have increased steadily over the years. While the absolute number of FTFT White students, who make up the majority of domestic students, has remained almost constant and their share has decreased, the absolute FTFT enrollment and share of domestic minority students has increased. The absolute number and proportion of Hispanic students have increased relatively more than other racial groups within the minority student population.

Finally, I use tourist visa (B visa) data from the U.S. Department of State to construct part of the instrumental variable, which is the interaction between each institution's historical FTFT enrollment share and B visa issuance of the year. The average B visa issuance is 4,296 thousand, and the trends in B visa issuance are described in Figure 3 with the trends in the total FTFT enrollment for international students in the US 4-year undergraduate universities. While the overall upward trends in both FTFT enrollments of international students and B-visa issuance demonstrate the positive relationship between the endogenous variable and part of the instrumental variable, it is important to note that B-visa issuance shows particular fluctuations related to exogenous shocks in immigration policy or agreements between countries. In fact, Chen et al. (2023) argues that while the issuance of B visas reflects policy restrictiveness on U.S. entry and shares many administrative features with student visas, B visas are not for international students. Thus, changes in B visa issuance isolate the variation in visa policy restrictiveness and likely satisfy the exclusion restriction subject to additional controls. For example, a deep drop in B visa issuance after

[3] Selective institutions are those ranked as ”Most Competitive,” ”Highly Competitive Plus,” or ”Highly Competitive” in the 2009 Barron's Profile of American Colleges

the year 2001 is likely to be affected by the September 11 terrorist attacks in 2001 (Wasem (2004)), and a big drop in B visa issuance after 2015 may be caused by heightened scrutiny by consular officers regarding financial documents and post-graduation plans of applicants, changes in consular staffing and guidance, all interrelated with the Trump administration.[4] In addition, student visa and B visa, both non-immigrant visas, share similar trends as described in Figure A1 in practice. According to Chen et al. (2023), the reason these two types of visas share similar features is that the State Department asks consular officers to adjust the strictness according to immediate and near-term intent and does not expect applicants to have a detailed long-term plan for both visa applicants.

2.3 Empirical Strategy

The main challenge in identifying the causal impact of the influx of new international students into U.S. higher education, particularly at the undergraduate level, on the enrollment of new domestic minority students stems from issues of omitted variable bias and the potential for reverse causation. Therefore, this section describes an empirical strategy that exploits exogenous variation to address this difficulty.

First, to estimate how the new influx of international enrollments affects the number of domestic minority enrollments, the following model is adopted:

$$Y_{it} = \beta_0 + \beta_1 \cdot International_{it} + \beta_2 \cdot Total_{i,t-1} + \lambda_t + \lambda_i + X_{it} \cdot \Phi + \epsilon_{it} \tag{2.1}$$

where the outcome variable Y_{it} is the first-time full-time enrollment of domestic minority students at institution i in year t, and the main explanatory variable is the first-time full-time enrollment of international students at institution i in year t. Time and institution fixed effects control for any time-invariant attributes in each institution, as well as common changes in the entire economy. Observable time-varying institutional and regional

[4] Source: Viggo Stacey, "Significant" US F-1 visa denial rise in 2022 amid calls for immigration reform, The PIE News, July 26, 2023, accessed April 3, 2024, https://thepienews.com/news/24-us-f1-visa-denials-in-2022-amid-calls-for-immigration-reform/.

characteristics are controlled in a covariate matrix X_{it}. All institution-year observations are weighted by the 12-month unduplicated headcount of undergraduate students at baseline year (2000) and standard errors are clustered at the state level. An institution's previous total new enrollment for each t is also controlled since the school that increases its total size will attract both international and domestic minorities. Thus, the coefficient β_1, which is the coefficient of interest, is interpreted as the estimated effect of one additional enrollment of international FTFT students on the number of domestic minority FTFT enrollments. A positive $\hat{\beta}_1$ indicates that an increase in the inflow of international students increases the new enrollment of domestic minorities on average.

However, the coefficient β_1 based on OLS regression cannot be interpreted as the causal effect of international student inflows on domestic minority enrollments, since both groups may be affected by unobserved factors included in ϵ_{it}. For example, both groups may be attracted by factors such as positive changes in institutional characteristics that are unobservable. In this case, the OLS estimate for β_1 is biased upward. On the other hand, if an institution is trying to replace declining enrollments of domestic groups, including minority students, with international students, this could lead to reverse causation and bias β_1 downward.

To address these concerns, I construct an instrument variable Z_{it} by interacting the initial distribution of international FTFT enrollment across institutions in the base year (=2000) with the tourist B visa issuance across years, on the assumption that changes in the issuance of B visa, which is a temporary visiting visa, are sufficiently correlated to student visa issuance, and at the same time arguably exogenous to domestic minority enrollment conditional on the baseline controls such as time-varying institutional and regional characteristics. Thus, the first stage follows the following equation:

$$International_{it} = \alpha_0 + \alpha_1 \cdot Z_{it} + \alpha_2 \cdot Total_{i,t-1} + \lambda_t + \lambda_i + X_{it} \cdot \Phi + \epsilon_{it} \tag{2.2}$$

where

$$Z_{it} = \frac{International_{i,2000}}{\sum_i International_{i,2000}} \times \text{B visa}_t. \tag{2.3}$$

The reason I employ an interacted instrumental variable is to flexibly control for time effects and to improve the strength of the first stage, as described in Nunn and Qian (2014). When using the plausibly exogenous variation in B visas as an instrumental variable, the exclusion restriction assumption requires that changes in tourist visa issuance affect domestic minority enrollment only through changes in international student FTFT enrollment. However, since B visa issuance varies only over time, there may be other nonlinear changes over time that are spuriously correlated with domestic minority enrollment. While time-fixed effects can control for this, they absorb all variation in B visa issuance. Therefore, by interacting it with the initial and historical share of international students for each institution in the base year, an instrumental variable Z_{it} now varies by institution and time period, allowing for the control of year-fixed effects. According to Borusyak et al. (2022), this type of instrumental variable can be viewed as a specific shift-share instrument (SSIV) in panel data leveraging purely time-series shocks and time-invariant shares. In the same paper, it is noted that in this particular case, when the shift-share instrument is with a single industry, the first-stage and reduced-form estimates are similar to conventional difference-in-differences (DD) with continuous treatment. In other words, it compares the difference in new international enrollments (or new domestic minority enrollments) in years of high non-immigrant visa restrictiveness to years of generous visa issuance at institutions that have historically had more international students relative to schools with low shares of international students.

Even after controlling for time and institutional fixed- effects, there remains a concern that any changes in B visa issuance or initial share of international students across universities may be related to the time-varying regional economic situation and directly affect domestic minority enrollment at that school, which would still violate the exclusion restriction. However, this concern is mitigated by the inclusion of a rich set of controls that capture the differential responses of institutions to time-varying regional factors such as unemployment rates, college-age population, percentage of males, percentage of blacks, percentage of Hispanics, median income, and poverty rates at both the state and county levels, as described in Section 2.

Using this instrumental variable, the 2SLS estimates for the coefficient β_1 can be interpreted as the causal effect of the inflow of new international students on the new enrollment of domestic minority students, with the caveat that the effects should be interpreted locally, in that these international students are more attracted to universities with a higher initial share and are more likely to enroll when overall non-immigrant visa issuance becomes generous. Indeed, Figure 4 describes that the institutions with a higher initial share of international students' FTFT enrollment in 2000 have a higher average number of FTFT enrollments for international students in the following years.

2.4 Results

2.4.1 OLS Estimates

I begin by discussing the OLS estimates of equation (1). The results are presented in the first panel of Table 2. Column 2 reports estimates of the correlation between international student inflows and first-time full-time domestic minority enrollments. All regressions include regional and institutional time-varying controls, as well as time and institutional fixed effects. The estimate is close to zero and statistically insignificant. In columns 3 and 4, the estimates show statistically significant but opposite correlations with the FTFT enrollment of white and Asian domestic students. The overall association with total domestic new enrollment shows statistically insignificant but negative point estimates.

2.4.2 First-Stage and 2SLS estimates

The first-stage estimates of equation (2) are shown in the third panel of Table 2 and also plotted in Figure 5. Samples for 2000 are dropped because there is no lagged total FTFT enrollment for that year. Again, the specification includes time and institutional fixed effects, as well as regional and institutional time-varying controls, to mitigate concerns about a direct correlation between different initial proportions of international students across schools. The first-stage estimates show a strong positive correlation between the instrumental variable and domestic minority enrollment. The first-stage Kleibergen-Paap F-statistic is 35.55, suggesting that the estimates are less likely to be biased by weak instruments. To interpret the first-stage results, consider that an institution with an assumed share of 0.1 percent of all international students in 2000 would experience, on average, approximately 1.1 additional FTFT enrollments of international students for a 100,000 increase in B-visa issuance for one year. Since the average historical share in the sample is 0.0006 (0.06 percent), a 100,000 increase in B visa issuance is predicted to increase international student enrollment by an average of 0.66 for each institution. Given that there are 1,539 institutions included in the regression, the total increase in first-time, full-time international students resulting from a 100,000 increase in B visa issuance is estimated to be 1,016.

Figure 6 shows the difference-in-differences style first-stage and reduced-form plots by dividing the sample into those with an initial international student enrollment share above the median and those with a share below the median. In both the first stage and the reduced form, the continuous treatment variable, B visa issuance appears to have a greater impact on FTFT enrollment of international students at institutions with a high initial share compared to schools with a low initial share of international students.

The 2SLS estimates are derived as the ratio of the reduced-form estimates to the first-stage estimates, which is interpreted as the causal effect of an additional inflow of first-time, full-time international students on the new FTFT enrollment of domestic minority students. In the second panel of Table 2, the estimates in column 2 show a positive estimate, but which is statistically insignificant. Using 95% confidence intervals, this insignificant effect on domestic minorities FTFT enrollment is interpreted to rule out the positive effect greater than 0.33 and the negative effect greater than -0.10. Since the average FTFT international enrollment is 26.94, this means that the average effect of the influx of international students cannot be larger than a 8.9 increase or a 2.7 decrease in new domestic minority students, where the possible maximum change is only 4.38 percent of the average FTFT minority enrollment (203.39). On the other hand, an average inflow of 27 international students reduces, on average, about 11 white students, which is 2.1 percent of the average white enrollment, consistent with Borjas (2004)'s finding. Table A1 shows that while the overall effect for minority students is close to zero, there is heterogeneity within minority groups. The positive point estimates for Hispanic FTFT enrollment, in contrast to the negative point estimates for Black, American Indian, and Alaska Native students, suggest that each racial minority group may be affected differently by the influx of international students. It appears that these heterogeneous effects in opposite directions result in a net zero effect on total minority enrollment.

2.4.3 Small Public and Private Institutions versus Large Public Institutions

While there is no statistically significant effect of international student inflows on new domestic minority enrollments in U.S. undergraduate higher education, a deeper analysis reveals that international student inflows increase new domestic minority enrollments among small public and private universities within the overall distribution. In Table 3, I divide the sample into two groups: large public schools which are public schools with more than 10,000 student enrollment (12-month unduplicated headcount), and small public and private schools, which comprise the remaining schools. The majority of the first group are private universities (79 percent), but one-fifth are still public colleges of smaller size. The 2SLS estimates in column 1 of Panel A for small public and private universities, which make up 84.1% of the sample, show statistically significant positive estimate, and it indicates that one additional international student inflow increases domestic minority FTFT enrollment by 0.358. Considering that the average number of international student inflows at small public and private universities is 15.99, this effect size implies an average increase of 5.72 domestic minority students caused by the inflow of international students, which accounts for 4.93 percent of the total domestic minority FTFT enrollment. On the other hand, column 2 of the same table shows a negative but statistically insignificant point estimate that is close to zero. This implies that while large public universities have a greater number of both international and domestic minority students, they do not experience the same positive effects on domestic minorities from the influx of international students.

Panel B shows the impact on each racial group. Interestingly, the crowding-out effect of international students stands out only in small public or private schools, mainly due to the displacement of white students. This result is consistent with the crowding-out effect of white students from Borjas (2004)'s study. In addition, the positive effect of the influx of international students on minority students within small public or private schools appears to be driven mainly by Hispanic students, as indicated by a statistically significant, strongly positive point estimate. Although not statistically significant, black students also appear to be positively affected by the inflow of international students in small public or private schools. On the other hand, within large public schools, while Hispanic students are positively affected by the influx of international students, albeit with statistically insignificant estimates and smaller magnitudes of the estimate, Black students are crowded out by the influx of international students. In both groups of institutions, American Indians and Alaska Natives are slightly crowded out by the influx of international students, although the effect size is close to zero. In addition, in both groups of institutions, there was a significant increase in the new enrollment of Asian domestic students affected by the increase in the inflow of international students, with a much larger effect size in large public schools.

2.4.4 Institutions with Low- versus High- Government Funding

I hypothesize that the reason for the differential impact of the influx of international students on domestic minority students by school type and size is the differential reliance on government funding for revenue. In fact, Bound et al. (2020) mention that while private institutions rely on tuition and endowment income for their sources of support, public universities rely on state appropriations and tuition, with a more modest role for endowment income. Because a school's decisions about student group composition may be influenced by its reliance on different sources of funding, particularly when public universities rely heavily on state appropriations and are mandated to align with state educational goals, I examine potential heterogeneity that may exist among subsamples of institutions with low and high state funding.

First, I define low government funding schools as those with less than \$9,000 per student from government sources (including state appropriations, local appropriations, and government grants and contracts), while high government funding schools are those with \$9,000 or more per student from government sources. Descriptive statistics for these two groups are presented in Table A2, with about three-quarters of the samples representing institutions with low government funding per student. Compared to high government funding institutions, low government funding institutions are relatively smaller, less selective, and have lower proportion of minority enrollments. Approximately one-fifth of them are public universities, and their revenue sources are highly concentrated in tuition. While they provide relatively more financial aid per student than high government funding schools, they allocate less to instruction, research, public service, and academic support. Conversely, these low government funding schools allocate more resources to student services and institutional support.

In Table 4, the 2SLS estimate in column 1 of Panel A shows statistically significant effects with a much larger point estimate. It indicates that at low government funding universities, one additional international student inflow increases the domestic minority FTFT enrollment by 0.65. Given that the average number of international student inflows at low government funding universities is 16.85, this effect size implies an average increase of 10.95 domestic minority students caused by the inflow of international students, which accounts for 8.7 percent of the total domestic minority FTFT enrollment. On the other hand, column 2 of the same table shows a negative but statistically insignificant point estimate that is close to zero, showing a similar pattern to that shown for large public universities in the previous subsection. In Panel B, the effect for each racial group shows a similar conclusion from Table 3. The increase in domestic minorities due to the influx of international students at schools with low government funding is mainly driven by Hispanic and black students, while there is a small but negative effect on American Indian and Alaska Native FTFT enrollment. Regardless of the level of government funding, a significant crowding-out effect was observed for White domestic students, while a significant crowding-in effect was observed for Asian domestic students.

Figure 7 delves further into the first-stage and reduced-form effects plots in a difference-in-differences format for these subgroups divided by the amount of government funding. Looking at universities with low government funding in the left panels, while there is a relatively small change in mean FTFT enrollment for both international students and domestic minorities with changes in B-visa issuance at universities with low initial shares, there are larger first-stage and reduced-form effects at schools with high initial shares of international students. In the right panels, universities with high government funding show larger differences in the first-stage effect of changes in B-visa issuance between the high and low initial share groups. However, these universities with high government funding show more sporadic patterns in the reduced-form graph.

In contrast to the other groups, high government funding schools with both low and high initial international enrollments already have much larger numbers of domestic minority students than the other groups. This suggests that these schools have selectively maintained high minority enrollments relative to international enrollments, probably because these universities have historically relied heavily on government funding and are mandated to educate domestic students, including minorities (Groen and White (2004)). This story is also consistent with Chen (2021)'s finding that states allocate more funds to universities that attract fewer international students. However, as the level of B visa issuance, which is closely correlated with the change in student visa issuance, changes in the range of more than 5000 thousand B visa issuances, it appears that there is a negative correlation between B visa issuance and FTFT enrollment of domestic minorities. This negative correlation is mainly from the year after 2015, as described in Figure 3, when there was a steep downward trend in nonimmigrant issuance levels, and it appears that high government funding schools compensate for the decrease in international students with more domestic minorities. Similar patterns are shown in the reduced form graph for low state funding schools, but not as large as for high government funding schools. Therefore, it can be interpreted that there is more substitution between international and domestic minority groups in schools with high government funding, which may have resulted in negative reduced-form estimates and a negative, though insignificant, 2SLS point estimate. In sum, the near-zero effects of international student inflows on domestic minority FTFT enrollments shown in Table 2 result from the canceled effects between low and high government funding institutions, while clear and strong positive effects of international student inflows on domestic minority FTFT enrollments are found among institutions with relatively low government funding.

2.4.5 Effects on Transfer-in Students

A large proportion of minority students enter the four-year college system through the transfer system rather than directly through first-time, full-time enrollment. According to IPEDS data from 2000 to 2019, the ratio of transfer students to FTFT enrollment is 0.42 for minority students, compared to 0.33 for white students and 0.32 for Asian students. Since the transfer system is an important pathway for domestic minority students to enroll in U.S. higher education, a natural question is whether a similar pattern can be found among transfer-in minority students. However, this question is likely to face limitations in statistical power due to the small number of transfer-in students, and even more so for the smaller number of minority students who represent a subsample of the total student population. Nonetheless, it remains an intriguing exploration to determine whether a positive point estimate can be detected for the number of domestic minority transfer-ins within schools with low government funding, similar to the effects observed for FTFT enrollment in previous subsections.

In Table 5, the 2SLS estimate in column 1 shows a positive point estimate, 0.150, albeit is statistically insignificant as expected. It suggests that we can rule out the positive effect greater than 0.33 and the negative effect greater than -0.03, with a 95 percent confidence level. Since the average FTFT international enrollment among schools with low government funding is 16.85, this means that the average effect of the influx of international students cannot be larger than a 5.6 increase or a 0.5 decrease in domestic minority transfer-ins, where the possible maximum change is 10 percent increase in the average minority transfer-ins (51.79). While the point estimate for schools with low government funding is positive, the point estimate for schools with high government funding is negative, statistically insignificant, and closer to zero. For institutions with high government funding, we can rule out the positive effect of greater than 0.08 and the negative effect of greater than -0.33 at a 95 percent confidence level. Since the average FTFT international enrollment at high government funding institutions is 63.08, this means that the average effect of the influx of international students cannot be greater than an increase of 5 or a decrease of 20.8 in domestic minority transfer-ins, where the maximum possible change is a 10.3 percent decrease in average minority transfer-ins (201.85). In summary, although it is not possible to statistically detect the effects on domestic minority transfer-ins, the sign and the size of the point estimates between low and high government funding institutions suggest that similar patterns are maintained in the group of minority transfer-ins.

2.4.6 Additional Heterogeneity Checks

In Table A3, further heterogeneity checks are performed by comparing public and private universities, selective and non-selective universities, universities offering bachelor's degrees as the highest level of education and those offering master's degrees or higher, universities located in counties with a median income above the mean of the entire distribution and those below the mean, and universities located in urban or suburban areas compared to those located in rural areas.

When examining institutions by sector, private universities have statistically significant positive effects on domestic minority FTFT enrollment, while public schools have effects close to zero with statistically insignificant negative point estimates. This finding mirrors the results presented in Table 3, likely because private universities are a subset of small public universities and private institutions. No statistically detectable effects are observed for groups categorized by different levels of selectivity, highest degree offered, median income, and regional status. In addition, certain subgroups show small F-statistics in the first stage, indicating that the results of the 2SLS are not interpretable.

2.5 Robustness Check

2.5.1 Unbalanced Sample

As described in the data section, the institutions included in the main analysis are restricted to those with data points in all years from 2000 to 2019. In this section, I report the robustness of the results when the sample is relaxed to an unbalanced panel. Table A4 presents the results of the estimation when institutions are included in the sample if they have at least 10 data points starting from 2000. Thus, slightly more observations with a larger number of institutions are included in the sample. The results generally show similar effect sizes. The 2SLS estimate in Column 1 suggests that an additional influx of international students increases the FTFT enrollment of domestic minority students by 0.66 in schools with low government funding, which is almost the same effect size as shown in Table 4 with a balanced sample. In high government funding universities, the statistically insignificant negative point estimate of -0.124 is also similar in magnitude and direction to the effect shown in the balanced sample, suggesting that restricting the sample to the balanced panel or not does not significantly affect the results.

2.6 Discussion

Based on the finding that the influx of international students increases the new enrollment of domestic minority students only at smaller public or private schools and at institutions with low government funding per student, one might ask what causes this difference in how the influx of international students affects certain demographic groups within the U.S. domestic student population. While there may be many mechanisms that lead to these heterogeneous effects, since the difference in effects stands out with the standard related to the type of institutions and specifically the reliance on government funding in revenue, I suggest a possible story that institutions that are less reliant on government funding and more reliant on tuition-based revenue are more sensitive to the influx of international students, since the financial resources that international students bring may play a critical role in these institutions' decisions about student composition or how domestic minorities view these institutions when they apply for schools. As described in Table A2, there is significantly more institutional grant aid per student at low government funding schools than at high government funding schools, which can be supported by the financial resources brought in by international students and can be used to support domestic minority groups. In addition, if the story suggested by Chen (2021) that states tend to allocate less funding to universities that attract more international students is true, an increase in the influx of international students may negatively affect the financial resources that can be directed to domestic minorities at high government funding schools.

To find more evidence to support this narrative, I ran the same regression on the outcomes of financial variables, including tuition revenue per student, institutional aid per student, and each type of institutional expenditure, to observe which financial channels are affected by the influx of international students and whether there are any different patterns between schools with low and high state funding. All financial variables are deflated by the Higher Education Price Index(HEPI) and presented in 2013 dollars. Unfortunately, IPEDS does not provide financial data disaggregated by race, so it is only possible to indirectly assess which channel is more related to support for domestic minorities. The results are shown in Table A5. First, an additional influx of international students increases tuition revenue per student by about \$8 to \$10 at both low and high government funding schools. Similarly, institutional aid per student is positively affected by an additional influx of international students by a similar amount in both groups of schools. There are some significant differences in how these revenues are allocated to different categories of expenditures. While institutions with high government funding increased spending primarily on instruction and academic support, schools with low government funding allocated relatively more to research and public service, although the effect size is not statistically significant. While public service expenditures include support for general advisory services and services provided to specific sectors of the community, which may be related to domestic minority groups, the evidence from the table does not appear to be sufficient to draw any conclusions.

2.7 Conclusion

Over the past two decades, the number of international undergraduate students in the United States has increased significantly, and there have been questions about whether the influx of international students is crowding out domestic students. As higher education is often viewed as a means for minority groups to improve their socioeconomic status, this study examines the impact of the influx of international students on the first-time, full-time enrollment of domestic minority students at four-year institutions in the United States.

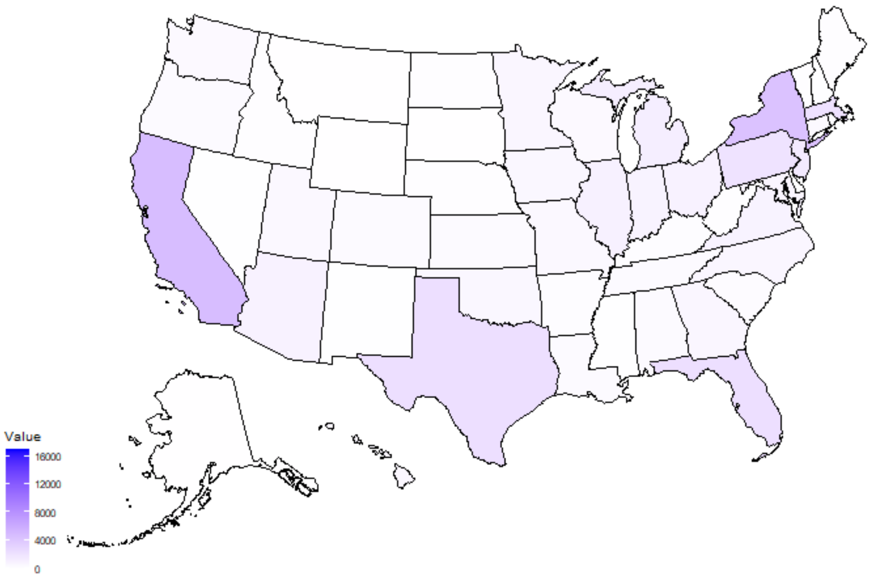
Analyzing data from the Integrated Postsecondary Education Data System (IPEDS) from 2000 to 2019, I use an instrumental variables approach to estimate the causal effect of international student inflows on domestic minority enrollments and find that the overall effect of international student inflows on new domestic minority enrollments is close to zero, with statistically insignificant estimates. However, when institutions are divided by the level of government funding per student, a statistically significant positive effect is found at institutions with low government funding per student. I suggest that this is because institutions with relatively little reliance on government funding are more sensitive to the financial resources that international students bring in terms of determining the supply and demand of domestic minority enrollment.

These findings contribute to the existing literature by focusing specifically on domestic minority groups, including Hispanic, Black, American Indian, and Alaska Native students, rather than examining the impact on the overall domestic student population. By addressing the limitations of previous research and employing a rigorous methodology, this study provides valuable insights into the complex relationship between international student enrollment and domestic minority enrollment in U.S. higher education, thereby advancing our understanding of social inequality within the educational landscape.

