ifo BEITRÄGE
zur Wirtschaftsforschung

109
2025

Global Challenges in the Economics of Education: Basic Skills, Natural Disasters, and Civic Engagement

Sarah Gust

ifo INSTITUTE
Leibniz Institute for Economic Research
at the University of Munich

ifo BEITRÄGE zur Wirtschaftsforschung

109
2025

Global Challenges in the Economics of Education: Basic Skills, Natural Disasters, and Civic Engagement

Sarah Gust

Herausgeber der Reihe: Clemens Fuest
Schriftleitung: Chang Woon Nam

ifo INSTITUTE
Leibniz Institute for Economic Research
at the University of Munich

Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation
in der Deutschen Nationalbibliografie; detaillierte bibliografische
Daten sind im Internet über
https://dnb.d-nb.de
abrufbar.

ISBN Nr. 978-3-95942-148-5

Druck: Pinsker Druck und Medien GmbH, Mainburg

ifo Institut im Internet:
https://www.ifo.de

Global Challenges in the Economics of Education: Basic Skills, Natural Disasters, and Civic Engagement

Inaugural-Dissertation
Zur Erlangung des Grades
Doctor oeconomiae publicae (Dr. oec. publ.)

eingereicht an der
Ludwig-Maximilians-Universität München
2024

vorgelegt von

Sarah Gust

Referent: Prof. Dr. Ludger Wößmann
Koreferent: Prof. Dr. Helmut Rainer
Promotionsabschlussberatung: 29. Januar 2025

Datum der mündlichen Prüfung:	13.01.2025
Namen der Berichterstatter:innen:	Prof. Dr. Ludger Wößmann
	Prof. Dr. Helmut Rainer
	Prof. Dr. Ingrid Hägele

Preface

Sarah Gust conducted this study while working at the ifo Institute. The study was completed in September 2024 and accepted as a doctoral thesis by the Department of Economics at LMU Munich. It comprises three distinct empirical essays that apply microeconometric methods. The study addresses three critical challenges faced by contemporary societies worldwide through the lenses of economics of education: ensuring basic skills, understanding the effects of natural disasters, and fostering civic engagement. The first chapter provides a general introduction to the topics and methods used throughout the study. Chapter 2 constructs a global database on the lack of basic skills and simulates the economic losses associated with failing to achieve universal basic skills worldwide. Chapter 3 examines the effects of natural disasters on student achievement and explores the underlying mechanisms, revealing persistent negative impacts. Chapter 4 investigates the relationship between civic education in schools and civic engagement in adulthood. The analysis shows that introducing civic education as a subject has a positive impact on civic engagement. However, findings regarding the impact of increased average instructional hours, are less definitive, showing negligible effects on average.

Keywords:	Skills, Student Achievement, Development Goals, Economic Growth, Natural Disasters, Education Economics, Disaster Resilience, Human Capital, Civic Education, Political Behavior and Attitudes, Non-Market Benefits to Education
JEL-No:	I25, O15, O47, Q54, I21, O44, I26, J24, H4

"Nothing in life is to be feared, it is only to be understood. Now is the time to understand more, so that we may fear less."

(Marie Curie)

Acknowledgement

I am deeply grateful to my supervisor, Ludger Wößmann, for giving me the opportunity to conduct research in such an inspiring environment and for his unwavering support throughout this journey. The privilege of doing research, attending conferences, connecting with the international research community, and spending two months abroad is something I do not take for granted - and Ludger made all of this possible. I had been used to providing recaps of previous discussions at the beginning of every meeting, but with Ludger, there was no need – he remembered everything. While I may never fully grasp these remarkable skills, I have certainly learned a great deal from him. Working with Ludger and Rick Hanushek on my first project was a fantastic experience. I especially enjoyed the time Rick visited Munich, when we made real progress during those intensive working sessions. Despite the tight schedules, our discussions were never rushed - there was always room for deep conversation, meticulous attention to detail, and a constant push to strive for excellence. Throughout, I always felt both valued and heard. Aside from being a great co-author and mentor, I also want to thank Rick for hosting me at the Hoover Institution at Stanford University. He and Macke Raymond made me feel very welcome, and it was a real highlight of my PhD. The generous financial support from the CESifo Young Ambassador Program made this research stay possible. I also want to thank my second supervisor, Helmut Rainer. I admire his research, his commitment to continually enhancing ifo as a welcoming and supportive environment, and his impressive skills as a cyclist. Looking back, I was fortunate to have encouraging mentors and supervisors throughout my education and career. Special thanks to René Böheim and Jesús Crespo for supporting me in my PhD application and guiding me on this path.

I am truly grateful to all my co-authors and colleagues. Being part of the ifo Center for the Economics of Education has been special, and sharing this journey with my fellow PhD students has been incredibly rewarding. Katharina Werner, my junior supervisor, but really all postdocs in our group have been an invaluable blend of mentors and companions. I always knew I could rely on them for both work-related advice and a listening ear. Simon ter Meulen and Florian Schoner are not only great colleagues but also the best cycling buddies anyone could wish for. I want to thank Larissa Zierow for being a great co-author, teaching me how to navigate conferences, and sharing countless essential survival strategies. A special thanks to Ulrike Baldi-Cohrs and Franziska Binder, who are the backbone of this team, and to Vera Freundl, for meticulously reviewing my entire dissertation. I also want to extend my appreciation to all the research assistants and interns – thank you for your contributions.

My dad deserves a shout-out for diving into my early bachelor's essays and helping me refine my academic writing. Whenever my to-do list seems overwhelming, I think of my mom and how she seems to manage twice as much in a single day. I am incredibly lucky to have such a loving, supportive, and fun family, which importantly includes my partner's family. Especially while starting my PhD during the lockdown, it was a real comfort to be close to them and some of my oldest friends, Anna, Lisa, and Magdalena. Thank you, Charlotte, Eva, Malla