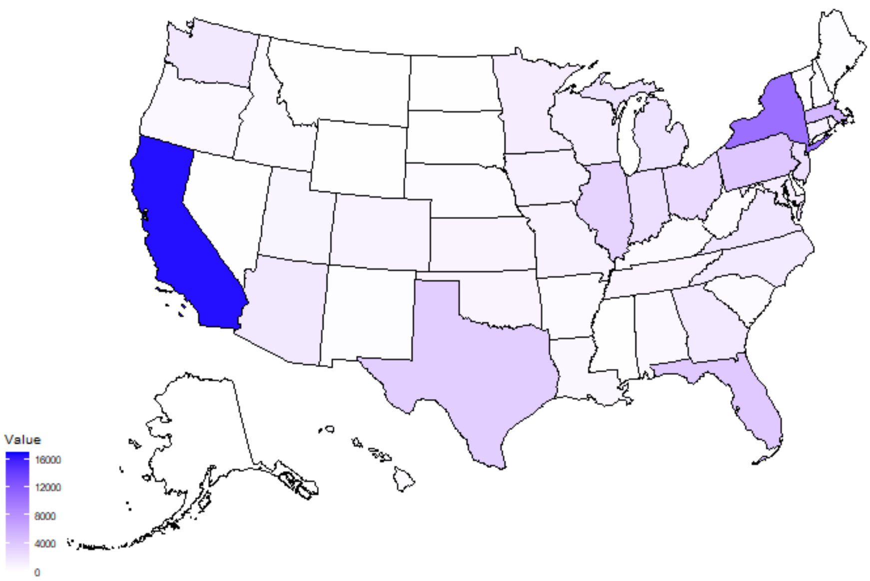
2.8 Figures and Tables

FIGURE 1: DISTRIBUTION OF INTERNATIONAL STUDENTS BY U.S. STATE

Year 2000

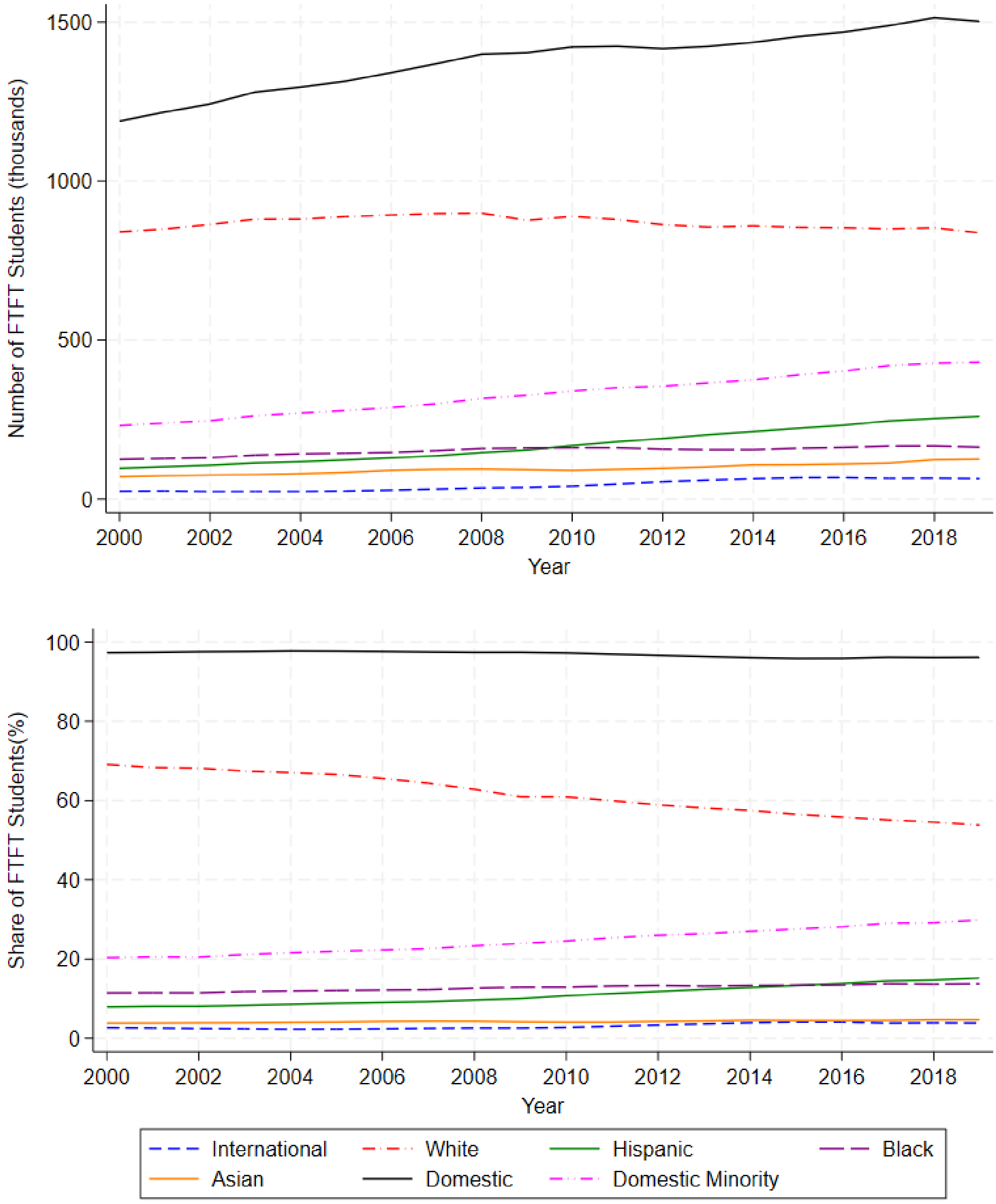


Year 2019



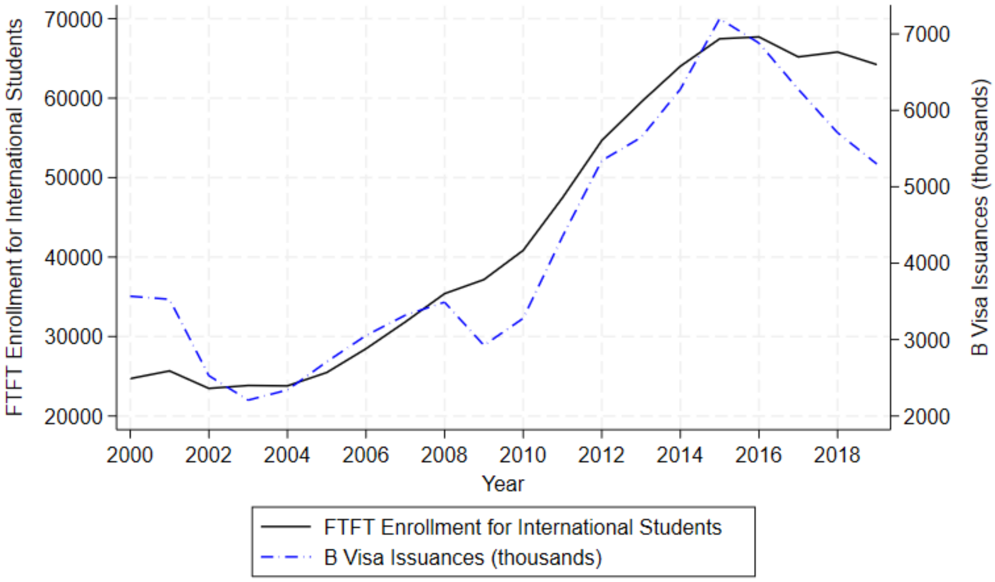
Source: IPEDS

FIGURE 2: TRENDS IN FIRST-TIME FULLE-TIME ENROLLMENT AND SHARE BY RACIAL GROUP



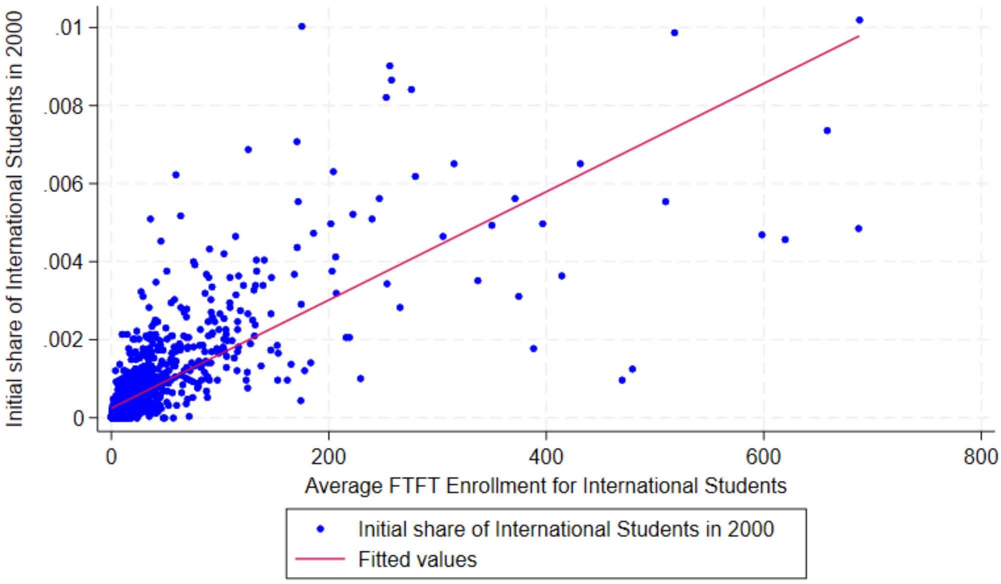
Source: IPEDS

Figure 3: Trends in F-visa Approval Rates and International FTFT Enrollment



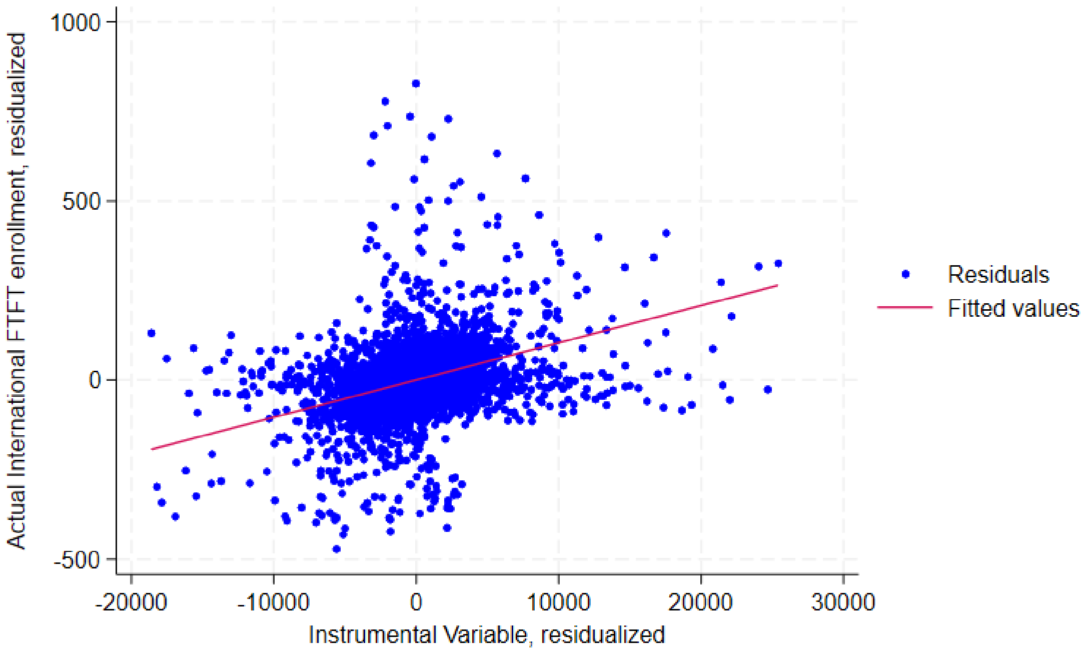
Data are obtained from the IPEDS and US Department of State.

Figure 4: Initial Share of International Students and the Average International FTFT Enrollment in the following years



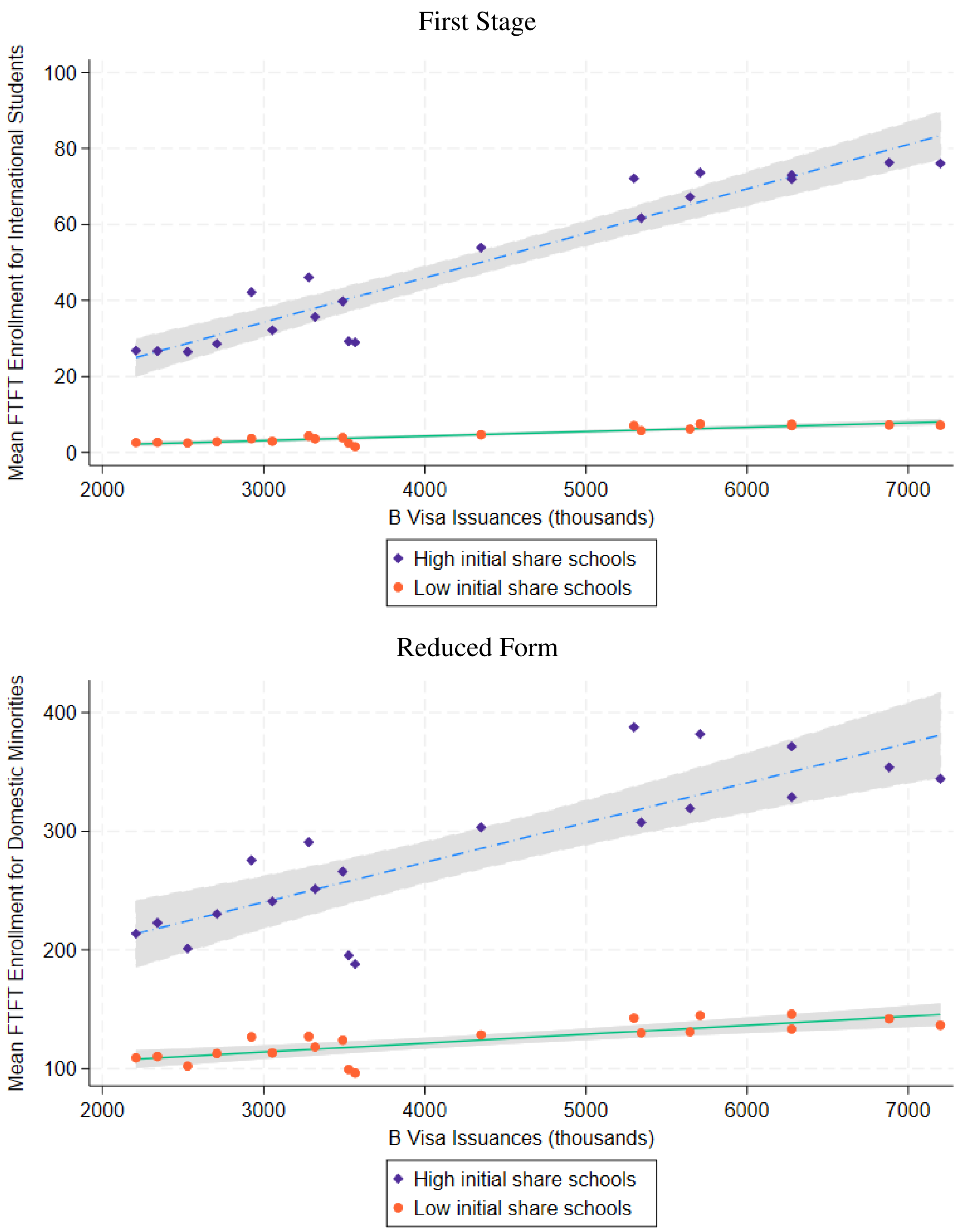
Source: IPEDS

FIGURE 5: FIRST STAGE RESULTS



Source: IPEDS

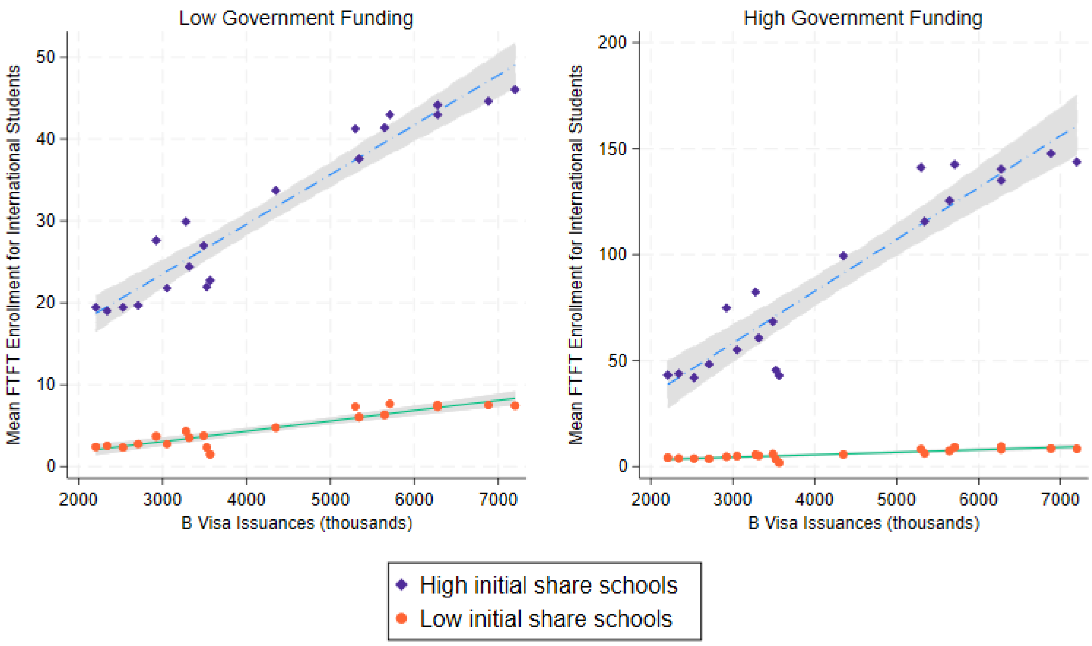
FIGURE 6: DD-STYLE FIRST STAGE AND REDUCED FORM GRAPHS BETWEEN LOW AND HIGH INITIAL SHARE SCHOOLS



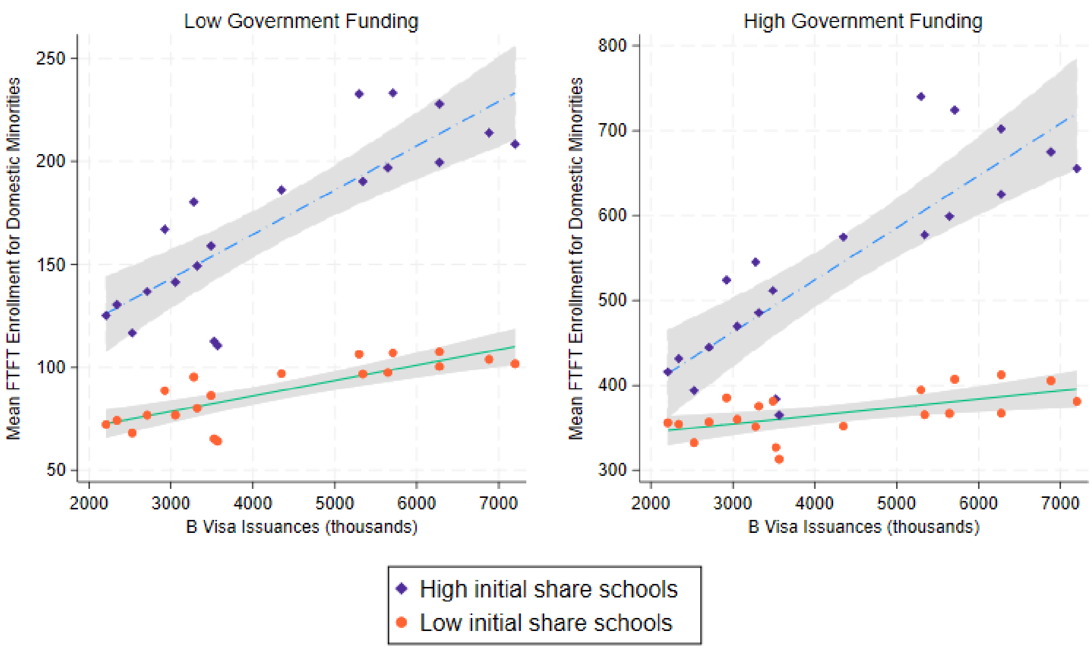
Source: IPEDS. High initial share schools are those with international student enrollment share at year 2000 above the median (.02 percent) and low initial share schools are those with a share below the median.

Figure 7: DD-style First Stage and Reduced Form Graphs between low and high initial share schools by the level of Government Funding

First Stage



Reduced Form



Source: IPEDS. High initial share schools are those with international student enrollment share at year 2000 above the median (.02 percent) and low initial share schools are those with a share below the median. Low government funding schools are those with less than $9,000 per student from government sources (including state appropriations, local appropriations, and government grants and contracts), while high government funding schools are those with $9,000 or more per student from government sources.

TABLE 1: DESCRIPTIVE STATISTICS

	Average (s.d. or percent)
N	32,540
Number of Universities	1,627
First-time Full-time Enrollment	
Grand total	875.02 (1170.31)
International	26.94 (76.53)
Domestic	848.08 (1126.27)
Minority	203.39 (347.78)
Hispanic	104.37 (268.53)
Black	93.40 (179.36)
American Indian	5.62 (18.33)
White	533.60 (790.78)
Asian	58.59 (181.03)
Minority Transfer-ins	85.66 (178.92)
12-month unduplicated headcount	5740.98 (7893.09)
B Visa Issuances	4,296,213.15 (1579534.40)
Sector of institution	
Public, 4-year or above	10,735 (33.0%)
Private not-for-profit, 4-year or above	21,030 (64.6%)
Private for-profit, 4-year or above	775 (2.4%)
Research schools	0.16 (0.37)
Selective schools	0.11 (0.32)
City/Suburb	0.49 (0.50)
Revenues per student	
Tuition Revenues	12,168.54 (7,128.23)
Government Fundings	5,997.34 (9,455.32)
Financial Aid per student	
State/Local grant aid	3,465.87 (2054.89)
Institutional grant aid	9,599.62 (8004.67)
Expenditures per student	
Instruction	9,785.69 (8306.50)
Research	3,146.63 (8826.48)
Public Service	992.32 (1667.06)
Academic Support	2,593.85 (3346.07)
Student Service	3,273.44 (2436.26)
Institutional Support	4,718.69 (4205.76)

Notes: Means of the samples for the years 2000-2019 are described. Standard deviations are shown in parentheses. Minority students include Black, Hispanic, and American Indian or Alaska Native as defined by the National Science Foundation (NSF). Research schools are those with Carnegie classifications of “research universities” and “doctoral/research universities.” Selective institutions are those ranked as ”Most Competitive,” ”Highly Competitive Plus,” or ”Highly Competitive” in the 2009 Barron’s Profile of American Colleges.

TABLE 2: OLS AND 2SLS RESULTS

	(1)	(2)	(3)	(4)
Dependent Var:	Domestic	Minority	White	Asian
OLS	-0.352	0.032	-0.601***	0.135*
	(0.232)	(0.060)	(0.192)	(0.074)
2SLS	-0.410*	0.113	-1.274***	0.436***
	(0.213)	(0.110)	(0.197)	(0.105)
1st stage		0.011***		
		(0.002)		
Observations	29218	29218	29218	29218
Mean $\bar{Y}$	875	198	559	62
F statistics	35.55	35.55	35.55	35.55
Number of Schools	1,539	1,539	1,539	1,539
Regional Controls	Yes	Yes	Yes	Yes
Institutional Controls	Yes	Yes	Yes	Yes

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: All observations are at the institution-year level. Minority students include Black, Hispanic, and American Indian or Alaska Native, as defined by the National Science Foundation (NSF). All specifications include year and institution fixed effects and the total number of FTFT enrollments in the previous year with the institution and regional time-varying controls. All observations are weighted by the 12-month unduplicated undergraduate enrollment in the base year (2000), and standard errors are clustered at the state level. Samples for 2000 are dropped because there is no lagged total FTFT enrollment for that year, nor is there information on F visa approval rates in 2000. The first-stage Kleibergen-Paap F-statistics are reported.

TABLE 3: DOMESTIC MINORITY ENROLLMENT IN SMALL PUBLIC AND PRIVATE SCHOOLS VS. LARGE PUBLIC INSTITUTIONS

	(1) Small Public and Private Schools	(2) Large Public Schools
Panel A: Dep. var:		
Minority FTFT Enrollment		
OLS	0.332***	-0.052
	(0.066)	(0.064)
1st stage	0.010***	0.011***
	(0.002)	(0.003)
2SLS	0.358***	-0.150
	(0.101)	(0.146)
Observations	24565	4653
Mean $\bar{Y}$	116	632
F statistics	22.32	17.00
Number of Schools	1,294	245
Regional Controls	Yes	Yes
Institutional Controls	Yes	Yes
Panel B: Dep. var.:		
Other racial groups FTFT Enrollment		
2SLS		
Total Domestic	-0.903***	-0.332
	(0.273)	(0.244)
White	-1.150***	-1.276***
	(0.056)	(0.241)
Hispanic	0.284***	0.090
	(0.069)	(0.095)
Black	0.099	-0.202**
	(0.068)	(0.080)
American Indian and Alaska Native	-0.025***	-0.038**
	(0.008)	(0.017)
Asian	0.217**	0.561***
	(0.085)	(0.158)

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: All observations are at the institution-year level. Large public schools are defined as public schools with more than 10,000 student enrollment (12-month unduplicated headcount), while small public and private schools comprise the remaining schools. Minority students include Black, Hispanic, and American Indian or Alaska Native, as defined by the National Science Foundation (NSF). All specifications include year and institution fixed effects and the total number of FTFT enrollments in the previous year with the institution and regional time-varying controls. All observations are weighted by the 12-month unduplicated undergraduate enrollment in the base year (2000), and standard errors are clustered at the state level. Samples for the year 2000 are dropped because there is no lagged total FTFT enrollment for that year, nor is there information on F visa approval rates in 2000. The first-stage Kleibergen-Paap F-statistics are reported.

TABLE 4: DOMESTIC MINORITY ENROLLMENT IN LOW VS. HIGH GOVERNMENT FUNDING INSTITUTIONS

	(1) Low Government Funding	(2) High Government Funding
Panel A: Dep. var:		
Minority FTFT Enrollment		
OLS	0.323***	-0.025
	(0.098)	(0.066)
1st stage	0.011***	0.010***
	(0.003)	(0.002)
2SLS	0.650**	-0.140
	(0.255)	(0.220)
Observations	21544	6515
Mean $\bar{Y}$	119	485
F statistics	15.67	16.56
Number of Schools	1,134	344
Regional Controls	Yes	Yes
Institutional Controls	Yes	Yes
Panel B: Dep. var.:		
Other racial groups FTFT Enrollment		
2SLS		
Total Domestic	-0.850***	-0.320
	(0.272)	(0.231)
White	-1.495***	-1.264***
	(0.302)	(0.263)
Hispanic	0.458***	0.042
	(0.167)	(0.188)
Black	0.221*	-0.140
	(0.124)	(0.096)
American Indian and Alaska Native	-0.029*	-0.042**
	(0.015)	(0.018)
Asian	0.227**	0.538***
	(0.097)	(0.147)

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: All observations are at the institution-year level. Low government funding schools are classified as those with less than $9,000 per student from government sources (including state appropriations, local appropriations, and government grants and contracts), while high government funding schools are those with $9,000 or more per pupil from government sources. Minority students include Black, Hispanic, and American Indian or Alaska Native, as defined by the National Science Foundation (NSF). All specifications include year and institution fixed effects and the total number of FTFT enrollments in the previous year with the institution and regional time-varying controls. All observations are weighted by the 12-month unduplicated undergraduate enrollment in the base year (2000), and standard errors are clustered at the state level. Samples for the year 2000 are dropped because there is no lagged total FTFT enrollment for that year, nor is there information on F visa approval rates in 2000. The first-stage Kleibergen-Paap F-statistics are reported.

TABLE 5: DOMESTIC MINORITY TRANSFER-INS IN LOW VS. HIGH GOVERNMENT FUNDING INSTITUTIONS

	(1) Low Government Funding	(2) High Government Funding
Panel A: Dep. var: Minority Full time Transfer-ins		
OLS	0.097	-0.055
	(0.142)	(0.043)
1st stage	0.011***	0.010***
	(0.003)	(0.002)
2SLS	0.150	-0.127
	(0.094)	(0.105)
Observations	15068	4603
Mean $\bar{Y}$	50	209
F statistics	15.66	13.28
Number of Schools	1,128	343
Regional Controls	Yes	Yes
Institutional Controls	Yes	Yes

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: All observations are at the institution-year level. Low government funding schools are classified as those with less than \$9,000 per student from government sources (including state appropriations, local appropriations, and government grants and contracts), while high government funding schools are those with \$9,000 or more per pupil from government sources. Minority students include Black, Hispanic, and American Indian or Alaska Native, as defined by the National Science Foundation (NSF). All specifications include year and institution fixed effects and the total number of FTFT enrollments in the previous year with the institution and regional time-varying controls. All observations are weighted by the 12-month unduplicated undergraduate enrollment in the base year (2000), and standard errors are clustered at the state level. Samples for the year 2000 are dropped because there is no lagged total FTFT enrollment for that year, nor is there information on F visa approval rates in 2000. The first-stage Kleibergen-Paap F-statistics are reported.

Chapter 3

Trends in Academic Achievement in OECD countries: Focus on Gender and Socioeconomic Disparities

Joint work with Kelly Bedard

This study examines how the academic achievement gap between different genders and socioeconomic groups within OECD countries has evolved over the years. Using Trends in International Mathematics and Science Study (TIMSS) data for eighth graders from eighteen OECD countries from 1995 to 2019, we first confirm that trends in academic achievement have progressed towards gender equality, particularly in science. Conversely, we find widening socioeconomic gaps, with high socioeconomic status (SES) groups showing greater improvements than low SES groups in both math and science test scores. When we examine the interactions between gender and socioeconomic groups to identify patterns driving these trends, we find that the SES gaps worsened for both males and females, and gender gaps similarly improved for both high and low SES students in most countries. Some countries show patterns that the worsening SES gaps are driven more by boys, and the improving gender gaps are driven more by the low SES students.

3.1 Introduction

Gender and socioeconomic (SES) gaps are known to exist in many educational settings. For example, Bedard and Cho (2010) show that boys outperform girls in math and science in most OECD countries, and Weinberger (2005) shows that girls are less likely to choose careers in sectors related to STEM. There is similarly a small number of studies analyzing the trend of SES-achievement gaps. For example, several studies looking at SES-achievement gaps in the United States suggest that the test score gap between the haves and have-nots has declined modestly or remained stable over four decades (Hanushek et al. (2022); Broer et al. (2019); Bai et al. (2021); Shakeel and Peterson (2022); Chmielewski (2019); Hashim et al. (2020)). In contrast, Reardon (2018) found that SES achievement gaps have widened over time.

While most studies of trends in SES achievement gaps focus on the U.S. context, there is an OECD report using PISA scores to track changes in the SES achievement gap between 2000 and 2015 (for Economic Co-operation and Development (2018)). They find no specific patterns in the changes in these gradients across countries. A book by Broer et al. (2019), probably the most similar analysis to ours, uses TIMSS data to estimate trends in SES achievement gaps, defined by the 75-25 gaps on an SES index constructed from indicators of parental education, books in the home, and the presence of a computer and study desk, for thirteen countries between 1995 and 2015. They found variation in the direction and magnitude of changes in SES achievement gaps. While Hungary, Iran, Singapore, and Lithuania showed statistically significant increasing SES achievement gaps, Norway, Slovenia, and the United States showed decreasing SES achievement gaps. In contrast to this study, our results construct the SES index slightly differently, as described in the data section, and we compare all populations with SES indexes above and below the median. As a result, we still observe an increasing SES achievement gap in Norway for both subjects, and in Slovenia and the US for either math or science. In contrast to Broer et al. (2019)'s study, which focused on a limited number of countries for data completeness, Chmielewski (2019) instead combined data from 30 international large-scale assessments, including 100 countries over 50 years. She found that SES achievement gaps (90-10 gaps) have increased in a majority of the sample countries, while there is considerable cross-country variation in the magnitude of the increases.

Compared to the literature on trends in the SES achievement gap, there are even fewer studies on trends in the global gender gap. First, in the U.S. context, Shakeel and Peterson (2022) estimated the gender gap trends in U.S. math and reading scores between 1954 and 2007 and found that there were no large differences in achievement trends by gender. For international studies, Meinck and Brese (2019) used TIMSS data from 1995 to 2015 and confirmed that gender equality in math and science has increased. This finding is similar to our findings, although their focus was on gender equality at the tails of the ability distributions. While they found that gender gaps have either persisted or widened since 1995 within the group of the top 20 percent of students in both math and science among fourth graders in the majority of countries in their sample, the gaps have closed or never existed among fourth graders at the lower end of the achievement distribution. Outside of STEM, Steinmann et al. (2023) focused on trends in gender gaps in reading at the end of primary school in 63 different education systems between 1970 and 2016. In contrast to STEM, they found an advantage of girls over boys in reading in almost all countries and observed a significant increase in this gap until 2001 and a slight decrease since then.

Our study is the first to think about the evolution of gender versus SES gaps across countries over time. We ask whether there are consistent patterns in these trends when these two critical variables – gender and a student's socioeconomic status – are interacted with the goal of identifying the precise subpopulations driving the trends. To address these questions, we use test scores of eighth graders from eighteen OECD countries with different education systems, obtained from the Trends in International Mathematics and Science Study (TIMSS). In our model, individual-level math and science test scores over the past two decades, from 1995 to 2019, standardized across countries, are regressed on time, gender, socioeconomic status, and their interactions.

The results reveal three interesting findings. First, we confirm that trends in academic achievement have been moving towards gender equality, especially in science. While female students performed less well than their male counterparts in math and science in 1995 and 1999, this gender gap has narrowed dramatically for science in most countries, except for one out of eighteen countries in 2015 and 2019. There was no such dramatic change in mathematics. Second, we find widening socioeconomic gaps, with high socioeconomic status (SES) groups showing greater improvements than low SES groups in both math and science test scores. While there are a few exceptions where the SES achievement gap has not changed or even decreased, most countries have experienced a widening of the SES achievement gap over the past two decades. Finally, when we examine the interaction between gender and socioeconomic groups to identify patterns driving these trends, we find that the low SES group shows greater improvements in the gender gap than the high SES group in some countries. Moreover, while a widening SES achievement gap is observed for both genders, boys show a more pronounced increase in inequality than girls in some countries.

Our study contributes to the literature in two ways. First, none of the previous studies have answered the question of which subgroup of students drives trends in gender and SES achievement gaps. To our knowledge, our study is the first to examine the interaction of these two demographic factors and which subgroup drives changes in gender and socioeconomic gaps over time. We believe it is necessary to clearly understand which subgroup of students is driving the convergence or divergence of gender and socioeconomic gaps in order to accurately diagnose the current status and properly target groups for further improvements in educational equity.

Another contribution of this study is that it examines trends in both gender and SES gaps in school achievement over time in several countries and discusses a general conclusion. There are two advantages to examining multiple country cases rather than focusing on a single country case. Although trends in a particular country may be of primary interest to decision-makers within that country, examining trends in achievement across different education systems in different countries provides additional information about how education systems are addressing the challenges faced by disadvantaged students (Broer et al. (2019)). This can have implications for decision-makers in terms of which education systems to target. If our goal is to reduce gender and SES gaps, we could learn from cases in certain countries where gender and SES gaps have been significantly reduced by examining patterns observed in those countries' educational or social support systems. Thus, the uniqueness of this study lies in its analysis of how gender and SES gaps have changed over time in each developed country over the past 20 years, using recently available internationally comparable test score data.

3.2 Data

We use Trends in International Mathematics and Science Study (TIMSS) data for eighth grade students[1] to analyze the trends in gender and socioeconomic (SES) achievement gaps. TIMSS has been administered every four years since 1995 to monitor trends in math and science achievement among students across countries. The countries participating in TIMSS vary from year to year, but always include a broad range of developed and less-developed countries. Our analysis is focused on eighteen Organization for Economic Co-operation and Development (OECD) countries that participated in 1995 and/or 1999 and were also present in 2015 and/or 2019. We refer to these two sample periods as 1995/99 and 2015/19 throughout the paper. The sample countries are listed in Table 1[2].

The math and science scores are standardized within each test book to a scale with a mean of 500 and a standard deviation of 100 across all countries in 1995. Subsequent assessment waves are linked to the 1995 scale, to allow for longitudinal comparisons across test years. Table 1 presents summary statistics for the countries included in our analysis. As expected, average math and science scores for OECD countries are above 500.

Hanushek et al. (2022) point out that survey questionnaires often bundle the right tail of distribution into broad categories that contain a large percentage of all observations, mak-

[1]TIMSS targets the grade that represents eight years of schooling.

[2]England and Japan are excluded because they do not report the data required to construct socioeconomic status.

ing it difficult to reliably estimate within-category differences. This is true of the family background data in TIMSS. For some countries, the parental education category "college degree or more" includes nearly half of the sample. For this reason, we construct an index of socioeconomic status using parental education, the number of books at home, whether a student has a desk at home, whether a student has a calculator or a dictionary in 1995/99, and whether a student has an internet connection or a cell phone in 2015/19. We use different variables for 1995/1999 and 2015/2019 to more accurately reflect socioeconomic status. Our SES index ranges from 0 to 11. Books at home range from 0 to 4, with categories: 0-15, 11-25, 26-100, 101-200, and 200+ books at home. Parental education ranges from 0 to 4, with categories: less than lower secondary, lower secondary, upper secondary, post-secondary, and university or higher. Has a desk, calculator, dictionary, and internet connection at home are all binary indicators for the student having the item at home. Figure A1 displays histograms for this index for each country in each sample period.

Conceptually, low (high) SES is defined as being below (at or above) the 50th percentile in a specific country and in a specific sample period.[3] Determining the exact cut-off is challenging because there are often many students with the same SES index as the median student. We therefore define high SES students as those whose SES score is greater than or equal to the SES score of the top 50% of students, and low SES is defined as the remaining students.[4] The distribution of high and low SES status across countries and sample periods are reported in Figure A2. To ensure that all results are robust to alternative SES definitions, we also define high SES to be anyone with an SES index of 8 or higher. The distribution of this measure is reported in Figure A3.

[3]Our comparison is broader than most other studies. For example, Hanushek et al. (2022), Broer et al. (2019), and Bai et al. (2021) compared the 75th and 25th SES percentiles, while Corak (2013), Chmielewski and Reardon (2016), and Chmielewski (2019) compared achievement in the top and bottom deciles.

[4]If a 50th percentile student is included in high SES, the proportion of high SES exceeds 50% as other students with the same SES score are also included in high SES group. Other studies such as Broer et al. (2019) avoid this problem by randomly splitting the sample of students around the cutoff to exactly fill the 25th and 75th percentiles.

3.3 Empirical Strategy

We examine the evolution of gender and SES gaps using a simple difference in difference in differences (DDD) model.

$$Score_{it} = \beta_0 + \beta_1 Y_{it} + \beta_2 F_{it} + \beta_3 L_{it} + \beta_4 F_{it} T_{it} + \beta_5 T_{it} L_{it} + \beta_6 F_{it} L_{it} + \beta_7 F_{it} T_{it} L_{it} + \epsilon_{it} \quad (3.1)$$

where $Score$ denotes the math or science test score for student i in year t, $T = 0$ for 1995/99 and 1 for 2015/19, $F = 1$ for female students and 0 otherwise, $L = 1$ if the student is low SES and 0 otherwise, and ϵ is the usual error term. The model includes grade indicators and a quadratic in age, but these are suppressed to focus on the key parameters. This parameterization allows changes in socioeconomic gaps to evolve differently by gender.

In all regressions, jackknife repeated replication (JRR) standard errors are used. According to Martin et al. (2020), TIMSS uses a variation of the Jackknife to estimate sampling variances. JRR was chosen because it is computationally straightforward and provides approximately unbiased estimates of the sampling variance means, totals, and percentages.[5] In addition, Hansen (2022) says conventional Heteroskedasticity-consistent (HC) and cluster-robust (CRVE) variance estimators and standard errors can be fully downward biased under standard conditions. He says this situation – full downward bias, unbounded size, and zero coverage probability – can be corrected by replacing conventional variance estimators with an appropriate jackknife estimator. The use of jackknife instead of conventional standard errors has the result that variance estimation is never downward biased and size distortion is bounded. We use the STATA package *pv* by Macdonald (2019), which includes the option to use the TIMSS-specific sampling zone and jackknife to generate the variance-covariance matrix (VCE) to estimate the sampling variance.

[5] According to Martin et al. (2020), at the core of the JRR technique is the repeated resampling from the one sample drawn, under identical sample design conditions. In the context of TIMSS, this entails the grouping of primary sampling units into sampling zones based on the TIMSS sample design and repeated draws of subsamples from these zones. The main features of the TIMSS sample design that JRR incorporates in its repeated replication are the stratification of schools and the clustering of students within schools. This was done by defining Jackknife sampling zones as pairs of successive schools to model the stratification and clustering from the national samples.

3.4 Results

Before turning to our main specification, it is helpful to show the gender and socioeconomic gaps in 1995/99 and how they changed by 2015/19 across countries. We do this using simple difference-in-difference models first including 2015/19 and female indicators and then including 2015/19 and low SES indicators. The gender gap and SES gap results are reported in Figure 1 and Figure 2 (and Appendix Table A3) for math and science, respectively.

The left panel of Figure 1 shows a range of female math gaps (female – male) in math, ranging from approximately 0 to -25 points on a 500-point mean. While there is some reduction in the magnitude of the gender gap over time, it is generally quite small, which is not surprising as the gaps were relatively small in 1995/99. In contrast, SES gaps (low SES – high SES) in math have gotten substantially worse as shown on the right panel of Figure 1. In country after country, we see that low SES students have fallen behind their high SES counterparts, often by large amounts.

For Science, the left panel of Figure 2 shows larger initial gender gaps in 1995/99 compared to math and a subsequent larger reduction in the gap by 2015/19. For example, the average female deficit in Ireland fell from -27 to 0 points. In seven of the 18 countries, the female science gap went to zero, or became positive. Overall, it generally shows substantial reductions in female test score gaps from 1995/99 to 2015/19, except for Chile and Hungary. For SES gaps in science, similar worsening patterns are shown except for Canada, Ireland, Korea, and the USA, on the right panel of Figure 2. It is important to note that these results are not an artifact of the SES measure. Similar results are reported in Appendix Table A4 using our alternate SES measure. These results are concerning as they suggest that educational inequality based on socioeconomic status has substantially worsened in most developed countries over the last few decades.

It is natural to ask if the SES patterns reported in Figure 1 and Figure 2 hide gender-specific patterns. Or if the improved gender gaps reflect overall changes across the SES distribution, or reflect SES-specific changes. We begin this exploration by graphing the change in the gender gap against the change in the SES gap. Figure 3 plots these for math and science. For math, countries with improved gender gaps had less worsening of their SES gap compared to countries with worsening gender gaps. In contrast, there is no systematic relationship between the two gaps in science.

Equation (1) allows us to examine these issues more directly. The results for equation (1) are reported in Table 2 and Table 3 for math and science, respectively. To aid in interpretation we have linearly combined coefficients to construct the group gaps, and gap changes, of interest. Columns (1) through (4) report the mean score change from 1995/99 to 2015/19 for high SES males, low SES males, high SES females, and low SES females. Columns (5) and (6) report the difference in the change for low SES males (female) compared to high SES males (female). Columns (7) and (8) are the differences in the change for high (low) SES females compared to high (low) SES males. Finally, column (9) reports of β_7, or how the difference in the gender gap changed differentially for high versus low SES students (column (8) minus column (7)). Or equivalently, if the SES gap for males changed differentially from the SES gap for females (column (6) minus column (5)).

Table 2 and Table 3 tell a simple story. The SES gaps worsened for both males and females in remarkably similar ways in both math and science. At the same time, gender gaps similarly improved for both high and low SES students in most countries. The patterns are most easily seen diagrammatically. The top panel of Figure 4 clearly shows that the change in the SES gap in math scores for girls and boys was similar in magnitude, with substantial worsening in most countries while the dots on the right side of the 45-degree line indicate that the worsening is slightly less for girls. In contrast, the bottom panel of Figure 4 shows that the gender gap in math scores for low and high SES students both improved in some countries, while it worsened for high SES students, despite improving for low SES students, in other countries. The DDD estimates shown in Column (9) of Table 2 indicate that 13 out of 18 countries show positive estimates, which means that some countries show patterns that the worsening SES gaps are driven more by boys and the improving gender gaps are driven more by the low SES students, although many DDD estimates are indistinguishable from zero. Let's compare the United States and Sweden for an example. For math in the United States, the male low SES minus male high SES gap increased by 17.32 points, while the same gap for women increased by 10.63 points resulting in a male-female differential SES gap change of 6.69; the change in the female SES gap worsened by 6.69 less than it did for males, albeit imprecisely estimated. The gaps changed by larger magnitudes in Sweden for math, but resulting in a similar male-female differential SES gap change of 7.90, again imprecisely estimated. Among such countries showing positive DDD estimates, 4 countries including Australia, Lithuania, New Zealand, and Portugal show these patterns at the 5% significance level, but we couldn't find any countries showing significant patterns in the opposite direction.

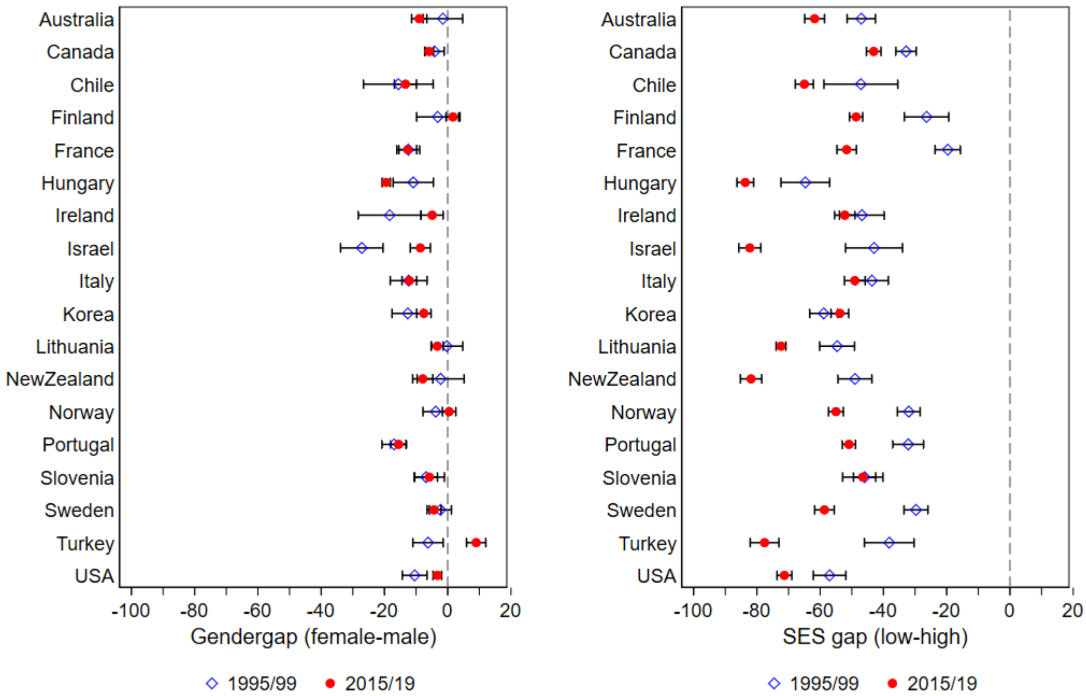
While the correlations have similar shapes for science (Figure 5), the magnitudes are quite different. The SES gap changes are large and negative for both genders, while the gender gap changes are positive, and more stood out among the low SES students, as indicated by the dots far from the 45-degree line on the second panel. Looking at Column (9) of Table 3, 15 out of 18 countries show positive DDD estimates, meaning that some countries show patterns that the worsening SES gaps are driven more by boys and the improving gender gaps are driven more by the low SES students, albeit imprecisely estimated. From the same example comparing the USA and Sweden in Table 3, the differences in changes of SES gaps between boys and girls are much larger for science in Sweden and smaller in the United States. For Sweden (U.S.), the male low SES minus male high SES gap increased by 48.67 (3.05) points, while the same gap for women increased by 35.07 (-2.62) points resulting in a male-female differential SES gap change of 13.60 (5.68), with the Swedish DDD estimate being statistically significant at the 5% level or better. 6 of 15 countries showing positive DDD estimates, including Australia, France, Israel, New Zealand, Portugal, and Sweden, show statistical significance in these patterns, while none of the countries show significant patterns in the opposite direction. To summarize, Figures 4 and 5 are consistent with worsening SES gaps driven by boys and improved gender gaps driven by low SES groups in some countries.

3.5 Conclusion

Taken as a whole, the results presented in Figures 1 to 5 and Tables 2 and 3 clearly show that in most countries there has been a substantial reduction of the gender gap over time, with some countries showing that the low socioeconomic status group is driving these patterns. Moreover, these patterns have coincided with a worsening of the socioeconomic status gap, with some countries showing that boys are driving it. The reduction in the gender gap over the past two decades, particularly among low SES students and in science, is encouraging for many efforts that have been made to reduce the gender gap in STEM fields, especially for low-income students who may not have sufficient resources to catch up with other students while they are behind. Nevertheless, we found that nearly half of the OECD countries in our sample show a worsening gender gap in math scores among the high SES students. Given that high SES students, both male and female, are the ones with the largest test score gains over the past two decades, this may be due to the sharp increase in math scores for high SES boys and the fact that high SES girls have lagged behind the rate of improvement for boys. Therefore, it implies that education policy focused on reducing the gender gap can focus on these disparities shown in math among high SES groups while maintaining efforts to increase test scores for both male and female students among low SES groups. On the other hand, we found the pervasiveness of the worsening SES gaps across subjects, genders, and countries except for a few countries in our sample, with boys showing worse trends in some countries. This result should be concerning to educators and policymakers and suggest the need for policies to reduce the socioeconomic status gaps across the population as a whole.

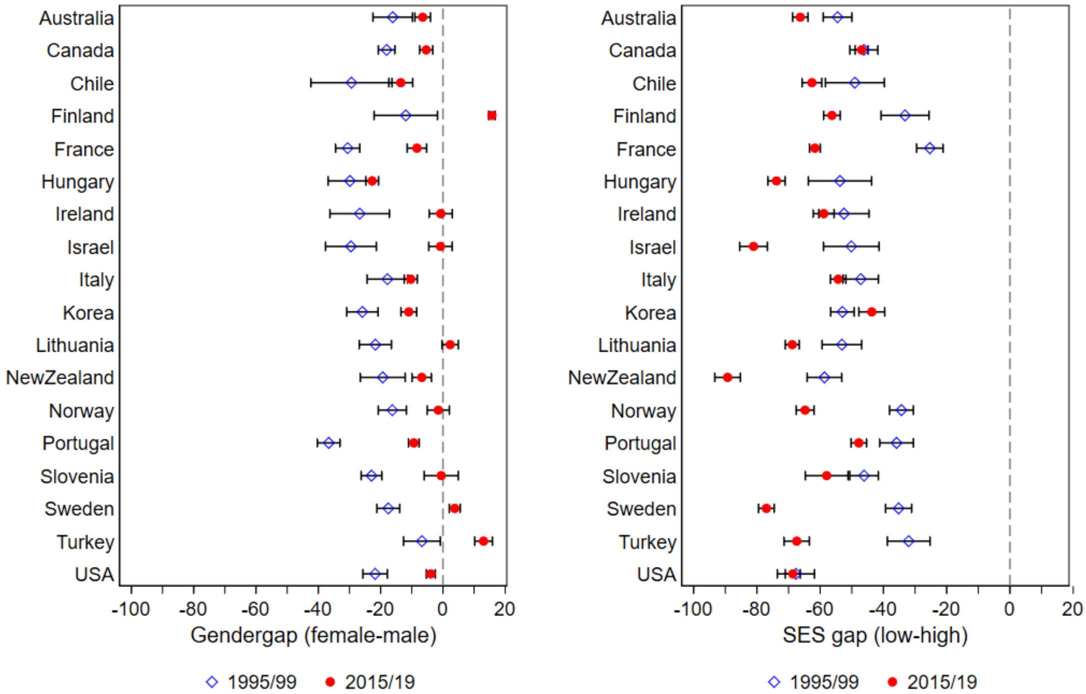
3.6 Figures and Tables

FIGURE 1: GENDER AND SOCIOECONOMIC STATUS GAPS FOR MATH



Source: TIMSS © IEA 1995-2019.

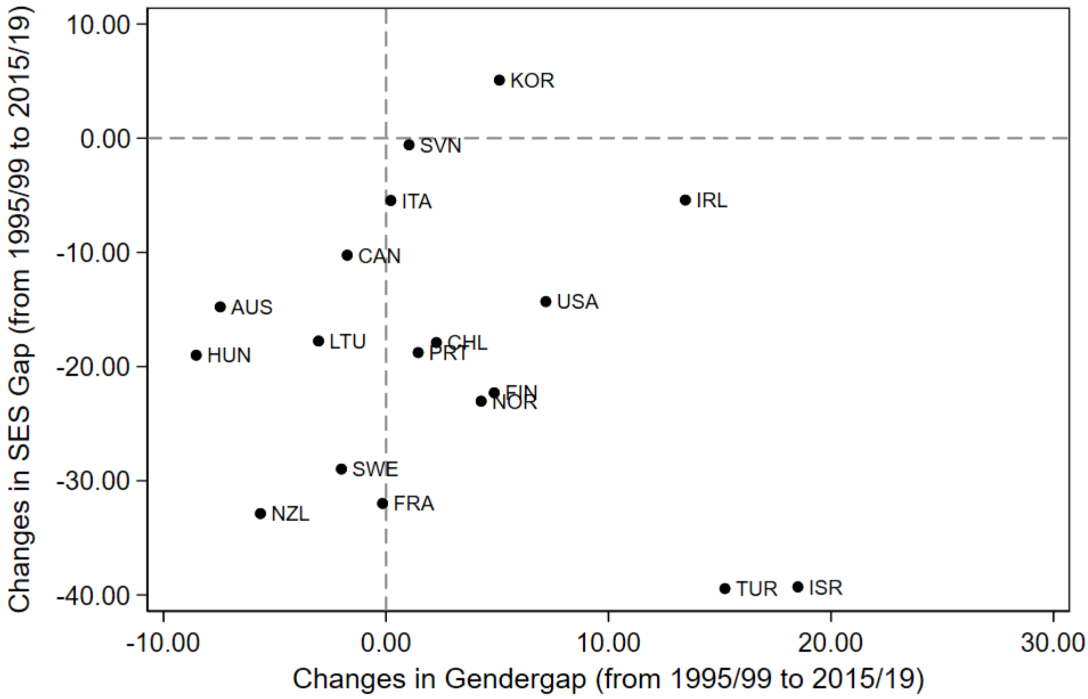
FIGURE 2: GENDER AND SOCIOECONOMIC STATUS GAPS FOR SCIENCE



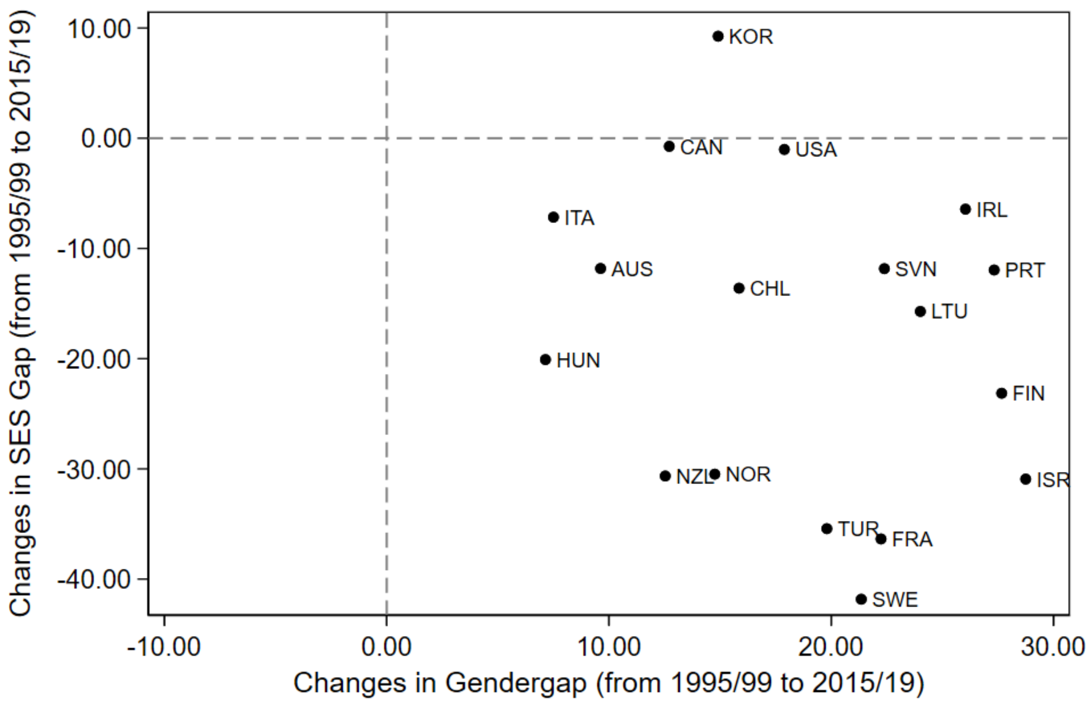
Source: TIMSS © IEA 1995-2019.

FIGURE 3: RELATIONSHIP BETWEEN GENDER AND SOCIOECONOMIC STATUS GAPS

Panel A: Math

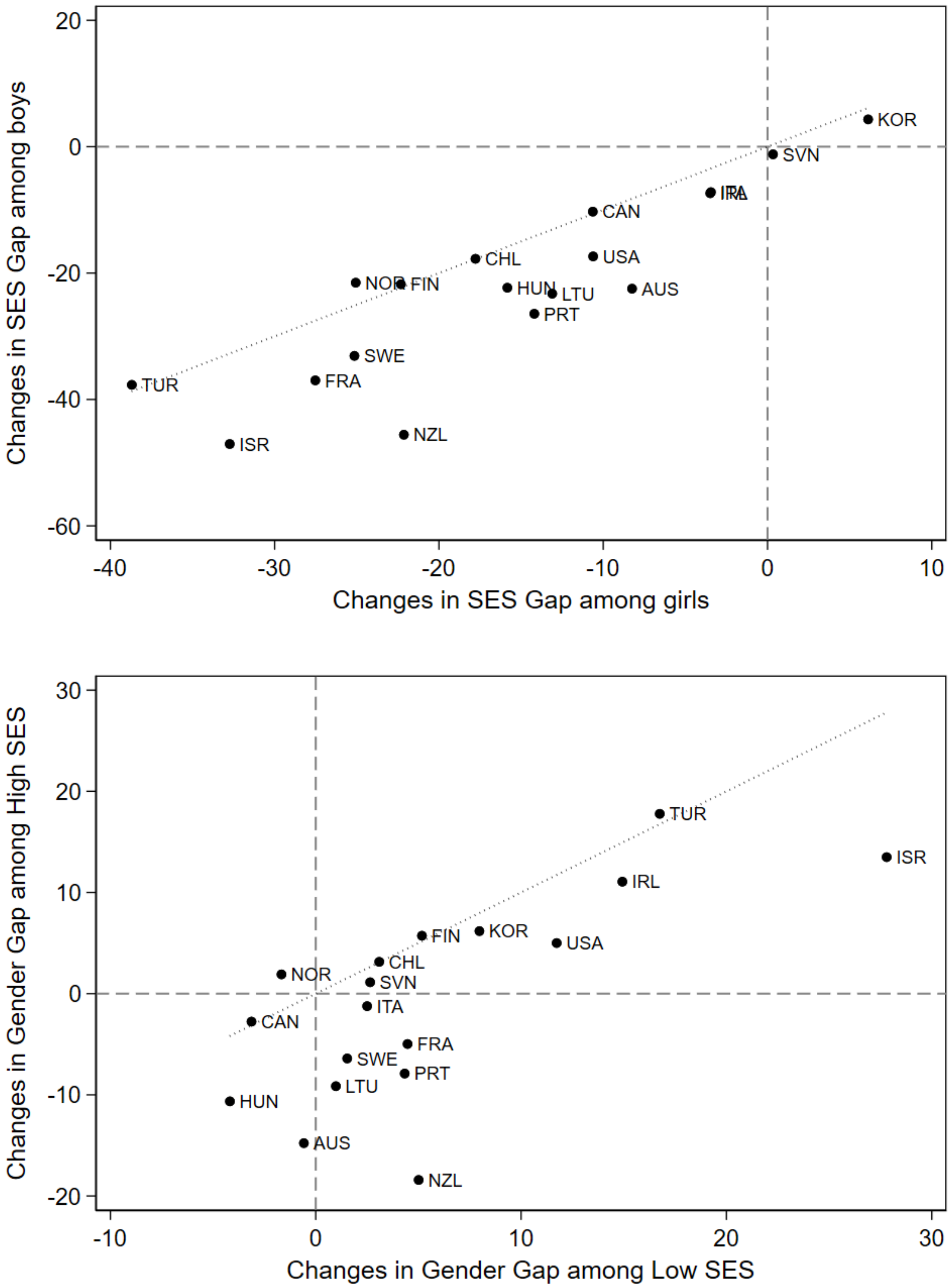


Panel B: Science



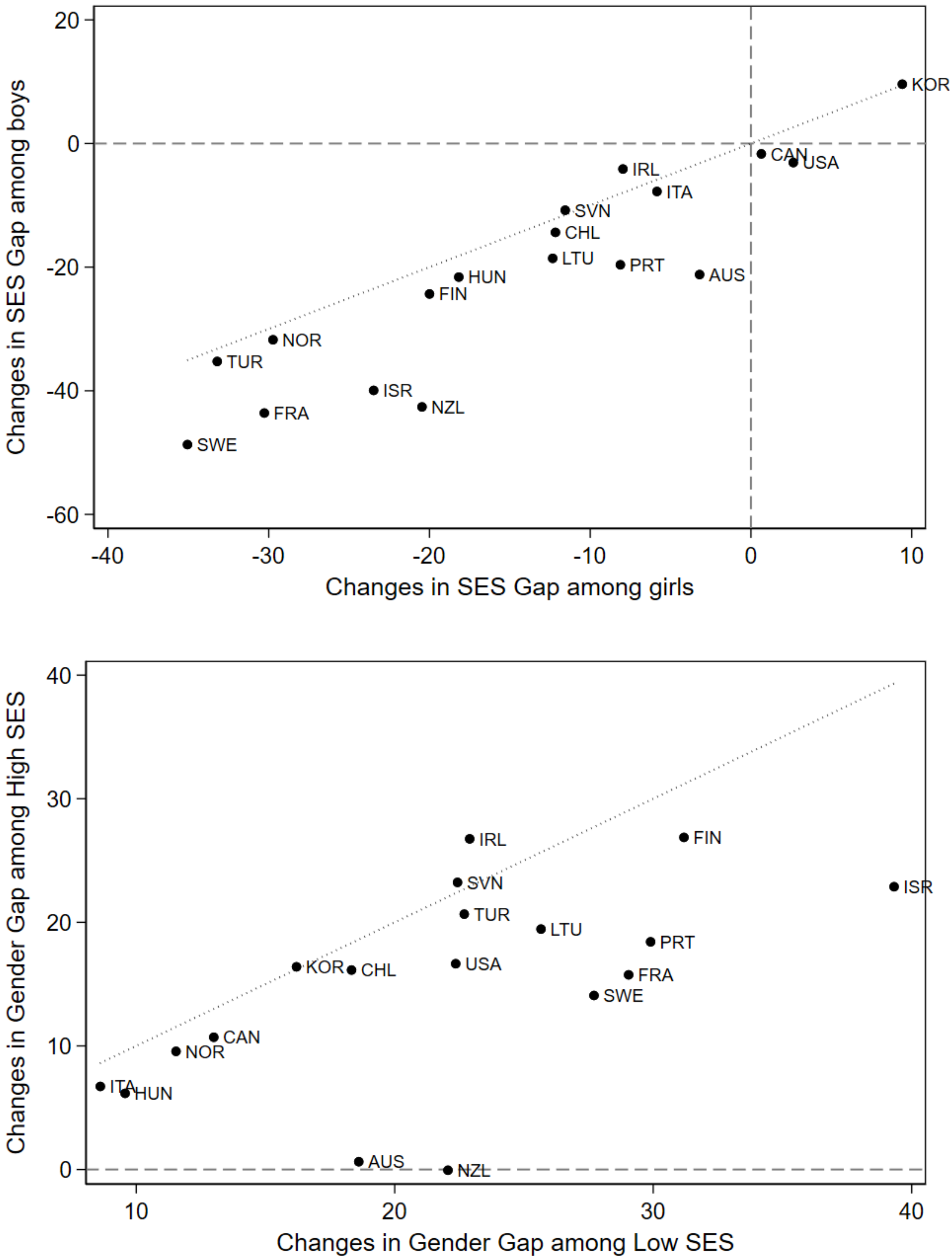
Source: TIMSS © IEA 1995-2019.

FIGURE 4: INTERACTION BETWEEN GENDER AND SOCIOECONOMIC STATUS GAPS FOR MATH



Source: TIMSS © IEA 1995-2019.

FIGURE 5: INTERACTION BETWEEN GENDER AND SOCIOECONOMIC STATUS GAPS FOR SCIENCE



Source: TIMSS © IEA 1995-2019.

Table 1: Descriptive Statistics

	(1) Math	(2) Science	(3) % Female	(4) % High SES	(5) Age	(6) % Later Period (Y=1)	(7) Sample Size
Australia	527.66	537.62	52.51	55.91	14.01	52.84	33,350
Canada	525.39	519.85	51.28	58.99	13.83	45.77	44,442
Chile	439.30	462.81	49.75	58.40	14.21	72.07	17,139
Finland	515.69	549.10	49.61	62.27	14.64	79.58	12,986
France	503.03	487.71	51.04	55.08	13.88	56.00	11,931
Hungary	531.35	543.66	50.06	59.27	14.61	88.93	27,183
Ireland	525.96	525.05	51.59	60.67	14.22	60.87	15,239
Israel	512.56	512.59	51.35	63.47	13.99	65.96	16,215
Italy	492.31	498.78	50.81	63.09	13.70	56.07	17,645
Korea	596.72	554.41	48.94	60.65	14.25	47.70	22,968
Lithuania	509.64	512.99	51.71	64.94	14.56	74.38	21,535
NewZealand	504.42	520.31	50.52	58.49	13.85	48.36	19,425
Norway	504.10	507.40	51.01	58.15	14.18	69.70	17,509
Portugal	476.30	490.89	51.47	56.78	14.00	55.66	14,433
Slovenia	521.08	537.78	51.18	64.10	14.30	25.36	11,303
Sweden	514.82	532.84	50.71	54.83	14.37	56.26	19,836
Turkey	462.40	479.97	47.34	59.33	14.02	64.06	20,469
USA	505.79	517.49	51.45	56.42	14.12	51.64	39,780
Total	512.53	518.33	50.78	59.12	14.14	58.86	383,388

Notes: Source: TIMSS © IEA 1995-2019. Columns (1) and (2) report country-specific average math and science scores. Column (3) reports the percentage of females and column (4) shows the percentage of high SES students. Students with high socioeconomic status are defined as those with SES score bigger or equal to top 50% student's SES score and low SES is defined as the rest. Column (5) shows the average age of students, column (6) shows the percentage of samples from the later period (Y=1, the year 2015 or 2019) among the entire samples. Earlier period is the year 1995 and 1999 (Y=0). England, Japan, and Slovenia are excluded since they don't have data points in either Y=0 or Y=1.

TABLE 2: GENDER AND SOCIOECONOMIC STATUS GAPS OVER TIME FOR MATH: 1995/99 TO 2015/19

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Difference over Time				SESgap Changes		Gendergap Changes		
	Δ Male High	Δ Male Low	Δ Female High	Δ Female Low	Δ Male Low - Δ Male High	Δ Female Low- Δ Female High	Δ Female High - Δ Male High	Δ Female Low - Δ Male Low	DDD
Australia	32.84*	10.42*	18.09*	9.83*	-22.43*	-8.26*	-14.75*	-0.59	14.17*
	(3.65)	(3.01)	(3.43)	(3.10)	(3.57)	(3.68)	(3.42)	(3.28)	(4.06)
Canada	24.93*	14.69*	22.19*	11.55*	-10.24*	-10.65*	-2.74	-3.14	-0.41
	(2.24)	(2.64)	(2.22)	(2.53)	(2.49)	(2.75)	(1.86)	(2.23)	(2.50)
Chile	67.37*	49.66*	70.53*	52.75*	-17.70*	-17.79*	3.17	3.08	-0.08
	(7.37)	(5.21)	(5.54)	(4.89)	(7.95)	(6.52)	(7.24)	(5.07)	(7.88)
Finland	-4.10	-25.83	1.66	-20.67	-21.73*	-22.33*	5.75	5.16	-0.59
	(16.05)	(16.38)	(15.58)	(16.22)	(4.99)	(4.05)	(5.22)	(4.95)	(6.29)
France	-34.15*	-71.09*	-39.09*	-66.63*	-36.94*	-27.54*	-4.95*	4.46	9.41
	(3.19)	(4.49)	(3.51)	(3.62)	(3.30)	(3.96)	(2.26)	(4.48)	(4.90)
Hungary	8.71*	-13.57*	-1.91	-17.76*	-22.28*	-15.85*	-10.62*	-4.19	6.43
	(4.33)	(4.43)	(4.26)	(4.21)	(5.44)	(5.42)	(4.07)	(5.40)	(6.51)
Ireland	6.04	-1.29	17.12*	13.63*	-7.33	-3.50	11.09*	14.92*	3.83
	(5.11)	(5.58)	(4.32)	(5.49)	(5.88)	(4.51)	(5.47)	(6.15)	(6.53)
Israel	41.78*	-5.23	55.31*	22.56*	-47.01*	-32.75*	13.52*	27.79*	14.27
	(4.76)	(6.18)	(3.85)	(5.57)	(6.49)	(6.14)	(4.40)	(6.33)	(7.55)
Italy	14.82*	7.67*	13.62*	10.16*	-7.15	-3.46	-1.21	2.49	3.69
	(3.44)	(3.85)	(2.84)	(4.14)	(3.87)	(4.48)	(3.05)	(4.53)	(4.65)
Korea	30.74*	35.09*	36.94*	43.05*	4.36	6.11	6.21*	7.96*	1.75
	(2.41)	(2.87)	(2.78)	(2.74)	(3.55)	(3.29)	(3.13)	(3.51)	(4.71)
Lithuania	50.31*	27.09*	41.19*	28.06*	-23.22*	-13.12*	-9.12*	0.97	10.10*
	(4.21)	(4.19)	(4.02)	(4.48)	(3.77)	(3.12)	(2.82)	(3.75)	(3.88)
New Zealand	38.76*	-6.78	20.37*	-1.78	-45.54*	-22.15*	-18.39*	5.00	23.39*
	(4.48)	(4.29)	(3.84)	(3.57)	(4.70)	(3.65)	(5.14)	(4.08)	(5.65)
Norway	8.43*	-13.04*	10.36*	-14.72*	-21.48*	-25.08*	1.93	-1.68	-3.61
	(3.64)	(3.85)	(3.49)	(3.89)	(3.18)	(3.26)	(3.06)	(3.48)	(4.63)
Portugal	57.13*	30.73*	49.26*	35.05*	-26.40*	-14.21*	-7.87*	4.32	12.19*
	(3.76)	(2.99)	(2.96)	(3.16)	(3.66)	(3.27)	(2.52)	(3.55)	(4.10)
Slovenia	-24.14*	-25.31*	-22.98*	-22.67*	-1.17	0.32	1.15	2.64	1.49
	(4.09)	(4.84)	(3.96)	(4.42)	(4.75)	(4.74)	(3.73)	(4.67)	(6.18)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Difference over Time				SESgap Changes		Gendergap Changes		
	Δ Male High	Δ Male Low	Δ Female High	Δ Female Low	Δ Male Low - Δ Male High	Δ Female Low- Δ Female High	Δ Female High - Δ Male High	Δ Female Low - Δ Male Low	DDD
Sweden	-6.29	-39.35*	-12.67*	-37.82*	-33.06*	-25.16*	-6.38*	1.52	7.90
	(4.71)	(5.11)	(4.64)	(4.79)	(3.56)	(2.57)	(2.79)	(3.09)	(4.18)
Turkey	57.57*	19.92*	75.37*	36.66*	-37.65*	-38.71*	17.80*	16.74*	-1.06
	(4.33)	(4.61)	(4.79)	(5.66)	(4.55)	(5.32)	(3.45)	(3.15)	(4.73)
USA	37.97*	20.65*	43.00*	32.37*	-17.32*	-10.63*	5.03*	11.72*	6.69
	(3.50)	(3.86)	(3.16)	(3.54)	(3.72)	(3.17)	(2.25)	(3.08)	(3.60)

Standard errors in parentheses
* $p < 0.05$

Notes: Source: TIMSS © IEA 1995-2019. Standard errors are in parentheses and * means the significance at the 5 percent level or better. The table shows the point estimates of the linear combinations of the estimates as written in the second row. Students with high socioeconomic status are defined as those with SES index bigger or equal to top 50% student's SES index and low SES is defined as the rest. Columns (1)-(4) show the difference in scores over time in each group and (5)-(8) show the Difference in differences results to show whether the high SES led the change when compared to low SES (Column (5) and (6)) and whether females led the change when compared to males (Column (7) and (8)). In the regression, jackknife repeated replication (JRR) standard errors are used, and covariates include indicator variables for grade, age and age^2. Sample sizes for each country are reported in Table 1.

Table 3: Gender and Socioeconomic Status Gaps over Time for Science: 1995/99 to 2015/19

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Difference over Time				SESgap Changes		Gendergap Changes		
	Δ Male High	Δ Male Low	Δ Female High	Δ Female Low	Δ Male Low - Δ Male High	Δ Female Low- Δ Female High	Δ Female High - Δ Male High	Δ Female Low - Δ Male Low	DDD
Australia	24.86*	3.70	25.51*	22.30*	-21.16*	-3.21	0.65	18.60*	17.95*
	(3.51)	(3.25)	(3.00)	(2.88)	(3.35)	(3.09)	(3.40)	(3.21)	(3.77)
Canada	12.12*	10.48*	22.84*	23.47*	-1.64	0.63	10.72*	12.99*	2.27
	(2.39)	(2.30)	(2.74)	(2.51)	(2.75)	(3.36)	(1.95)	(2.39)	(2.94)
Chile	54.64*	40.30*	70.80*	58.62*	-14.34*	-12.18*	16.16*	18.32*	2.16
	(6.38)	(4.49)	(4.25)	(5.79)	(6.93)	(5.97)	(6.98)	(7.89)	(8.43)
Finland	0.98	-23.32*	27.87*	7.85	-24.30*	-20.01*	26.89*	31.18*	4.29
	(11.48)	(10.58)	(11.84)	(12.15)	(5.62)	(4.93)	(6.60)	(6.63)	(7.44)
France	7.06*	-36.51*	22.83*	-7.46*	-43.56*	-30.29*	15.77*	29.04*	13.27*
	(3.35)	(3.75)	(3.63)	(3.58)	(3.31)	(3.20)	(3.26)	(3.69)	(4.77)
Hungary	-8.84*	-30.41*	-2.66	-20.85*	-21.57*	-18.19*	6.18	9.56	3.38
	(4.37)	(5.71)	(4.26)	(5.52)	(6.83)	(6.04)	(4.04)	(7.75)	(8.95)
Ireland	1.59	-2.50	28.37*	20.39*	-4.09	-7.98	26.77*	22.89*	-3.88
	(5.10)	(5.61)	(4.84)	(5.69)	(5.72)	(5.11)	(5.29)	(6.13)	(6.72)
Israel	28.29*	-11.61	51.19*	27.70*	-39.90*	-23.49*	22.90*	39.31*	16.41*
	(4.87)	(7.20)	(4.08)	(5.46)	(7.53)	(5.69)	(4.39)	(6.87)	(7.96)
Italy	5.49	-2.23	12.23*	6.37	-7.72	-5.86	6.74	8.60	1.86
	(4.05)	(4.33)	(2.60)	(4.11)	(4.20)	(3.67)	(3.91)	(4.89)	(5.14)
Korea	11.07*	20.70*	27.49*	36.89*	9.63*	9.40*	16.42*	16.19*	-0.23
	(2.86)	(2.30)	(2.58)	(2.93)	(3.65)	(3.38)	(3.65)	(3.79)	(5.27)
Lithuania	50.77*	32.23*	70.24*	57.88*	-18.54*	-12.35*	19.47*	25.65*	6.18
	(4.63)	(4.81)	(4.82)	(5.01)	(5.23)	(3.59)	(3.54)	(3.67)	(5.59)
New Zealand	33.91*	-8.66	33.87*	13.39*	-42.57*	-20.48*	-0.04	22.05*	22.08*
	(4.16)	(4.75)	(3.55)	(4.17)	(4.90)	(3.92)	(4.19)	(4.21)	(4.79)
Norway	-2.69	-34.40*	6.89	-22.87*	-31.71*	-29.75*	9.58*	11.53*	1.95
	(4.26)	(4.43)	(4.10)	(5.33)	(3.78)	(3.20)	(3.48)	(4.30)	(5.07)
Portugal	39.12*	19.54*	57.56*	49.43*	-19.58*	-8.14	18.44*	29.89*	11.45*
	(3.43)	(3.29)	(3.26)	(3.47)	(3.50)	(4.34)	(2.49)	(4.16)	(5.44)
Slovenia	3.20	-7.55	26.45*	14.88*	-10.75*	-11.57*	23.25*	22.42*	-0.83
	(4.27)	(5.31)	(4.57)	(4.48)	(5.11)	(5.22)	(4.34)	(4.85)	(6.81)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Difference over Time				SESgap Changes		Gendergap Changes		
	Δ Male High	Δ Male Low	Δ Female High	Δ Female Low	Δ Male Low - Δ Male High	Δ Female Low- Δ Female High	Δ Female High - Δ Male High	Δ Female Low - Δ Male Low	DDD
Sweden	-0.94	-49.61*	13.16*	-21.91*	-48.67*	-35.07*	14.10*	27.70*	13.60*
	(4.57)	(5.10)	(4.58)	(4.79)	(3.13)	(2.94)	(2.55)	(3.05)	(3.87)
Turkey	77.81*	42.58*	98.49*	65.27*	-35.22*	-33.22*	20.68*	22.68*	2.00
	(4.13)	(4.07)	(5.09)	(4.95)	(4.24)	(4.42)	(4.28)	(3.61)	(4.34)
USA	21.43*	18.38*	38.11*	40.73*	-3.05	2.62	16.67*	22.35*	5.68
	(3.31)	(4.51)	(3.46)	(3.96)	(3.81)	(3.79)	(2.31)	(3.23)	(3.84)

Standard errors in parentheses
* $p < 0.05$

Notes: Source: TIMSS © IEA 1995-2019. Standard errors are in parentheses and * means the significance at the 5 percent level or better. The table shows the point estimates of the linear combinations of the estimates as written in the second row. Students with high socioeconomic status are defined as those with SES index bigger or equal to top 50% student's SES index and low SES is defined as the rest. Columns (1)-(4) show the difference in scores over time in each group and (5)-(8) show the Difference in differences results to show whether the high SES led the change when compared to low SES (Column (5) and (6)) and whether females led the change when compared to males (Column (7) and (8)). In the regression, jackknife repeated replication (JRR) standard errors are used, and covariates include indicator variables for grade, age and age^2. Sample sizes for each country are reported in Table 1.

Appendix A

Appendix for Chapter 1

Figure A1: Distribution of birth months and assigned relative age/observed age by month of birth in developed countries

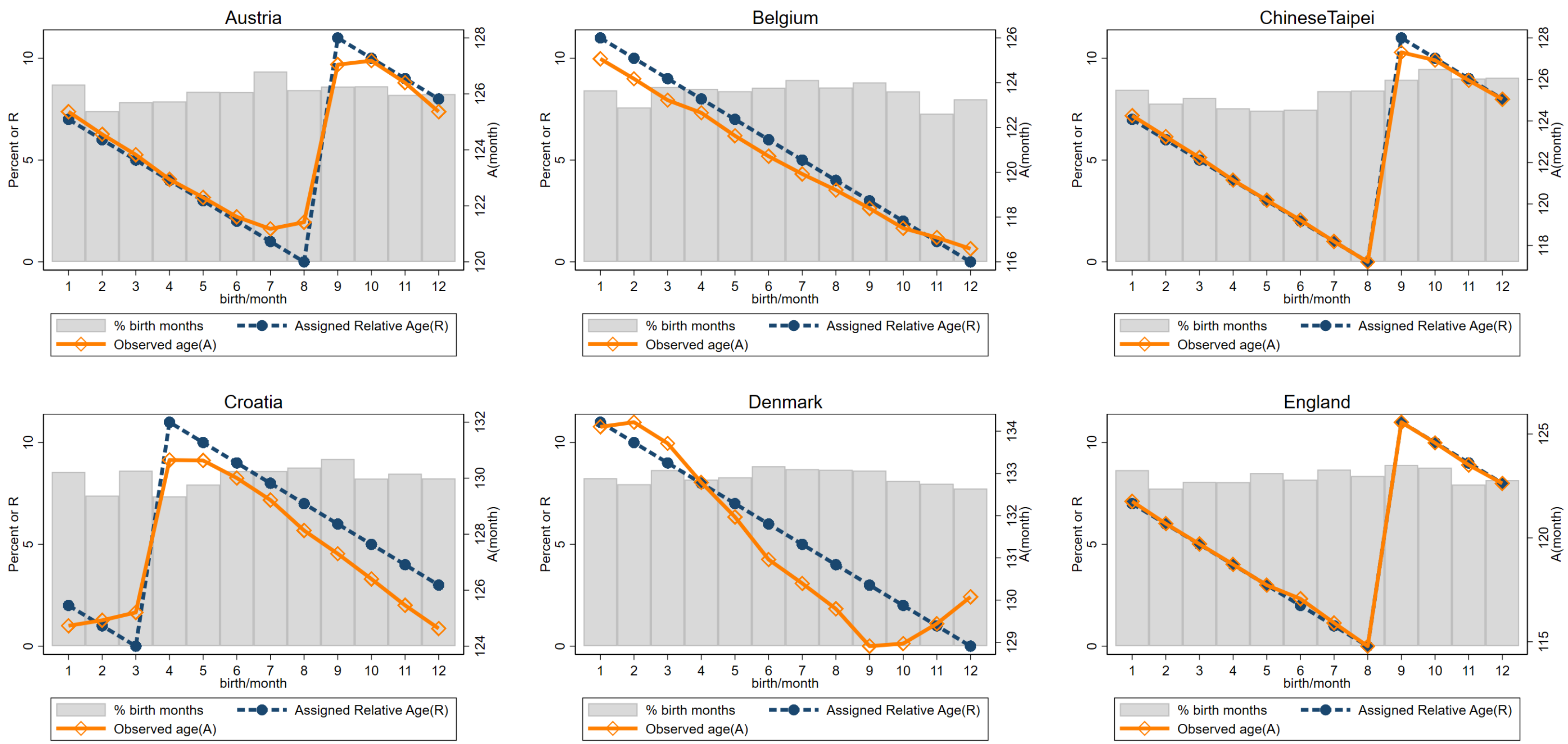


FIGURE A1: DISTRIBUTION OF BIRTH MONTHS AND ASSIGNED RELATIVE AGE/OBSERVED AGE BY MONTH OF BIRTH IN DEVELOPED COUNTRIES(CONTINUED)

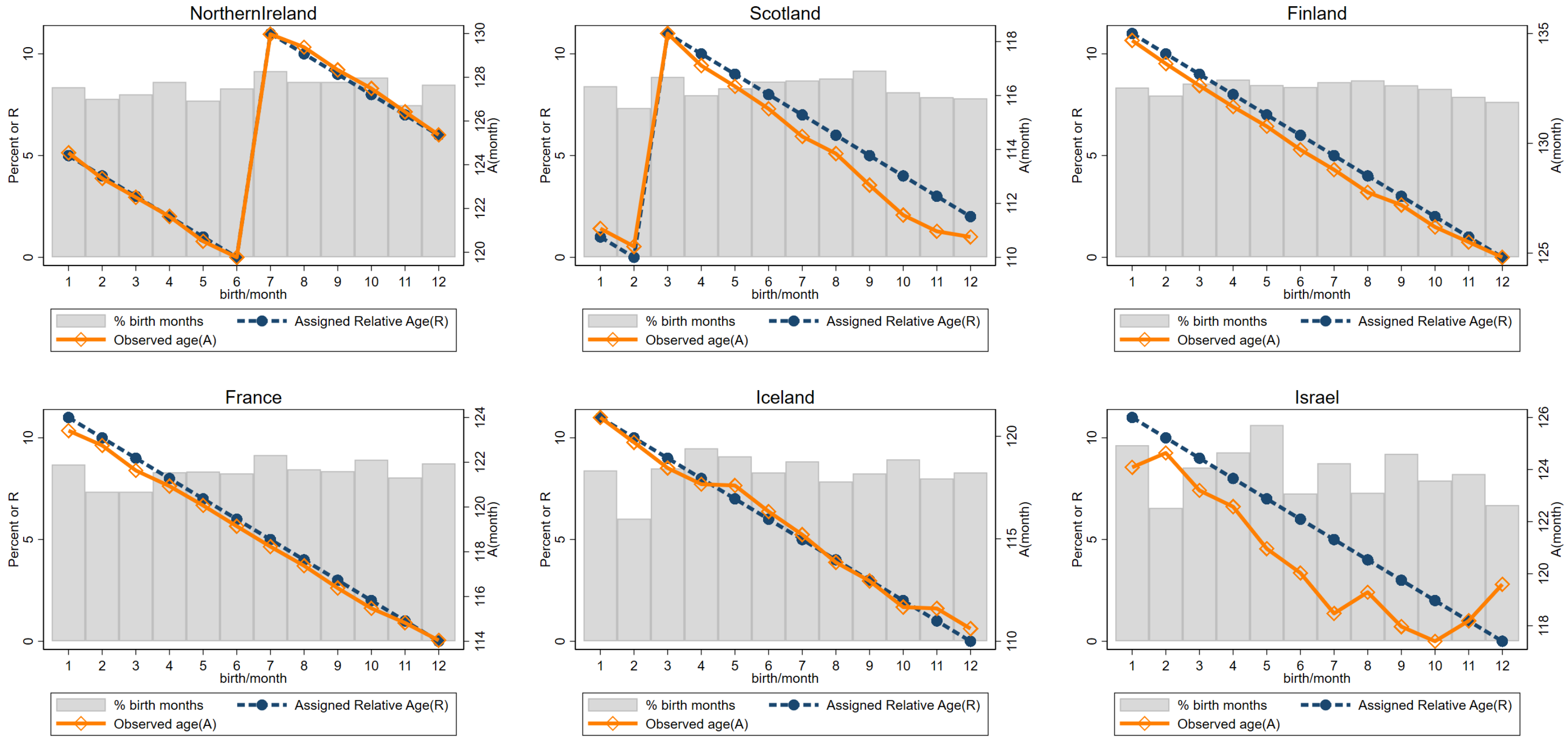


FIGURE A1: DISTRIBUTION OF BIRTH MONTHS AND ASSIGNED RELATIVE AGE/OBSERVED AGE BY MONTH OF BIRTH IN DEVELOPED COUNTRIES(CONTINUED)

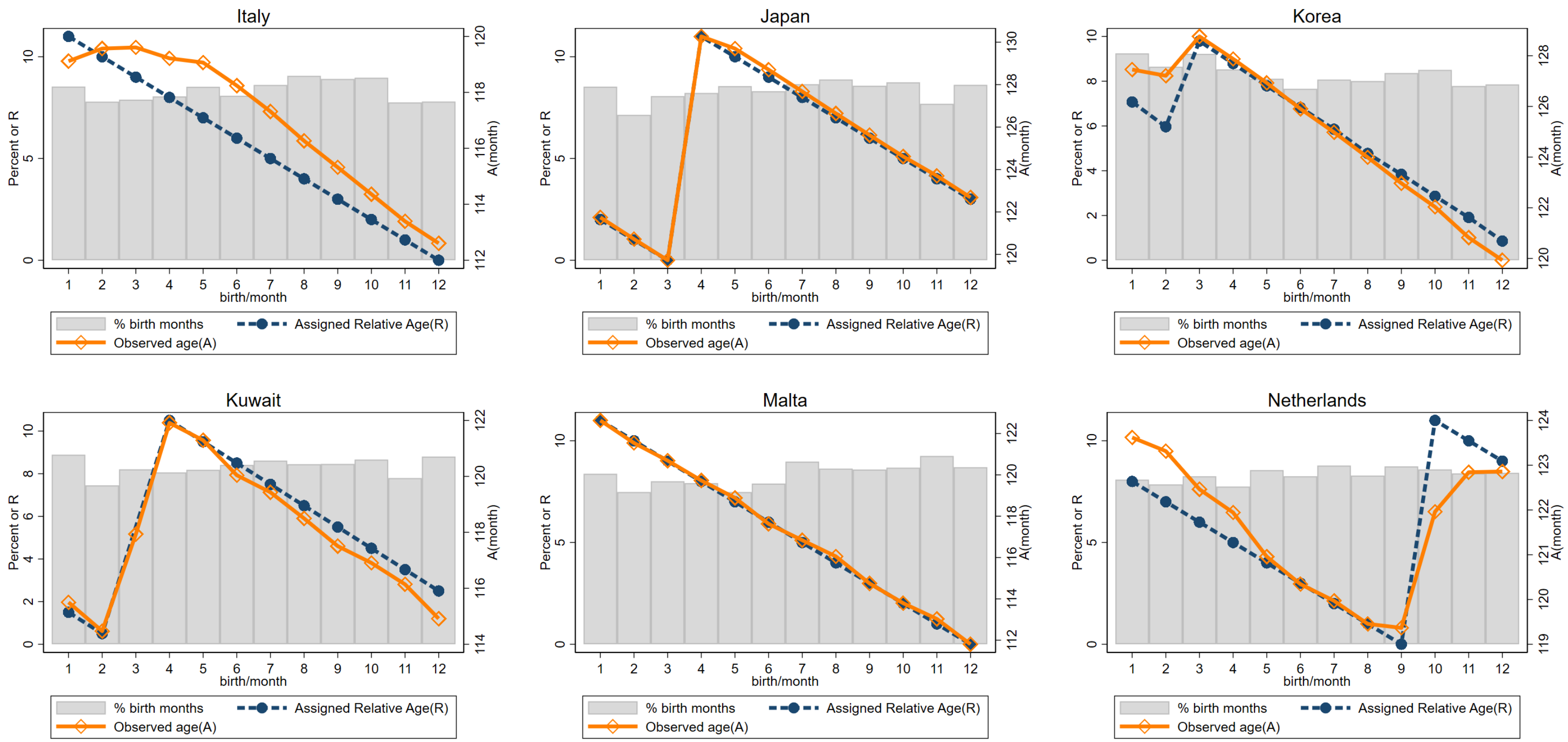


FIGURE A1: DISTRIBUTION OF BIRTH MONTHS AND ASSIGNED RELATIVE AGE/OBSERVED AGE BY MONTH OF BIRTH IN DEVELOPED COUNTRIES(CONTINUED)

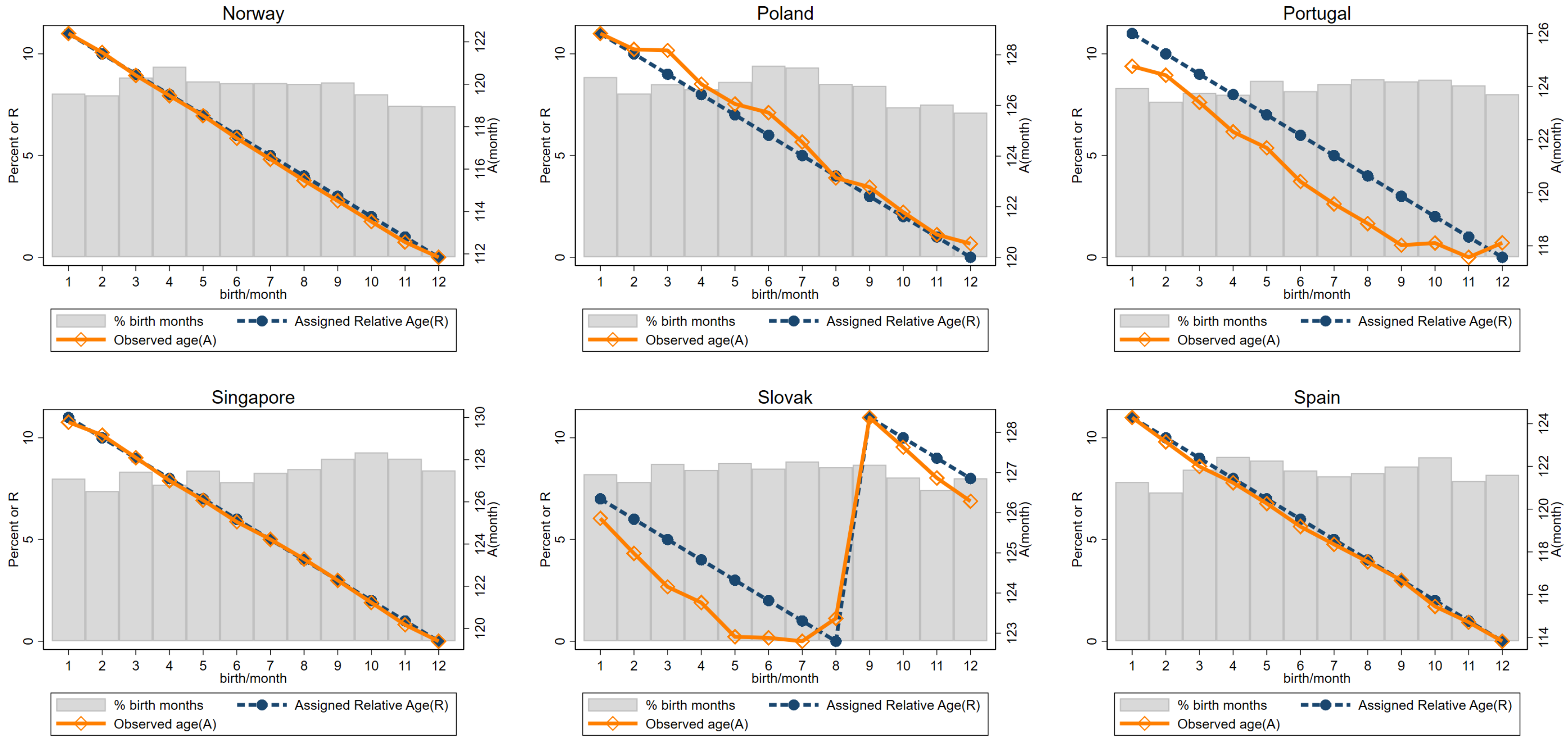
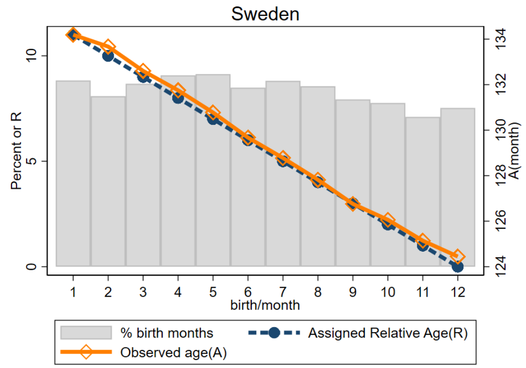
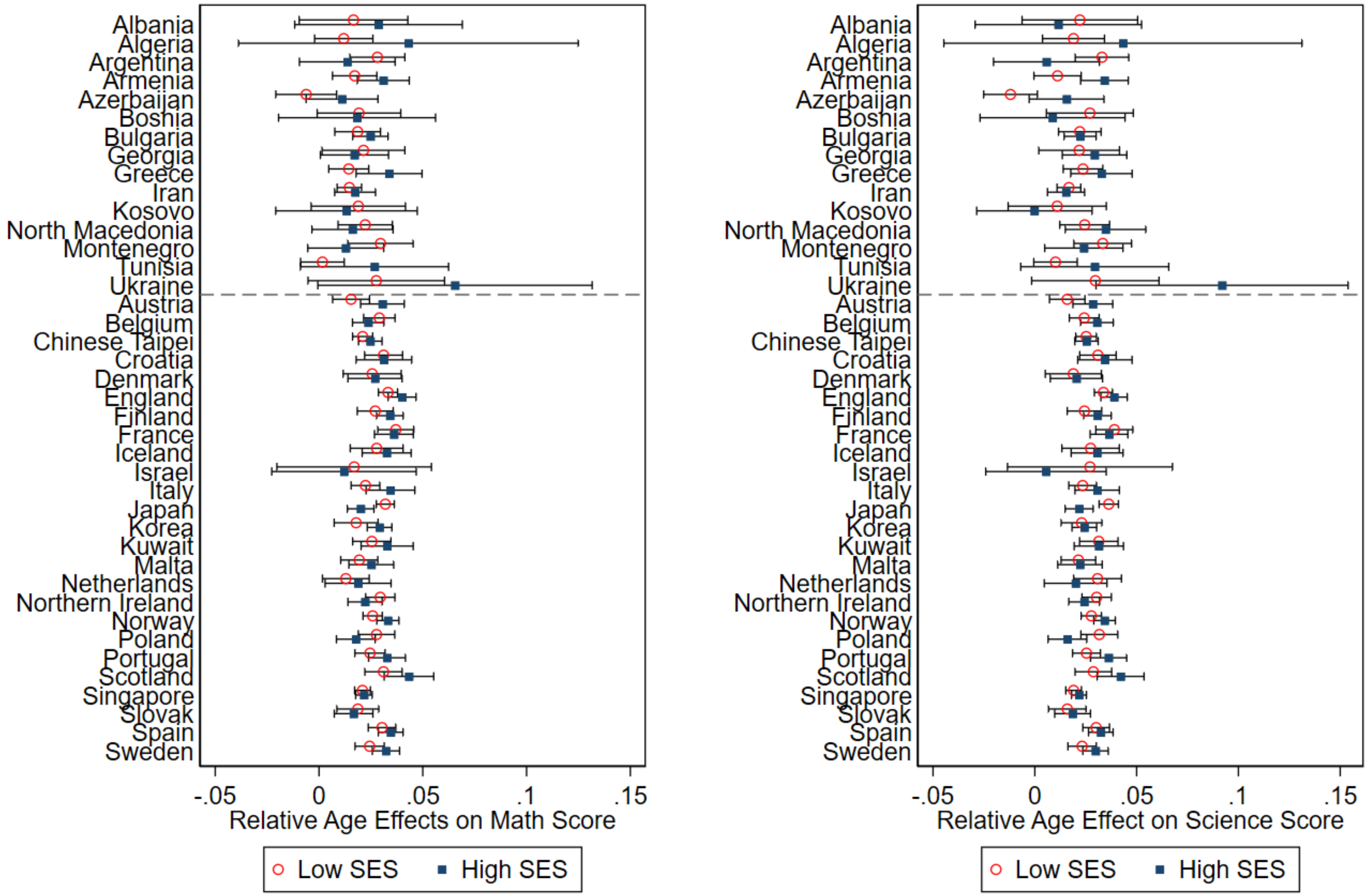


FIGURE A1: DISTRIBUTION OF BIRTH MONTHS AND ASSIGNED RELATIVE AGE/OBSERVED AGE BY MONTH OF BIRTH IN DEVELOPED COUNTRIES(CONTINUED)



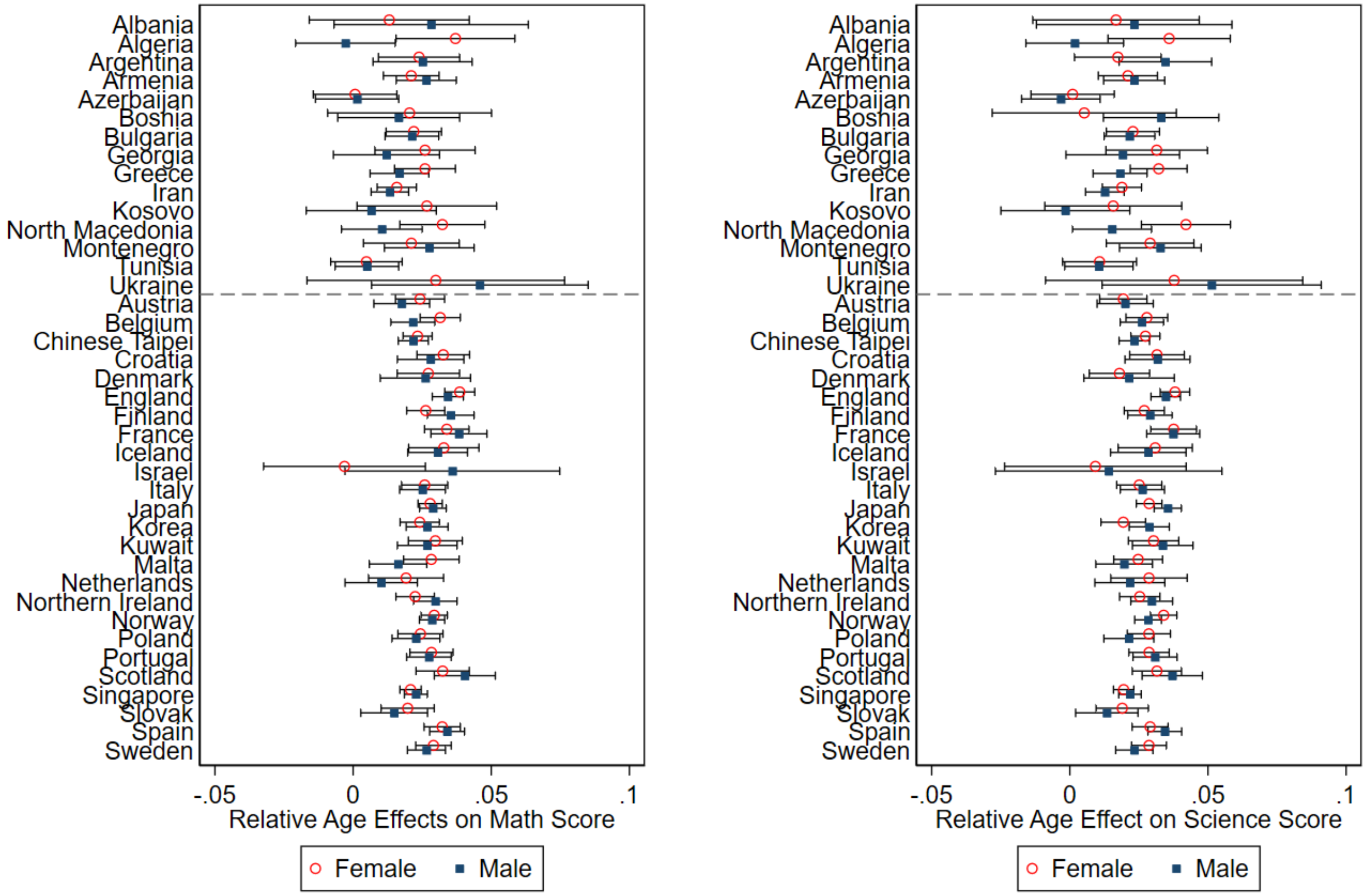
Notes: Source: TIMSS © IEA 1995-2019. Assigned relative age R_{it} is decided by each country's cutoff. A student who is the youngest kid in the cohort is assigned $R = 0$, and who is the oldest kid is assigned $R = 11$. Only grade 4 is included. Data points for Korea include years after 2015 and those for Hungary include years until 2015 due to the change of school entry cutoff.

FIGURE A2: DIFFERENCE IN AGE EFFECTS BY SOCIOECONOMIC STATUS IN DEVELOPING AND DEVELOPED COUNTRIES



Notes: Source: TIMSS © IEA 1995-2019. 95% Confidence Interval.

FIGURE A3: DIFFERENCE IN AGE EFFECTS BY GENDER IN DEVELOPING AND DEVELOPED COUNTRIES



Notes: Source: TIMSS © IEA 1995-2019. 95% Confidence Interval.

TABLE A1-1: PROPORTIONS OF BIRTHS IN EACH QUARTER AND SOCIOECONOMIC STATUS DIFFERENCES IN DEVELOPING COUNTRIES

	Proportions of births in each quarter				Socioeconomic status difference (HighSES-LowSES)			
Country	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
Albania	0.266	0.227	0.223	0.284	0.012	0.030	-0.039*	-0.003
Algeria	0.290	0.247	0.249	0.214	-0.001	0.065*	-0.017	-0.047
Argentina	0.254	0.253	0.247	0.246	-0.028	-0.010	-0.013	0.052*
Armenia	0.227	0.218	0.286	0.269	0.005	0.013	0.004	-0.023*
Azerbaijan	0.245	0.256	0.222	0.278	-0.004	-0.009	0.006	0.007
Bosnia	0.245	0.244	0.239	0.272	-0.015	0.002	0.002	0.011
Bulgaria	0.239	0.248	0.261	0.252	-0.008	0.004	0.002	0.002
Georgia	0.250	0.252	0.261	0.238	-0.009	0.014	-0.010	0.005
Greece	0.272	0.279	0.240	0.210	0.018	-0.015	-0.009	0.006
Iran	0.231	0.249	0.272	0.248	0.011	-0.012	-0.012	0.013
Kosovo	0.235	0.254	0.282	0.228	0.020	0.005	-0.014	-0.011
Macedonia	0.247	0.235	0.269	0.248	-0.030	-0.015	0.015	0.031
Montenegro	0.233	0.240	0.268	0.259	-0.025	0.010	-0.001	0.016
Tunisia	0.255	0.272	0.256	0.216	-0.035*	0.016	-0.005	0.023
Ukraine	0.252	0.233	0.252	0.263	-0.048*	0.006	0.023	0.019

Notes: Source: TIMSS © IEA 1995-2019. Q1 is the first three months after the school entry cutoff, and Q4 is the three months before the cutoff. Therefore, the oldest children are born in Q1, and the youngest is born in Q4 in each country. Since the purpose of this analysis is to see whether birth proportions are equally distributed every three months, September and October/November are included as Q1 for Iran and Azerbaijan, respectively. All proportions are population weighted. High SES is defined as those with either parental education is at least a bachelor's degree holder or the number of books in the household is more or equal to 100. The starred coefficients indicate that the fraction of births to high SES children differs from the fraction of births to low SES children at the 5 percent level or better.

TABLE A1-2: PROPORTIONS OF BIRTHS IN EACH QUARTER AND SOCIOECONOMIC STATUS DIFFERENCES IN DEVELOPED COUNTRIES

	Proportions of births in each quarter				Socioeconomic status difference (HighSES-LowSES)			
Country	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
Austria	0.224	0.256	0.267	0.252	-0.003	0.005	-0.006	0.005
Belgium	0.247	0.248	0.269	0.235	0.014*	0.002	-0.003	-0.013
Chinese Taipei	0.264	0.251	0.232	0.253	-0.006	0.012	-0.001	-0.005
Croatia	0.241	0.270	0.248	0.240	0.001	-0.005	-0.009	0.014
Denmark	0.259	0.243	0.265	0.233	-0.005	0.018*	-0.010	-0.002
England	0.263	0.249	0.251	0.237	0.008	0.013*	-0.006	-0.015*
Finland	0.250	0.271	0.237	0.242	0.006	0.005	-0.006	-0.005
France	0.236	0.249	0.259	0.256	0.006	-0.005	-0.003	0.002
Iceland	0.224	0.270	0.261	0.245	-0.000	0.018	0.004	-0.021
Israel	0.247	0.272	0.253	0.228	0.042*	-0.018	-0.014	-0.010
Italy	0.236	0.248	0.262	0.255	-0.002	-0.004	-0.005	0.011
Japan	0.253	0.272	0.240	0.235	0.007	-0.007	0.000	0.000
Korea	0.215	0.242	0.253	0.289	0.014	-0.002	-0.000	-0.011
Kuwait	0.239	0.262	0.250	0.249	-0.002	-0.017*	0.002	0.018*
Malta	0.247	0.250	0.255	0.248	0.020	-0.012	-0.002	-0.005
Netherlands	0.252	0.252	0.246	0.250	0.004	0.003	-0.003	-0.004
Northern Ireland	0.263	0.238	0.249	0.249	0.012	-0.001	0.004	-0.015
Norway	0.248	0.256	0.260	0.236	0.010	-0.003	-0.004	-0.003
Poland	0.256	0.262	0.258	0.224	-0.015	0.007	0.009	-0.001
Portugal	0.234	0.267	0.264	0.235	-0.003	0.015*	-0.008	-0.004
Scotland	0.247	0.254	0.257	0.242	-0.001	0.013	-0.010	-0.003
Singapore	0.231	0.238	0.259	0.271	-0.001	-0.006	-0.007	0.013*
Slovak	0.239	0.235	0.261	0.265	-0.005	-0.003	0.003	0.005
Spain	0.236	0.268	0.237	0.259	0.004	0.004	-0.009	0.001
Sweden	0.257	0.271	0.252	0.220	-0.001	0.007	-0.001	-0.004

Notes: Source: TIMSS © IEA 1995-2019. Q1 is the first three months after the school entry cutoff, and Q4 is the three months before the cutoff. Therefore, the oldest children are born in Q1, and the youngest is born in Q4 in each country. All proportions are population weighted. High SES is defined as those with either parental education is at least a bachelor's degree holder or the number of books in the household is more or equal to 100. The starred coefficients indicate that the fraction of births to high SES children differs from the fraction of births to low SES children at the 5 percent level or better.

TABLE A2-1: MONOTONICITY FOR DEVELOPING COUNTRIES

Country	Male × High SES	Female × High SES	Male × Low SES	Female × Low SES
Albania	0.397*	0.559*	0.484*	0.451*
	(0.061)	(0.053)	(0.034)	(0.034)
Algeria	0.347*	0.192	0.358*	0.324*
	(0.166)	(0.116)	(0.034)	(0.037)
Argentina	0.912*	0.869*	0.817*	0.852*
	(0.036)	(0.039)	(0.031)	(0.028)
Armenia	0.844*	0.798*	0.760*	0.702*
	(0.017)	(0.021)	(0.019)	(0.019)
Azerbaijan	0.588*	0.600*	0.565*	0.565*
	(0.040)	(0.041)	(0.031)	(0.029)
Bosnia	0.538*	0.424*	0.573*	0.438*
	(0.041)	(0.038)	(0.025)	(0.032)
Bulgaria	0.837*	0.793*	0.842*	0.767*
	(0.017)	(0.022)	(0.017)	(0.022)
Georgia	0.528*	0.504*	0.435*	0.418*
	(0.022)	(0.026)	(0.023)	(0.025)
Greece	0.922*	0.902*	0.902*	0.875*
	(0.024)	(0.027)	(0.018)	(0.018)
Iran	0.894*	0.933*	0.846*	0.877*
	(0.016)	(0.013)	(0.009)	(0.010)
Kosovo	0.506*	0.602*	0.522*	0.514*
	(0.046)	(0.047)	(0.030)	(0.033)
Macedonia	0.797*	0.780*	0.837*	0.843*
	(0.053)	(0.046)	(0.025)	(0.025)
Montenegro	0.727*	0.741*	0.707*	0.690*
	(0.035)	(0.039)	(0.027)	(0.026)
Tunisia	0.311*	0.401*	0.388*	0.410*
	(0.065)	(0.052)	(0.023)	(0.023)
Ukraine	0.436*	0.439*	0.438*	0.410*
	(0.059)	(0.055)	(0.029)	(0.032)
% Observation	15.2	14.6	34.7	31.6

Robust standard errors in parentheses and clustered at the school level.
* $p < 0.05$

Notes: Source: TIMSS © IEA 1995-2019. The estimates show the results from First-stage regression using assigned relative age as an instrument for each group. High SES is defined as those with either parental education is at least a bachelor's degree holder or the number of books in the household is more or equal to 100. Low SES is non-High SES. All regressions are population-weighted and include controls for students' gender, the index for books, the index for desks, the index for parental education, and year and grade fixed effects.

TABLE A2-2: MONOTONICITY FOR DEVELOPED COUNTRIES

Country	Male × High SES	Female × High SES	Male × Low SES	Female × Low SES
Austria	0.656*	0.700*	0.555*	0.642*
	(0.020)	(0.018)	(0.015)	(0.015)
Belgium	0.844*	0.840*	0.666*	0.725*
	(0.013)	(0.014)	(0.017)	(0.015)
ChineseTaipei	0.968*	0.968*	0.956*	0.954*
	(0.006)	(0.006)	(0.005)	(0.006)
Croatia	0.659*	0.714*	0.679*	0.741*
	(0.023)	(0.022)	(0.016)	(0.016)
Denmark	0.581*	0.734*	0.484*	0.631*
	(0.021)	(0.019)	(0.019)	(0.018)
England	0.974*	0.976*	0.968*	0.983*
	(0.006)	(0.005)	(0.004)	(0.004)
Finland	0.920*	0.945*	0.841*	0.898*
	(0.009)	(0.008)	(0.017)	(0.013)
France	0.881*	0.887*	0.802*	0.863*
	(0.017)	(0.016)	(0.018)	(0.015)
Iceland	0.964*	0.961*	0.984*	0.958*
	(0.020)	(0.013)	(0.009)	(0.019)
Israel	0.710*	0.721*	0.519*	0.635*
	(0.070)	(0.053)	(0.058)	(0.046)
Italy	0.725*	0.683*	0.765*	0.711*
	(0.020)	(0.020)	(0.013)	(0.015)
Japan	0.985*	0.985*	0.978*	0.993*
	(0.004)	(0.005)	(0.003)	(0.002)
Korea	0.812*	0.844*	0.745*	0.762*
	(0.013)	(0.011)	(0.018)	(0.018)
Kuwait	0.814*	0.852*	0.810*	0.828*
	(0.022)	(0.023)	(0.021)	(0.018)
Malta	0.942*	0.936*	0.915*	0.928*
	(0.015)	(0.015)	(0.014)	(0.015)
Netherlands	0.436*	0.440*	0.434*	0.435*
	(0.022)	(0.025)	(0.017)	(0.015)
NorthernIreland	0.974*	0.974*	0.942*	0.951*
	(0.008)	(0.007)	(0.010)	(0.009)
Norway	0.976*	0.972*	0.976*	0.975*
	(0.004)	(0.005)	(0.004)	(0.004)
Poland	0.735*	0.742*	0.819*	0.843*
	(0.025)	(0.026)	(0.020)	(0.018)
Portugal	0.724*	0.751*	0.515*	0.561*
	(0.020)	(0.020)	(0.019)	(0.018)
Scotland	0.731*	0.775*	0.746*	0.830*
	(0.020)	(0.019)	(0.019)	(0.017)
Singapore	0.950*	0.959*	0.939*	0.947*
	(0.005)	(0.005)	(0.004)	(0.004)
Slovak	0.589*	0.695*	0.513*	0.633*

Continued on next page

Table A2-2 – *Continued from previous page*

Country	Male × High SES	Female × High SES	Male × Low SES	Female × Low SES
	(0.020)	(0.016)	(0.016)	(0.016)
Spain	0.936*	0.950*	0.849*	0.854*
	(0.008)	(0.008)	(0.011)	(0.012)
Sweden	0.930*	0.928*	0.909*	0.925*
	(0.009)	(0.009)	(0.010)	(0.010)
% Observation	20.9	20.7	28.0	27.1

Robust standard errors in parentheses and clustered at the school level.

* $p < 0.5$

Notes: Source: TIMSS © IEA 1995-2019. The second row for each country shows F-statistics for each first stage. High SES is defined as those with either parental education is at least a bachelor's degree holder or the number of books in the household is more or equal to 100. Low SES is non-High SES. All regressions are population-weighted and include controls for students' gender, the index for books, the index for desks, the index for parental education, and year and grade fixed effects.

TABLE A3: NONLINEAR SPECIFICATION

Developing	**Math** 2SLS	J-statistics	P-value	**Science** 2SLS	J-statistics	P-value
Albania	0.018*	1.43	0.49	0.021**	0.42	0.81
	(0.009)			(0.010)		
Algeria	0.013**	2.01	0.37	0.014**	3.10	0.21
	(0.007)			(0.007)		
Argentina	0.026***	0.66	0.72	0.027***	5.39	0.07
	(0.005)			(0.006)		
Armenia	0.025***	3.92	0.14	0.022***	2.11	0.35
	(0.004)			(0.004)		
Azerbaijan	0.003	1.19	0.55	0.003	0.10	0.95
	(0.006)			(0.006)		
Bosnia	0.019***	0.03	0.99	0.021***	0.18	0.91
	(0.007)			(0.007)		
Bulgaria	0.022***	0.84	0.66	0.023***	0.47	0.79
	(0.004)			(0.004)		
Georgia	0.016***	1.40	0.50	0.022***	1.60	0.45
	(0.006)			(0.006)		
Greece	0.021***	3.37	0.19	0.025***	2.37	0.31
	(0.004)			(0.004)		
Iran	0.014***	3.87	0.14	0.015***	3.15	0.21
	(0.002)			(0.002)		
Kosovo	0.016*	0.25	0.88	0.007	0.15	0.93
	(0.008)			(0.008)		
Macedonia	0.021***	0.71	0.70	0.029***	1.45	0.48
	(0.005)			(0.005)		
Montenegro	0.028***	10.19	0.01	0.032***	3.08	0.21
	(0.006)			(0.005)		
Tunisia	0.010**	19.28	0.00	0.016***	16.65	0.00
	(0.005)			(0.005)		
Ukraine	0.035***	0.35	0.84	0.037***	1.37	0.50
	(0.011)			(0.011)		

Robust standard errors in parentheses and clustered at the school level.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: Source: TIMSS © IEA 1995-2019. All regressions are population-weighted and include controls for students' gender, the index for books, the index for desks, the index for parental education, and year and grade fixed effects. The data points for students born in September in Iran and October/November in Azerbaijan are excluded from the 2SLS analysis as they are not assigned with IV values. J-statistics are from the Sargan-Hansen test of over-identifying restriction.

TABLE A4: RESTRICTED SAMPLES AROUND THE SCHOOL-ENTRY CUTOFF

	Math			Science		
	(1)	(2)	(3)	(4)	(5)	(6)
Developing	0.021***	0.021***	0.021***	0.021***	0.023***	0.023***
	(0.004)	(0.003)	(0.003)	(0.004)	(0.003)	(0.002)
F-statistics	1,467	2,643	3,533	1,467	2,643	3,533
Observation	9,031	17,298	25,922	9,031	17,298	25,922
Developed	0.025***	0.027***	0.027***	0.024***	0.025***	0.026***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
F-statistics	25,664	43,093	51,406	25,664	43,093	51,406
Observation	44,521	87,906	133,593	44,521	87,906	133,593
Difference	-0.005	-0.006	-0.006*	-0.003	-0.003	-0.002
	(0.004)	(0.003)	(0.003)	(0.004)	(0.003)	(0.003)
F-statistics	13,207	22,394	27,150	13,207	22,394	27,150
Observation	53,552	105,204	159,515	53,552	105,204	159,515
Window of 1 month	✓			✓		
Window of 2 months		✓			✓	
Window of 3 months			✓			✓

Robust standard errors in parentheses and clustered at the school level.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: Source: TIMSS © IEA 1995-2019. The estimates show the results from 2SLS regression using assigned relative age as an instrument. All regressions are population-weighted and include controls for the child's gender, the index for books, the index for desk, the index for parental education, year, grade, and country fixed effects. Column (1) and (4) restricts the sample to children who were born in 1 month right after and before the school-entry cutoff date. Column (2) and (5) restricts the sample to children who were born in 2 months right after and before the school-entry cutoff date. Column (3) and (6) restricts the sample to children who were born within three months right after and before the school-entry cutoff date. Iran and Azerbaijan are excluded from this analysis since data points near to the cutoff are not assigned IV values. F-statistics are from the weak instrument test in the first stage.

TABLE A5-1: HETEROGENEITY IN RELATIVE AGE EFFECTS BY SOCIOECONOMIC STATUS IN DEVELOPING COUNTRIES

	Math			**Science**		
	Low SES	High SES	Δ(L-H)	Low SES	High SES	Δ(L-H)
Albania	0.017	0.029	-0.012	0.022	0.012	0.011
	(0.013)	(0.021)	(0.024)	(0.014)	(0.021)	(0.025)
Algeria	0.012*	0.043	-0.031	0.019**	0.043	-0.024
	(0.007)	(0.042)	(0.042)	(0.008)	(0.045)	(0.045)
Argentina	0.028***	0.014	0.015	0.033***	0.006	0.027*
	(0.007)	(0.012)	(0.015)	(0.007)	(0.013)	(0.016)
Armenia	0.017***	0.031***	-0.014	0.011*	0.034***	-0.023***
	(0.005)	(0.006)	(0.009)	(0.006)	(0.006)	(0.009)
Azerbaijan	-0.006	0.011	-0.017	-0.012*	0.015	-0.027**
	(0.007)	(0.009)	(0.012)	(0.007)	(0.009)	(0.012)
Bosnia	0.019*	0.018	0.001	0.027**	0.009	0.018
	(0.010)	(0.019)	(0.023)	(0.011)	(0.018)	(0.022)
Bulgaria	0.019***	0.025***	-0.006	0.022***	0.022***	-0.000
	(0.006)	(0.004)	(0.007)	(0.005)	(0.004)	(0.007)
Georgia	0.021**	0.017**	0.004	0.022**	0.029***	-0.008
	(0.010)	(0.008)	(0.013)	(0.010)	(0.008)	(0.012)
Greece	0.014***	0.034***	-0.020**	0.024***	0.033***	-0.009
	(0.005)	(0.009)	(0.010)	(0.005)	(0.008)	(0.010)
Iran	0.015***	0.017***	-0.003	0.017***	0.015***	0.001
	(0.003)	(0.005)	(0.006)	(0.003)	(0.005)	(0.005)
Kosovo	0.019	0.013	0.006	0.011	-0.000	0.011
	(0.012)	(0.017)	(0.022)	(0.012)	(0.014)	(0.020)
Macedonia	0.022***	0.016	0.006	0.024***	0.035***	-0.010
	(0.007)	(0.010)	(0.012)	(0.006)	(0.010)	(0.012)
Montenegro	0.030***	0.013	0.017	0.033***	0.024**	0.009
	(0.008)	(0.009)	(0.013)	(0.007)	(0.010)	(0.013)
Tunisia	0.002	0.027	-0.025	0.010*	0.030	-0.019
	(0.005)	(0.018)	(0.019)	(0.005)	(0.018)	(0.019)
Ukraine	0.028*	0.066*	-0.038	0.030*	0.092***	-0.062*
	(0.017)	(0.034)	(0.036)	(0.016)	(0.032)	(0.034)

Robust standard errors in parentheses and clustered at the school level.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: Source: TIMSS © IEA 1995-2019. All regressions are population-weighted and include controls for students' gender, the index for books, the index for desks, the index for parental education, and year and grade fixed effects. High SES is defined as those with either parental education is at least a bachelor's degree holder or the number of books in the household is more or equal to 100. Low SES is defined as non-high SES.

TABLE A5-2: HETEROGENEITY IN RELATIVE AGE EFFECTS BY SOCIOECONOMIC STATUS IN DEVELOPED COUNTRIES

	Math			Science		
	Low SES	High SES	Δ(L-H)	Low SES	High SES	Δ(L-H)
Austria	0.015***	0.031***	-0.015**	0.016***	0.029***	-0.013**
	(0.004)	(0.005)	(0.007)	(0.004)	(0.005)	(0.006)
Belgium	0.029***	0.024***	0.005	0.024***	0.031***	-0.006
	(0.004)	(0.004)	(0.005)	(0.004)	(0.004)	(0.006)
ChineseTaipei	0.021***	0.025***	-0.003	0.025***	0.025***	0.000
	(0.002)	(0.003)	(0.004)	(0.003)	(0.003)	(0.004)
Croatia	0.031***	0.031***	0.000	0.031***	0.034***	-0.003
	(0.005)	(0.007)	(0.008)	(0.005)	(0.007)	(0.008)
Denmark	0.026***	0.027***	-0.002	0.019***	0.021***	-0.002
	(0.007)	(0.007)	(0.010)	(0.007)	(0.007)	(0.010)
England	0.033***	0.040***	-0.007*	0.034***	0.039***	-0.005
	(0.002)	(0.003)	(0.004)	(0.002)	(0.003)	(0.004)
Finland	0.027***	0.034***	-0.007	0.024***	0.031***	-0.007
	(0.004)	(0.003)	(0.005)	(0.004)	(0.003)	(0.005)
France	0.037***	0.036***	0.001	0.039***	0.036***	0.003
	(0.004)	(0.005)	(0.006)	(0.005)	(0.005)	(0.007)
Iceland	0.028***	0.033***	-0.005	0.027***	0.031***	-0.003
	(0.007)	(0.006)	(0.010)	(0.007)	(0.007)	(0.009)
Israel	0.017	0.012	0.005	0.027	0.005	0.022
	(0.019)	(0.018)	(0.025)	(0.021)	(0.015)	(0.026)
Italy	0.022***	0.035***	-0.013*	0.024***	0.031***	-0.008
	(0.004)	(0.006)	(0.007)	(0.003)	(0.006)	(0.007)
Japan	0.032***	0.020***	0.012***	0.036***	0.022***	0.015***
	(0.002)	(0.003)	(0.004)	(0.002)	(0.004)	(0.004)
Korea	0.017***	0.029***	-0.012**	0.022***	0.025***	-0.003
	(0.005)	(0.003)	(0.006)	(0.005)	(0.003)	(0.006)
Kuwait	0.025***	0.033***	-0.007	0.031***	0.031***	0.000
	(0.005)	(0.006)	(0.008)	(0.005)	(0.006)	(0.007)
Malta	0.020***	0.025***	-0.005	0.022***	0.022***	-0.000
	(0.005)	(0.005)	(0.007)	(0.004)	(0.006)	(0.007)
Netherlands	0.013**	0.019**	-0.005	0.031***	0.019**	0.012
	(0.006)	(0.008)	(0.010)	(0.006)	(0.008)	(0.010)

Continued on next page

Table A5-2 – *Continued from previous page*

	Math			Science		
	Low SES	High SES	Δ(L-H)	Low SES	High SES	Δ(L-H)
NorthernIreland	0.030***	0.022***	0.007	0.030***	0.024***	0.006
	(0.004)	(0.004)	(0.005)	(0.004)	(0.004)	(0.005)
Norway	0.026***	0.033***	-0.007**	0.028***	0.034***	-0.007*
	(0.002)	(0.003)	(0.004)	(0.003)	(0.003)	(0.004)
Poland	0.028***	0.018***	0.010	0.032***	0.016***	0.016**
	(0.004)	(0.005)	(0.007)	(0.005)	(0.005)	(0.007)
Portugal	0.025***	0.033***	-0.008	0.025***	0.036***	-0.011*
	(0.004)	(0.005)	(0.006)	(0.003)	(0.005)	(0.006)
Scotland	0.031***	0.043***	-0.012*	0.029***	0.042***	-0.013*
	(0.005)	(0.006)	(0.007)	(0.005)	(0.006)	(0.007)
Singapore	0.021***	0.022***	-0.001	0.019***	0.022***	-0.003
	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)
Slovak	0.019***	0.017***	0.002	0.016***	0.018***	-0.003
	(0.005)	(0.005)	(0.007)	(0.005)	(0.005)	(0.006)
Spain	0.030***	0.035***	-0.004	0.030***	0.032***	-0.002
	(0.003)	(0.003)	(0.004)	(0.003)	(0.003)	(0.005)
Sweden	0.024***	0.032***	-0.008*	0.023***	0.030***	-0.007
	(0.004)	(0.003)	(0.005)	(0.004)	(0.003)	(0.005)

Robust standard errors in parentheses and clustered at the school level.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: Source: TIMSS © IEA 1995-2019. All regressions are population-weighted and include controls for students' gender, the index for books, the index for desks, the index for parental education, and year and grade fixed effects. High SES is defined as those with either parental education is at least a bachelor's degree holder or the number of books in the household is more or equal to 100. Low SES is defined as non-high SES.

TABLE A6-1: HETEROGENEITY IN RELATIVE AGE EFFECTS BY GENDER IN DEVELOPING COUNTRIES

	Math			Science		
	Female	Male	Δ(F-M)	Female	Male	Δ(F-M)
Albania	0.013	0.028	-0.015	0.017	0.023	-0.007
	(0.015)	(0.018)	(0.023)	(0.015)	(0.018)	(0.023)
Algeria	0.037***	-0.003	0.040***	0.036***	0.002	0.034**
	(0.011)	(0.009)	(0.015)	(0.011)	(0.009)	(0.015)
Argentina	0.024***	0.025***	-0.001	0.017**	0.035***	-0.017
	(0.007)	(0.009)	(0.013)	(0.008)	(0.009)	(0.013)
Armenia	0.021***	0.026***	-0.005	0.021***	0.023***	-0.002
	(0.005)	(0.006)	(0.007)	(0.005)	(0.006)	(0.007)
Azerbaijan	0.001	0.001	-0.001	0.001	-0.003	0.004
	(0.008)	(0.008)	(0.011)	(0.008)	(0.007)	(0.011)
Bosnia	0.020	0.016	0.004	0.005	0.033***	-0.028
	(0.015)	(0.011)	(0.020)	(0.017)	(0.011)	(0.021)
Bulgaria	0.022***	0.021***	0.001	0.023***	0.022***	0.001
	(0.005)	(0.005)	(0.007)	(0.005)	(0.005)	(0.007)
Georgia	0.026***	0.012	0.014	0.032***	0.019*	0.012
	(0.009)	(0.010)	(0.014)	(0.009)	(0.011)	(0.015)
Greece	0.026***	0.017***	0.009	0.033***	0.018***	0.015**
	(0.006)	(0.006)	(0.008)	(0.005)	(0.005)	(0.007)
Iran	0.016***	0.013***	0.003	0.019***	0.013***	0.006
	(0.004)	(0.003)	(0.005)	(0.004)	(0.004)	(0.005)
Kosovo	0.027**	0.007	0.020	0.016	-0.002	0.017
	(0.013)	(0.012)	(0.017)	(0.013)	(0.012)	(0.017)
Macedonia	0.032***	0.010	0.022*	0.042***	0.015**	0.027**
	(0.008)	(0.007)	(0.011)	(0.008)	(0.007)	(0.012)
Montenegro	0.021**	0.028***	-0.006	0.029***	0.033***	-0.004
	(0.009)	(0.008)	(0.012)	(0.008)	(0.008)	(0.011)
Tunisia	0.005	0.005	-0.001	0.010	0.011*	-0.000
	(0.007)	(0.006)	(0.009)	(0.007)	(0.006)	(0.009)
Ukraine	0.030	0.046**	-0.016	0.038	0.051**	-0.014
	(0.024)	(0.020)	(0.031)	(0.024)	(0.020)	(0.033)

Robust standard errors in parentheses and clustered at the school level.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: Source: TIMSS © IEA 1995-2019. All regressions are population-weighted and include controls for students' gender, the index for books, the index for desks, the index for parental education, and year and grade fixed effects.

TABLE A6-2: HETEROGENEITY IN RELATIVE AGE EFFECTS BY GENDER IN DEVELOPED COUNTRIES

	Math			Science		
	Female	Male	Δ(F-M)	Female	Male	Δ(F-M)
Austria	0.024***	0.018***	0.006	0.019***	0.020***	-0.001
	(0.005)	(0.005)	(0.006)	(0.004)	(0.005)	(0.006)
Belgium	0.031***	0.022***	0.010*	0.028***	0.026***	0.002
	(0.004)	(0.004)	(0.006)	(0.004)	(0.004)	(0.006)
ChineseTaipei	0.023***	0.022***	0.002	0.027***	0.023***	0.004
	(0.003)	(0.003)	(0.004)	(0.003)	(0.003)	(0.004)
Croatia	0.033***	0.028***	0.005	0.031***	0.032***	-0.000
	(0.005)	(0.006)	(0.008)	(0.005)	(0.006)	(0.008)
Denmark	0.027***	0.026***	0.001	0.018***	0.021***	-0.003
	(0.006)	(0.008)	(0.010)	(0.006)	(0.008)	(0.010)
England	0.039***	0.034***	0.004	0.038***	0.035***	0.003
	(0.003)	(0.003)	(0.004)	(0.003)	(0.003)	(0.004)
Finland	0.027***	0.035***	-0.009	0.027***	0.029***	-0.002
	(0.004)	(0.004)	(0.006)	(0.004)	(0.004)	(0.006)
France	0.034***	0.038***	-0.004	0.038***	0.037***	0.000
	(0.004)	(0.005)	(0.007)	(0.004)	(0.005)	(0.006)
Iceland	0.033***	0.031***	0.002	0.031***	0.028***	0.003
	(0.006)	(0.006)	(0.009)	(0.007)	(0.007)	(0.009)
Israel	-0.003	0.036*	-0.039*	0.009	0.014	-0.005
	(0.015)	(0.020)	(0.023)	(0.017)	(0.021)	(0.027)
Italy	0.026***	0.025***	0.001	0.025***	0.026***	-0.001
	(0.004)	(0.004)	(0.006)	(0.004)	(0.004)	(0.006)
Japan	0.028***	0.029***	-0.001	0.029***	0.035***	-0.007*
	(0.002)	(0.002)	(0.003)	(0.002)	(0.003)	(0.004)
Korea	0.024***	0.026***	-0.002	0.020***	0.028***	-0.008
	(0.004)	(0.004)	(0.005)	(0.004)	(0.004)	(0.005)
Kuwait	0.030***	0.027***	0.003	0.030***	0.033***	-0.003
	(0.005)	(0.005)	(0.007)	(0.005)	(0.006)	(0.007)
Malta	0.028***	0.016***	0.012	0.025***	0.020***	0.005
	(0.005)	(0.005)	(0.008)	(0.005)	(0.005)	(0.007)
Netherlands	0.019***	0.010	0.009	0.029***	0.022***	0.007
	(0.007)	(0.007)	(0.010)	(0.007)	(0.007)	(0.010)

Continued on next page

Table A6-2 – *Continued from previous page*

	Math			Science		
	Female	Male	Δ(F-M)	Female	Male	Δ(F-M)
NorthernIreland	0.022***	0.030***	-0.007	0.025***	0.030***	-0.004
	(0.004)	(0.004)	(0.005)	(0.004)	(0.004)	(0.005)
Norway	0.029***	0.029***	0.001	0.034***	0.028***	0.006*
	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)
Poland	0.024***	0.023***	0.002	0.029***	0.021***	0.007
	(0.004)	(0.004)	(0.006)	(0.004)	(0.005)	(0.006)
Portugal	0.028***	0.027***	0.001	0.029***	0.031***	-0.002
	(0.004)	(0.004)	(0.006)	(0.004)	(0.004)	(0.006)
Scotland	0.033***	0.040***	-0.008	0.032***	0.037***	-0.006
	(0.005)	(0.006)	(0.007)	(0.005)	(0.006)	(0.006)
Singapore	0.021***	0.023***	-0.002	0.020***	0.022***	-0.002
	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)
Slovak	0.020***	0.015**	0.005	0.019***	0.013**	0.005
	(0.005)	(0.006)	(0.008)	(0.005)	(0.006)	(0.008)
Spain	0.032***	0.034***	-0.002	0.029***	0.034***	-0.005
	(0.003)	(0.003)	(0.005)	(0.003)	(0.003)	(0.004)
Sweden	0.029***	0.027***	0.002	0.029***	0.023***	0.005
	(0.003)	(0.004)	(0.005)	(0.003)	(0.003)	(0.004)

Robust standard errors in parentheses and clustered at the school level.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: Source: TIMSS © IEA 1995-2019. All regressions are population-weighted and include controls for students' gender, the index for books, the index for desks, the index for parental education, and year and grade fixed effects.

TABLE A7: ADDITIONAL DESCRIPTIVE STATISTICS FOR DEVELOPING AND DEVELOPED COUNTRIES

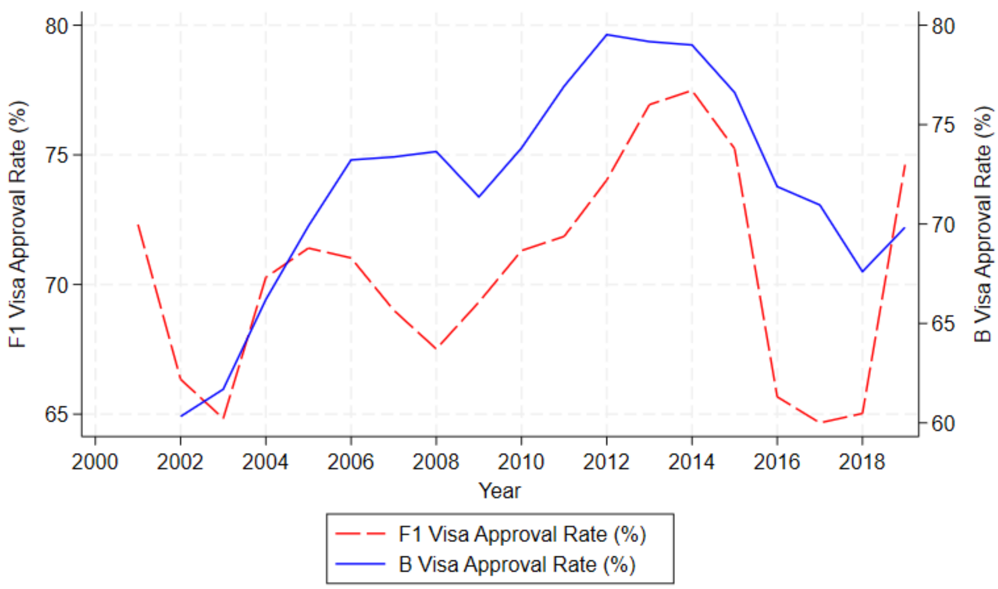
	Developing			Developed		
	mean	sd	n	mean	sd	n
GDP per capita, PPP (constant 2011 $)	11309.09	5299.57	15	37725.57	13975.96	24
Government funding per primary student (% of GDP per capita)	17.76	8.84	10	21.13	5.54	21
Government funding per primary student, constant PPP$	2508.09	1588.08	10	8781.97	2453.88	21
Government funding per pre-primary student (% of GDP per capita)	15.11	10.45	11	14.96	4.85	21
Government funding per pre-primary student, constant PPP$	2144.08	1654.32	11	6119.58	2464.23	21
Total number of computers per school	14.37	12.11	15	29.29	23.37	25
Shortage in instructional material	2.27	0.56	15	1.69	0.33	25
Shortage in supplies	2.17	0.47	15	1.56	0.29	25
Shortage in school buildings	2.37	0.45	15	2.06	0.28	25
Shortage in instructional space	2.16	0.45	15	2.04	0.27	25
Shortage in computer software	2.67	0.35	15	2.17	0.30	25
Shortage in library resources	2.52	0.38	15	2.11	0.26	25
Shortage in calculators	2.13	0.37	15	1.68	0.34	25
Repetition rate in primary education (all grades,%)	2.24	3.32	14	0.93	1.05	21
Number of students in the class	24.80	2.93	15	24.66	4.37	25
Years been teaching	19.40	3.85	15	16.38	3.56	25
Level of formal education completed (Teacher)	3.65	0.68	15	3.86	0.58	25
Attended preschool	0.79	0.15	11	0.96	0.05	20
Length of attending preschool (year)	1.78	0.67	11	2.43	0.47	20

Note: Source: TIMSS © IEA 1995-2019; World Bank EdStats. Each shortage in certain resource has value 1 if there is no shortage and 4 if there is extreme shortage. Level of formal education completed is a categorical variable where 1 means 'didn't complete the secondary school', 2 means 'completed secondary school', 3 means 'post-secondary but not bachelor', 4 means 'bachelor degree', and 5 means 'above bachelor'.

Appendix B

Appendix for Chapter 2

FIGURE A1: APPROVAL RATES OF STUDENT VISAS AND B-VISAS



Source: IPEDS.

TABLE A1: 2SLS RESULTS WITHIN MINORITY GROUP

	(1)	(2)	(3)
Dependent Var:	Hispanic	Black	American Indian & Alaska Native
	0.260***	-0.107	-0.041***
	(0.092)	(0.067)	(0.013)
Observations	29218	29218	29218
Mean $\bar{Y}$	94	98	6
F statistics	35.55	35.55	35.55
Number of Schools	1,539	1,539	1,539
Regional Controls	Yes	Yes	Yes
Institutional Controls	Yes	Yes	Yes

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: All observations are at the institution-year level. Minority students include Black, Hispanic, and American Indian or Alaska Native, as defined by the National Science Foundation (NSF). All specifications include year and institution fixed effects and the total number of FTFT enrollments in the previous year with the institution and regional time-varying controls. All observations are weighted by the 12-month unduplicated undergraduate enrollment in the base year (2000), and standard errors are clustered at the state level. Samples for 2000 are dropped because there is no lagged total FTFT enrollment for that year, nor is there information on F visa approval rates in 2000. The first-stage Kleibergen-Paap F-statistics are reported.

TABLE A2: COMPARISON OF INSTITUTIONS WITH LOW VS HIGH GOVERNMENT FUNDING PER STUDENT

	Low Government Funding	High Government Funding	Test
Observations	23,940 (76.7%)	7,260 (23.3%)	
Number of Universities	1197 (76.7%)	363 (23.3%)	
Public Institutions	0.18 (0.38)	0.88 (0.33)	<0.001
Research Institutions	0.08 (0.27)	0.45 (0.50)	<0.001
Selective Institutions	0.10 (0.30)	0.17 (0.37)	<0.001
12-month unduplicated Headcount	3,993.86 (5,727.42)	12,103.22 (10,489.91)	<0.001
FTFT Enrollment			
International	16.85 (43.03)	63.08 (134.91)	<0.001
Domestic	597.02 (733.50)	1,794.65 (1,646.96)	<0.001
Minority	125.93 (217.49)	486.35 (527.85)	<0.001
Hispanic	62.15 (164.97)	257.54 (448.98)	<0.001
Black	59.97 (114.99)	216.50 (283.28)	<0.001
American Indian & Alaska Natives	3.81 (14.34)	12.31 (27.68)	<0.001
White	404.70 (536.97)	1,034.44 (1,228.34)	<0.001
Asian	26.07 (75.50)	175.16 (332.38)	<0.001
Minority Transfer ins	51.79 (109.17)	201.85 (283.35)	<0.001
Regional Characteristics (State Level)			
City& Suburbs	0.46 (0.50)	0.56 (0.50)	<0.001
% black	12.83 (8.21)	13.69 (9.56)	<0.001
% Hispanic	11.12 (10.01)	13.74 (11.94)	<0.001
Median Income	51,436.79 (9,815.51)	51,313.54 (10,925.03)	0.371
Revenues per student			
Tuition Revenues	13,300.65 (6,926.88)	8,358.65 (6,405.35)	<0.001
Government Fundings	2,348.25 (2,881.52)	16,733.26 (13,175.23)	<0.001
Financial Aid per student			
State/Local grant aid	3,537.58 (2,053.51)	3,213.60 (1,970.67)	<0.001
Institutional grant aid	10,761.35 (7,890.16)	6,503.05 (7,804.06)	<0.001
Expenditures per student			
Instruction	9,032.27 (5,825.72)	12,742.70 (13,396.80)	<0.001
Research	815.48 (1,510.34)	6,674.72 (13,107.50)	<0.001
Public Service	590.74 (852.27)	1,688.73 (2,355.12)	<0.001
Academic Support	2,305.99 (2,092.82)	3,638.68 (5,709.31)	<0.001
Student Service	3,579.62 (2,424.97)	2,297.94 (2,238.13)	<0.001
Institutional Support	4,860.73 (4,042.37)	4,162.88 (4,553.90)	<0.001

Notes: Means of the samples for the years 2000-2019 are described. Standard deviations are shown in parentheses. Low government funding schools are classified as those with less than $9,000 per student from government sources (including state appropriations, local appropriations, and government grants and contracts), while high government funding schools are those with $9,000 or more per pupil from government sources. Minority students include Black, Hispanic, and American Indian or Alaska Native as defined by the National Science Foundation (NSF). Research schools are those with Carnegie classifications of “research universities” and “doctoral/research universities.” Selective institutions are those ranked as ”Most Competitive,” ”Highly Competitive Plus,” or ”Highly Competitive” in the 2009 Barron’s Profile of American Colleges.

TABLE A3: OTHER HETEROGENEITY CHECKS

	Sector		**Selectiveness**		**Highest Degree**		**Median Income**		**Region**	
	Public	Private	Selective	Non-Selective	Bachelor	Master or Higher	Above-mean	Below-mean	City	Rural
Dep. Var:	-0.045	0.422***	0.396	0.289	2.056	0.106	-0.053	0.236	0.121	0.628
Minority Enrollment	(0.114)	(0.111)	(0.288)	(0.190)	(2.490)	(0.114)	(0.190)	(0.163)	(0.136)	(0.416)
Observations	9872	19345	3477	25741	6057	22911	12386	16832	14206	15012
Mean $\bar{Y}$	415	87	217	195	77	231	200	196	269	131
F-Statistics	19.12	22.27	3.84	22.77	1.04	34.83	13.50	27.16	30.18	3.88
Number of Institutions	521	1,019	183	1,356	449	1,321	652	887	748	791
Regional Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Institutional Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: All observations are at the institution-year level. Minority students include Black, Hispanic, and American Indian or Alaska Native, as defined by the National Science Foundation (NSF). All specifications include year and institution fixed effects and the total number of FTFT enrollments in the previous year with the institution and regional time-varying controls. All observations are weighted by the 12-month unduplicated undergraduate enrollment in the base year (2000), and standard errors are clustered at the state level. Samples for 2000 are dropped because there is no lagged total FTFT enrollment for that year, nor is there information on F visa approval rates in 2000. The first-stage Kleibergen-Paap F-statistics are reported. The regional category 'city' includes suburbs.

TABLE A4: DOMESTIC MINORITY ENROLLMENT IN LOW VS. HIGH GOVERNMENT FUNDING INSTITUTIONS (UNBALANCED SAMPLE)

	(1) Low Government Funding	(2) High Government Funding
Panel A: Dep. var: Minority FTFT Enrollment		
OLS	0.321***	-0.025
	(0.099)	(0.066)
1st stage	0.660**	-0.124
	(0.258)	(0.223)
2SLS	0.660**	-0.124
	(0.258)	(0.223)
Observations	22551	6580
Mean $\bar{Y}$	116	484
F statistics	15.56	16.43
Number of Schools	1,194	348
Regional Controls	Yes	Yes
Institutional Controls	Yes	Yes
Panel B: Dep. var.: Other racial groups FTFT Enrollment		
2SLS		
Total Domestic	-0.839***	-0.307
	(0.276)	(0.239)
White	-1.502***	-1.268***
	(0.304)	(0.264)
Hispanic	0.465***	0.057
	(0.170)	(0.189)
Black	0.225*	-0.140
	(0.124)	(0.096)
American Indian and Alaska Native	-0.030*	-0.042**
	(0.015)	(0.018)
Asian	0.228**	0.539***
	(0.097)	(0.148)

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: All observations are at the institution-year level. Institutions with 10 or more observations are included in the sample. Low government funding schools are classified as those with less than $9,000 per student from government sources (including state appropriations, local appropriations, and government grants and contracts), while high government funding schools are those with $9,000 or more per pupil from government sources. Minority students include Black, Hispanic, and American Indian or Alaska Native, as defined by the National Science Foundation (NSF). All specifications include year and institution fixed effects and the total number of FTFT enrollments in the previous year with the institution and regional time-varying controls. All observations are weighted by the 12-month unduplicated undergraduate enrollment in the base year (2000), and standard errors are clustered at the state level. Samples for the year 2000 are dropped because there is no lagged total FTFT enrollment for that year, nor is there information on F visa approval rates in 2000. The first-stage Kleibergen-Paap F-statistics are reported.

TABLE A5: EFFECTS ON TUITION REVENUE, INSTITUTIONAL AID, AND EXPENDITURES

	(1) Total	(2) Low Government Funding	(3) High Government Funding
Dep. var:			
Tuition Revenue per student	9.171***	10.642***	8.041***
	(1.625)	(3.736)	(2.351)
Institutional Aid per student	8.877**	10.124***	13.652**
	(3.523)	(2.916)	(5.926)
Expenditures			
Instruction	8.609***	3.912*	9.667***
	(2.287)	(2.213)	(3.518)
Research	3.160*	6.352	1.430
	(1.760)	(4.000)	(2.238)
Public Service	-0.162	0.283	-1.154
	(0.887)	(1.672)	(1.168)
Academic Support	4.092***	-0.435	5.823***
	(1.224)	(0.480)	(1.441)
Student Service	-0.273	-0.089	-0.209
	(0.510)	(0.852)	(0.611)
Institutional Support	2.736**	2.196	2.703
	(1.156)	(1.765)	(1.740)

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: All observations are at the institution-year level. Low government funding schools are classified as those with less than $9,000 per student from government sources (including state appropriations, local appropriations, and government grants and contracts), while high government funding schools are those with $9,000 or more per pupil from government sources. Minority students include Black, Hispanic, and American Indian or Alaska Native, as defined by the National Science Foundation (NSF). All specifications include year and institution fixed effects and the total number of FTFT enrollments in the previous year with the institution and regional time-varying controls. All observations are weighted by the 12-month unduplicated undergraduate enrollment in the base year (2000), and standard errors are clustered at the state level. Samples for the year 2000 are dropped because there is no lagged total FTFT enrollment for that year, nor is there information on F visa approval rates in 2000. The first-stage Kleibergen-Paap F-statistics are reported. All financial variables are deflated by the Higher Education Price Index(HEPI) and presented in 2013 dollars.

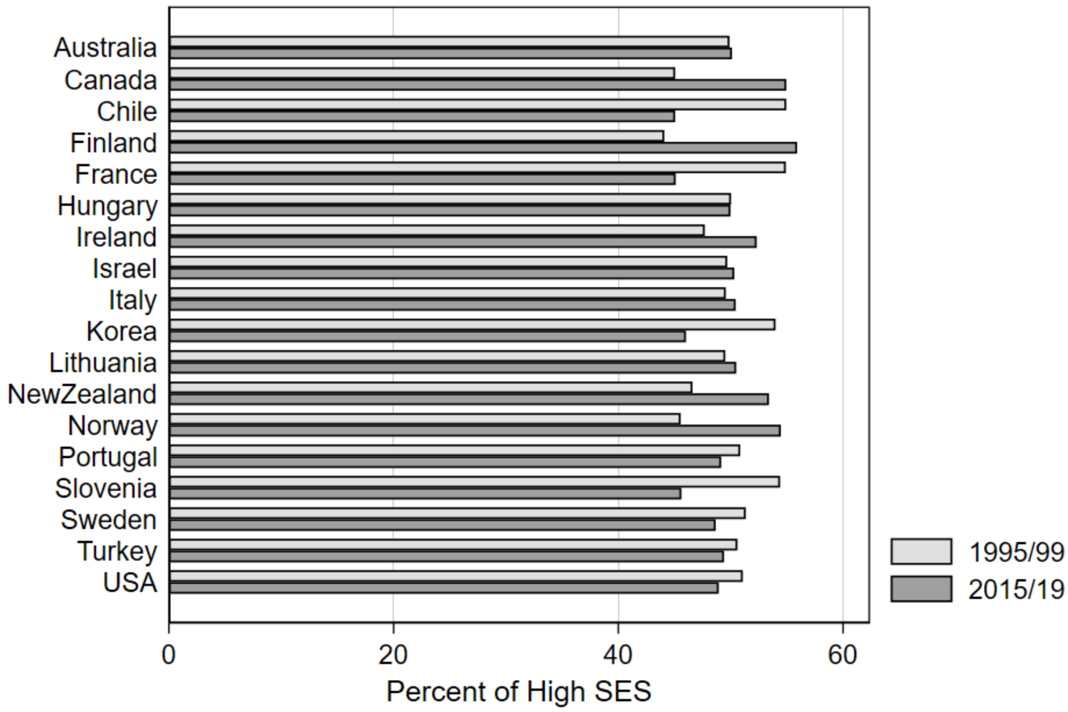
Appendix C

Appendix for Chapter 3

Figure A1: Distribution of SES Scores by country

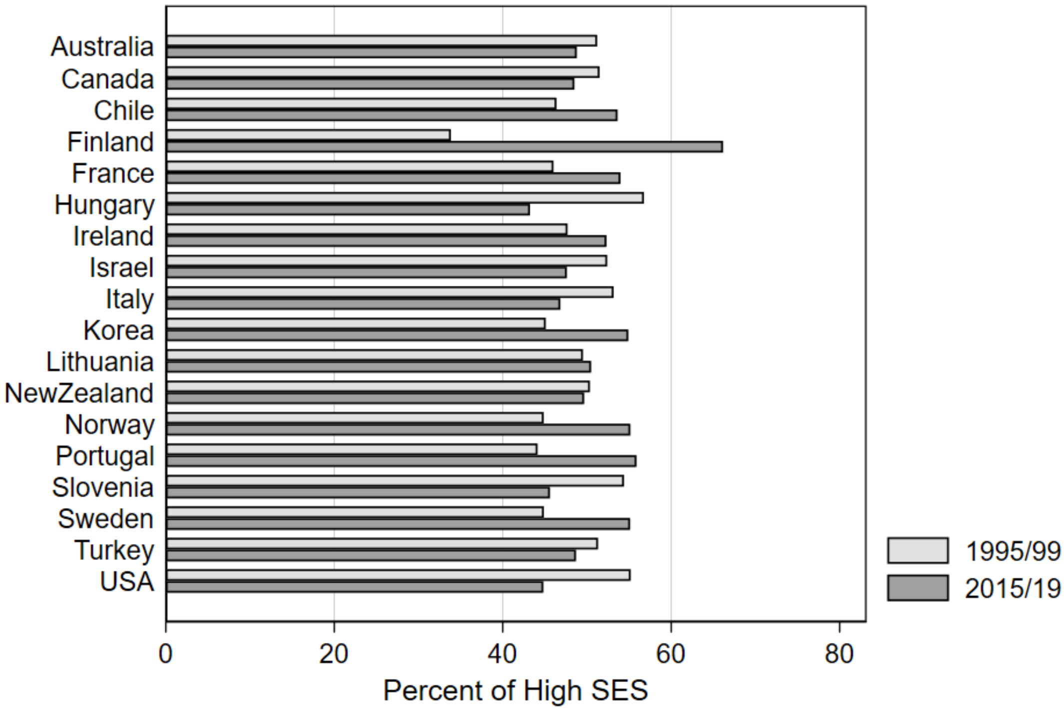
Source: TIMSS © IEA 1995-2019.

FIGURE A2: PERCENTAGE OF HIGH SES STUDENTS BY COUNTRY (HIGH SES: SES SCORE ≥ 50TH PERCENTILE SES SCORE)



Source: TIMSS © IEA 1995-2019.

FIGURE A3: PERCENTAGE OF HIGH SES STUDENTS BY COUNTRY (HIGH SES: SES SCORE ≥ 8



Source: TIMSS © IEA 1995-2019.

TABLE A1: DESCRIPTIVE STATISTICS FOR VARIABLES USED TO CREATE SES INDEX IN 1995/99

	(1) Avg SES index	(2) Books	(3) Parental Education	(4) Desk	(5) Calculator	(6) Dictionary
Australia	8.50	2.99	2.65	0.97	0.98	0.91
Canada	8.39	2.68	2.87	0.89	0.98	0.97
Chile	6.54	1.58	2.27	0.77	0.94	0.97
Finland	6.67	2.43	1.38	0.97	1.00	0.89
France	7.13	2.34	1.85	0.96	0.99	0.99
Hungary	8.76	2.85	3.02	0.95	0.99	0.95
Ireland	7.77	2.39	2.56	0.86	0.97	0.99
Israel	8.32	2.46	2.92	0.98	0.98	0.98
Italy	7.71	2.23	2.59	0.94	0.97	0.99
Korea	8.06	2.38	2.80	0.96	0.94	0.99
Lithuania	8.05	2.44	2.87	0.94	0.90	0.88
NewZealand	8.16	2.86	2.43	0.90	0.98	0.98
Norway	8.27	3.04	2.32	0.98	0.98	0.96
Portugal	7.00	1.99	2.20	0.83	0.99	0.97
Slovenia	8.06	2.38	2.83	0.94	0.98	0.93
Sweden	7.97	2.96	2.09	0.99	0.99	0.93
Turkey	6.06	1.42	2.25	0.68	0.83	0.88
USA	8.26	2.45	2.97	0.89	0.98	0.97
Total	7.95	2.50	2.61	0.91	0.97	0.96
Observations	157738	157738	157738	157738	157738	157738

Notes: Source: TIMSS © IEA 1995-2019. Bookscore is 0 if the number of books is 0-15, 1 for 11-25 books, 2 for 26-100 books, 3 for 101-200 books, and 4 for more than 200 books. Parental Education is 0 if parental education is less than lower secondary, 1 for lower secondary, 2 for upper secondary, 3 for post secondary, and 4 for university or higher. Desk is 1 if a student has a desk, 0 otherwise. Calculator and Dictionary are defined in the same way. Therefore, the maximum of the SES index is 11.

TABLE A2: DESCRIPTIVE STATISTICS FOR VARIABLES USED TO CREATE SES INDEX IN 2015/19

	(1) Avg SES index	(2) Books	(3) Parental Education	(4) Desk	(5) Internet	(6) Mobile
Australia	8.28	2.37	3.17	0.89	0.96	0.89
Canada	8.14	2.14	3.35	0.92	0.99	0.74
Chile	6.84	1.39	2.80	0.83	0.89	0.93
Finland	8.31	2.24	3.14	0.94	0.99	1.00
France	7.48	1.84	2.82	0.92	0.99	0.92
Hungary	8.13	2.31	2.89	0.95	0.99	0.98
Ireland	8.02	2.11	3.07	0.88	0.99	0.98
Israel	8.13	2.06	3.24	0.94	0.96	0.93
Italy	7.36	2.02	2.52	0.88	0.96	0.98
Korea	9.08	2.86	3.29	0.97	0.98	0.97
Lithuania	8.15	1.92	3.28	0.98	0.99	0.99
NewZealand	8.05	2.36	2.97	0.85	0.97	0.91
Norway	8.80	2.34	3.54	0.93	1.00	0.99
Portugal	7.03	1.71	2.39	0.97	0.99	0.98
Slovenia	7.82	1.96	2.94	0.95	0.99	0.97
Sweden	8.44	2.18	3.30	0.97	0.99	0.99
Turkey	5.27	1.65	1.62	0.76	0.67	0.57
USA	7.65	1.89	3.07	0.81	0.96	0.92
Total	7.87	2.09	3.00	0.90	0.96	0.91
Observations	225650	225650	225650	225650	225650	225650

Notes: Source: TIMSS © IEA 1995-2019. Bookscore is 0 if the number of books is 0-15, 1 for 11-25 books, 2 for 26-100 books, 3 for 101-200 books, and 4 for more than 200 books. Parental Education is 0 if parental education is less than lower secondary, 1 for lower secondary, 2 for upper secondary, 3 for post secondary, and 4 for university or higher. Desk is 1 if a student has a desk, 0 otherwise. Internet and Mobile are defined in the same way. Therefore, the maximum of the SES index is 11.

TABLE A3: GENDER AND SOCIOECONOMIC STATUS GAPS

	Gender gaps				SES gaps			
	Math		Science		Math		Science	
	(1) Gap	(2) Δ Gap	(3) Gap	(4) Δ Gap	(5) Gap	(6) Δ Gap	(7) Gap	(8) Δ Gap
Australia	-1.53	-7.47*	-16.12*	9.61*	-47.00*	-14.75*	-54.50*	-11.79*
	(3.19)	(3.33)	(3.23)	(3.45)	(2.29)	(3.00)	(2.31)	(2.61)
Canada	-4.12*	-1.76	-18.07*	12.70*	-32.85*	-10.22*	-46.19*	-0.71
	(1.54)	(1.63)	(1.36)	(1.62)	(1.65)	(2.29)	(2.26)	(2.70)
Chile	-15.60*	2.25	-29.37*	15.84*	-47.13*	-17.87*	-49.02*	-13.58*
	(5.62)	(5.96)	(6.64)	(6.62)	(5.96)	(6.18)	(4.75)	(5.02)
Finland	-3.17	4.85	-11.92*	27.66*	-26.36*	-22.27*	-33.17*	-23.11*
	(3.39)	(3.86)	(5.19)	(5.36)	(3.58)	(3.25)	(3.87)	(3.70)
France	-12.46*	-0.17	-30.58*	22.23*	-19.66*	-31.96*	-25.32*	-36.33*
	(1.84)	(2.40)	(1.98)	(2.55)	(2.05)	(2.70)	(2.14)	(2.20)
Hungary	-10.89*	-8.55*	-29.84*	7.13	-64.69*	-18.98*	-53.74*	-20.06*
	(3.25)	(3.45)	(3.58)	(3.95)	(3.93)	(4.30)	(5.10)	(4.60)
Ireland	-18.35*	13.44*	-26.72*	26.03*	-46.80*	-5.39	-52.47*	-6.41
	(5.05)	(5.48)	(4.87)	(5.39)	(3.61)	(4.12)	(4.05)	(4.34)
Israel	-27.15*	18.50*	-29.53*	28.74*	-42.97*	-39.27*	-50.16*	-30.90*
	(3.42)	(3.95)	(4.16)	(4.12)	(4.60)	(5.05)	(4.48)	(5.39)
Italy	-12.33*	0.20	-17.81*	7.49*	-43.62*	-5.43	-47.17*	-7.14*
	(2.97)	(2.99)	(3.33)	(3.57)	(2.64)	(3.47)	(2.85)	(3.02)
Korea	-12.63*	5.08	-25.85*	14.90*	-58.88*	5.11*	-52.95*	9.28*
	(2.53)	(2.59)	(2.57)	(2.83)	(2.26)	(2.51)	(1.90)	(2.36)
Lithuania	-0.25	-3.05	-21.69*	24.00*	-54.64*	-17.74*	-53.13*	-15.69*
	(2.55)	(2.71)	(2.63)	(2.49)	(2.81)	(2.86)	(3.19)	(3.41)
New Zealand	-2.23	-5.66	-19.31*	12.52*	-49.00*	-32.85*	-58.64*	-30.62*
	(3.78)	(4.13)	(3.68)	(3.92)	(2.73)	(3.14)	(2.79)	(3.75)
Norway	-3.80	4.26	-16.24*	14.75*	-31.99*	-23.01*	-34.30*	-30.45*
	(2.05)	(2.44)	(2.30)	(3.03)	(1.83)	(2.19)	(1.92)	(2.37)
Portugal	-16.95*	1.43	-36.65*	27.32*	-32.15*	-18.75*	-35.84*	-11.93*
	(1.94)	(2.28)	(1.85)	(1.98)	(2.48)	(2.80)	(2.71)	(2.93)
Slovenia	-6.83*	1.02	-22.93*	22.38*	-45.96*	-0.56	-46.12*	-11.81*
	(1.84)	(2.85)	(1.69)	(3.25)	(1.79)	(3.66)	(2.31)	(3.96)
Sweden	-2.29	-2.02	-17.53*	21.34*	-29.70*	-28.95*	-35.20*	-41.80*
	(1.78)	(2.09)	(1.87)	(2.12)	(1.93)	(2.29)	(2.11)	(2.34)
Turkey	-6.21*	15.22*	-6.74*	19.79*	-38.16*	-39.42*	-32.01*	-35.40*
	(2.44)	(2.52)	(3.02)	(3.48)	(4.01)	(4.28)	(3.44)	(3.75)
USA	-10.44*	7.17*	-21.77*	17.88*	-57.02*	-14.28*	-67.66*	-0.99
	(1.99)	(1.95)	(2.00)	(2.03)	(2.63)	(2.96)	(3.00)	(3.34)

Standard errors in parentheses
* $p < 0.05$

Notes: Source: TIMSS © IEA 1995-2019. Standard errors are in parentheses and * means the significance at the 5 percent level or better. Students with high socioeconomic status are defined as those with SES index bigger or equal to top 50% student's SES index and low SES is defined as the rest. Columns (1)-(4) show gender gaps and Columns (5)-(8) show SES gaps for both subjects. The original gender and SES gaps are shown in Columns with odd numbers, and the changes in gaps over time are shown in Columns with even numbers. In the regression, jackknife repeated replication (JRR) standard errors are used, and covariates include indicator variables for grade, age and age^2. Sample sizes for each country are reported in Table 1.

TABLE A4: SES GAP FOR MATH AND SCIENCE USING SES INDEX≥8

	Math		Science	
	Gap	Δ Gap	Gap	Δ Gap
	(1)	(2)	(3)	(4)
Australia	-47.70*	-22.21*	-55.69*	-20.09*
	(2.37)	(3.07)	(2.44)	(2.86)
Canada	-34.26*	-8.77*	-46.08*	-0.77
	(1.67)	(2.19)	(2.42)	(2.90)
Chile	-60.86*	-4.60	-59.70*	-2.35
	(6.57)	(6.63)	(5.60)	(5.73)
Finland	-28.04*	-20.59*	-38.00*	-18.28*
	(4.06)	(3.57)	(3.90)	(3.74)
France	-22.27*	-29.40*	-28.44*	-33.29*
	(2.15)	(2.89)	(2.11)	(2.37)
Hungary	-66.12*	-17.56*	-57.62*	-16.20*
	(4.14)	(4.53)	(6.52)	(5.99)
Ireland	-46.80*	-5.39	-52.47*	-6.41
	(3.61)	(4.12)	(4.05)	(4.34)
Israel	-54.01*	-28.24*	-62.67*	-18.40*
	(4.74)	(5.20)	(4.90)	(5.77)
Italy	-44.09*	-1.37	-46.95*	-3.26
	(2.56)	(3.06)	(2.92)	(2.83)
Korea	-58.85*	-2.48	-52.93*	4.50
	(2.26)	(2.50)	(1.89)	(3.01)
Lithuania	-54.64*	-17.74*	-53.13*	-15.69*
	(2.81)	(2.86)	(3.19)	(3.41)
New Zealand	-48.72*	-33.13*	-57.54*	-31.71*
	(2.77)	(3.18)	(2.92)	(3.89)
Norway	-27.09*	-30.78*	-28.63*	-41.85*
	(1.77)	(1.94)	(2.03)	(2.54)
Portugal	-35.87*	-15.68*	-43.27*	-5.27
	(2.86)	(3.41)	(2.70)	(2.94)
Slovenia	-45.96*	-0.56	-46.12*	-11.81*
	(1.79)	(3.66)	(2.31)	(3.96)
Sweden	-29.67*	-30.92*	-35.16*	-47.24*
	(1.93)	(2.24)	(2.11)	(2.34)
Turkey	-46.01*	-50.94*	-36.73*	-47.35*
	(5.21)	(5.24)	(4.18)	(4.48)
USA	-57.46*	-13.84*	-71.21*	2.55
	(2.55)	(2.92)	(3.10)	(3.52)

Standard errors in parentheses
* $p < 0.05$

Notes: Source: TIMSS © IEA 1995-2019. Standard errors are in parentheses and * means the significance at the 5 percent level or better. Students with high socioeconomic status are defined as those with SES index bigger than 8 and low SES is defined as the rest. The original gender and SES gaps are shown in Columns with odd numbers, and the changes in gaps over time are shown in Columns with even numbers. In the regression, jackknife repeated replication (JRR) standard errors are used, and covariates include indicator variables for grade, age and age^2. Sample sizes for each country are reported in Table 1.

Bibliography

Agan, A. Y., Doleac, J. L. and Harvey, A. (2021), Misdemeanor prosecution, Technical report, National Bureau of Economic Research.

Altwicker-Hámori, S. and Köllő, J. (2012), 'Whose children gain from starting school later?–evidence from hungary', *Educational Research and Evaluation* **18**(5), 459–488.

Ando, S., Usami, S., Matsubayashi, T., Ueda, M., Koike, S., Yamasaki, S., Fujikawa, S., Sasaki, T., Hiraiwa-Hasegawa, M., Patton, G. et al. (2019), 'Age relative to school class peers and emotional well-being in 10-year-olds', *Plos one* **14**(3), e0214359.

Angrist, J. D. and Imbens, G. W. (1995), 'Two-stage least squares estimation of average causal effects in models with variable treatment intensity', *Journal of the American statistical Association* **90**(430), 431–442.

Angrist, J. D., Imbens, G. W. and Rubin, D. B. (1996), 'Identification of causal effects using instrumental variables', *Journal of the American statistical Association* **91**(434), 444–455.

Angrist, J. D. and Krueger, A. B. (1992), 'The effect of age at school entry on educational attainment: an application of instrumental variables with moments from two samples', *Journal of the American statistical Association* **87**(418), 328–336.

Bai, Y., Straus, S. and Broer, M. (2021), 'Us national and state trends in educational inequality due to socioeconomic status: Evidence from the 2003-17 naep. air-naep working paper 2021-01.', *American Institutes for Research* .

Barua, R. and Lang, K. (2016), 'School entry, educational attainment, and quarter of birth: A cautionary tale of a local average treatment effect', *Journal of Human Capital* **10**(3), 347–376.

Baum, S., Ma, J. and Payea, K. (2013), 'Education pays, 2013: The benefits of higher education for individuals and society. trends in higher education series.', *College Board* .

Bedard, K. and Cho, I. (2010), 'Early gender test score gaps across oecd countries', *Economics of Education Review* **29**(3), 348–363.

Bedard, K. and Dhuey, E. (2006), 'The persistence of early childhood maturity: International evidence of long-run age effects', *The Quarterly Journal of Economics* **121**(4), 1437–1472.

Bedard, K. and Dhuey, E. (2012), 'School-entry policies and skill accumulation across directly and indirectly affected individuals', *Journal of Human Resources* **47**(3), 643–683.

Bhuller, M., Dahl, G. B., Løken, K. V. and Mogstad, M. (2020), 'Incarceration, recidivism, and employment', *Journal of Political Economy* **128**(4), 1269–1324.

Black, S. E., Devereux, P. J. and Salvanes, K. G. (2011), 'Too young to leave the nest? the effects of school starting age', *The review of economics and statistics* **93**(2), 455–467.

Borjas, G. J. (2004), 'Do foreign students crowd out native students from graduate programs?'.

Borusyak, K., Hull, P. and Jaravel, X. (2022), 'Quasi-experimental shift-share research designs', *The Review of Economic Studies* **89**(1), 181–213.

Bound, J., Braga, B., Khanna, G. and Turner, S. (2020), 'A passage to america: University funding and international students', *American Economic Journal: Economic Policy* **12**(1), 97–126.

Bound, J. and Jaeger, D. A. (2000), Do compulsory school attendance laws alone explain the association between quarter of birth and earnings?, *in* 'Research in labor economics', Vol. 19, Emerald Group Publishing Limited, pp. 83–108.

Bound, J., Jaeger, D. A. and Baker, R. M. (1995), 'Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak', *Journal of the American statistical association* **90**(430), 443–450.

Broer, M., Bai, Y. and Fonseca, F. (2019), *Socioeconomic inequality and educational outcomes: Evidence from twenty years of TIMSS*, Springer Nature.

Buckles, K. S. and Hungerman, D. M. (2013), 'Season of birth and later outcomes: Old questions, new answers', *Review of Economics and Statistics* **95**(3), 711–724.

Cascio, E. U. and Schanzenbach, D. W. (2016), 'First in the class? age and the education production function', *Education Finance and Policy* **11**(3), 225–250.

Chen, J. and Park, A. (2021), 'School entry age and educational attainment in developing countries: Evidence from china's compulsory education law', *Journal of Comparative Economics* **49**(3), 715–732.

Chen, M. (2021), 'The impact of international students on us colleges: Higher education as a service export', *Available at SSRN 3859798* .

Chen, M., Howell, J. and Smith, J. (2023), 'Best and brightest? the impact of student visa restrictiveness on who attends college in the us', *Labour Economics* **84**, 102385.

Chmielewski, A. K. (2019), 'The global increase in the socioeconomic achievement gap, 1964 to 2015', *American sociological review* **84**(3), 517–544.

Chmielewski, A. K. and Reardon, S. F. (2016), 'Patterns of cross-national variation in the association between income and academic achievement', *Aera Open* **2**(3), 2332858416649593.

Corak, M. (2013), 'Income inequality, equality of opportunity, and intergenerational mobility', *Journal of Economic Perspectives* **27**(3), 79–102.

Cordero, J. A. (1985), 'Relación de la variable edad con las deficiencias de escolarización en la egb', *Infancia y Aprendizaje* **8**(29), 95–104.

Crawford, C., Dearden, L. and Greaves, E. (2014), 'The drivers of month-of-birth differences in children's cognitive and non-cognitive skills', *Journal of the Royal Statistical Society. Series A,(Statistics in Society)* **177**(4), 829.

Crawford, C., Dearden, L. and Meghir, C. (2010), When you are born matters: the impact of date of birth on educational outcomes in england, Technical report, IFS working papers.

Datar, A. (2006), 'Does delaying kindergarten entrance give children a head start?', *Economics of Education review* **25**(1), 43–62.

De Chaisemartin, C. (2017), 'Tolerating defiance? local average treatment effects without monotonicity', *Quantitative Economics* **8**(2), 367–396.

Dee, T. S. and Sievertsen, H. H. (2018), 'The gift of time? school starting age and mental health', *Health economics* **27**(5), 781–802.

Deming, D. and Dynarski, S. (2008), 'The lengthening of childhood', *Journal of economic perspectives* **22**(3), 71–92.

Dhuey, E., Figlio, D., Karbownik, K. and Roth, J. (2019), 'School starting age and cognitive development', *Journal of Policy Analysis and Management* **38**(3), 538–578.

Dobbie, W., Goldin, J. and Yang, C. S. (2018), 'The effects of pre-trial detention on conviction, future crime, and employment: Evidence from randomly assigned judges', *American Economic Review* **108**(2), 201–240.

Dobkin, C. and Ferreira, F. (2010), 'Do school entry laws affect educational attainment and labor market outcomes?', *Economics of education review* **29**(1), 40–54.

Eide, E. R. and Showalter, M. H. (2001), 'The effect of grade retention on educational and labor market outcomes', *Economics of Education review* **20**(6), 563–576.

Elder, T. E. and Lubotsky, D. H. (2009), 'Kindergarten entrance age and children's achievement impacts of state policies, family background, and peers', *Journal of human Resources* **44**(3), 641–683.

Fletcher, J. and Kim, T. (2016), 'The effects of changes in kindergarten entry age policies on educational achievement', *Economics of education review* **50**, 45–62.

for Economic Co-operation, O. and Development (2018), *Equity in education: Breaking down barriers to social mobility*, OECD Publishing.

Fredriksson, P. and Ockert, B. (2005), 'Is early learning really more productive? the effect of school starting age on school and labor market performance'.

Fredriksson, P. and Öckert, B. (2014), 'Life-cycle effects of age at school start', *The Economic Journal* **124**(579), 977–1004.

Groen, J. A. and White, M. J. (2004), 'In-state versus out-of-state students: the divergence of interest between public universities and state governments', *Journal of Public Economics* **88**(9-10), 1793–1814.

Hansen, B. E. (2022), 'Jackknife standard errors for clustered regression', *University of Wisconsin* .

Hansen, M., Levesque, E. M., Quintero, D. and Valant, J. (2018), 'Have we made progress on achievement gaps? looking at evidence from the new naep results', *Washington, DC: Brookings Institution. April* **17**.

Hanushek, E. A., Light, J. D., Peterson, P. E., Talpey, L. M. and Woessmann, L. (2022), 'Long-run trends in the us ses—achievement gap', *Education Finance and Policy* **17**(4), 608–640.

Hanushek, E. A., Piopiunik, M. and Wiederhold, S. (2019), 'The value of smarter teachers international evidence on teacher cognitive skills and student performance', *Journal of Human Resources* **54**(4), 857–899.

Hanushek, E. A. and Woessmann, L. (2011), 'The economics of international differences in educational achievement', *Handbook of the Economics of Education* **3**, 89–200.

Hashim, S. A., Kane, T. J., Kelley-Kemple, T., Laski, M. E. and Staiger, D. O. (2020), Have income-based achievement gaps widened or narrowed?, Technical report, National Bureau of Economic Research.

Hauser, R. (2005), 'Kl alexander, dr entwisle, sl dauber, on the success of failure, a re-assessment of the effects of retention in the primary school grades, cambridge university press, new york (2003), 314 pp.'.

Hoxby, C. M. (1998), 'Do immigrants crowd disadvantaged', *Help or hindrance* p. 282.

Imbens, G. W. and Angrist, J. D. (1994), 'Identification and estimation of local average treatment effects', *Econometrica: journal of the Econometric Society* pp. 467–475.

Kawaguchi, D. (2011), 'Actual age at school entry, educational outcomes, and earnings', *Journal of the Japanese and International Economies* **25**(2), 64–80.

Kitagawa, T. (2015), 'A test for instrument validity', *Econometrica* **83**(5), 2043–2063.

Macdonald, K. (2019), 'Pv: Stata module to perform estimation with plausible values'.

Martin, M. O., von Davier, M. and Mullis, I. V. (2020), 'Methods and procedures: Timss 2019 technical report.', *International Association for the Evaluation of Educational Achievement* .

McEwan, P. J. and Shapiro, J. S. (2008), 'The benefits of delayed primary school enrollment discontinuity estimates using exact birth dates', *Journal of human Resources* **43**(1), 1–29.

Meinck, S. and Brese, F. (2019), 'Trends in gender gaps: Using 20 years of evidence from timss', *Large-Scale Assessments in Education* **7**(1), 1–23.

Morales, M. (2020), *School-entry Eligibility Effects in Developing Countries*, Columbia University.

Nunn, N. and Qian, N. (2014), 'Us food aid and civil conflict', *American economic review* **104**(6), 1630–1666.

Peña, P. A. (2017), 'Creating winners and losers: Date of birth, relative age in school, and outcomes in childhood and adulthood', *Economics of Education Review* **56**, 152–176.

Peña, P. A. (2020), 'Relative age and investment in human capital', *Economics of Education Review* **78**, 102039.

Pierson, L. H. and Connell, J. P. (1992), 'Effect of grade retention on self-system processes, school engagement, and academic performance.', *Journal of Educational psychology* **84**(3), 300.

Pokropek, A. (2016), 'Introduction to instrumental variables and their application to large-scale assessment data', *Large-scale Assessments in Education* **4**, 1–20.

Puhani, P. A. and Weber, A. M. (2008), *Does the early bird catch the worm? Instrumental variable estimates of early educational effects of age of school entry in Germany*, Springer.

Reardon, S. F. (2018), The widening academic achievement gap between the rich and the poor, *in* 'Social stratification', Routledge, pp. 536–550.

Rumberger, R. W. (1987), 'High school dropouts: A review of issues and evidence', *Review of educational research* **57**(2), 101–121.

Ryu, H., Helfand, S. M. and Moreira, R. B. (2020), 'Starting early and staying longer: The effects of a brazilian primary schooling reform on student performance', *World Development* **130**, 104924.

Sargan, J. D. (1958), 'The estimation of economic relationships using instrumental variables', *Econometrica: Journal of the econometric society* pp. 393–415.

Schwab, B., Janzen, S., Magnan, N. P. and Thompson, W. M. (2020), 'Constructing a summary index using the standardized inverse-covariance weighted average of indicators', *The Stata Journal* **20**(4), 952–964.

Shakeel, M. D. and Peterson, P. E. (2022), 'A half century of progress in us student achievement: Agency and flynn effects, ethnic and ses differences', *Educational Psychology Review* **34**(3), 1255–1342.

Shen, Y. (2016), 'The impacts of the influx of new foreign undergraduate students on us higher education', *Journal of Economic Literature* .

Shih, K. (2017), 'Do international students crowd-out or cross-subsidize americans in higher education?', *Journal of Public Economics* **156**, 170–184.

Singh, A. (2020), 'Learning more with every year: School year productivity and international learning divergence', *Journal of the European Economic Association* **18**(4), 1770–1813.

Smith, J. (2010), ‘How valuable is the gift of time? the factors that drive the birth date effect in education’, *Education Finance and Policy* **5**(3), 247–277.

Staiger, D. O. and Stock, J. H. (1994), ‘Instrumental variables regression with weak instruments’.

Steinmann, I., Strietholt, R. and Rosén, M. (2023), ‘International reading gaps between boys and girls, 1970–2016’, *Comparative Education Review* **67**(2), 298–330.

Stock, J. H. and Yogo, M. (2002), ‘Testing for weak instruments in linear iv regression’.

Urruticoechea, A., Oliveri, A., Vernazza, E., Giménez-Dasí, M., Martínez-Arias, R. and Martín-Babarro, J. (2021), ‘The relative age effects in educational development: A systematic review’, *International journal of environmental research and public health* **18**(17), 8966.

Wasem, R. E. (2004), Visa issuances: Policy, issues, and legislation, Congressional Research Service, the Library of Congress.

Weinberger, C. (2005), Is the science and engineering workforce drawn from the far upper tail of the math ability distribution, Technical report, Working Paper, UCSB.

Wilson, M. (1990), ‘Flunking grades: Research and policies on retention’.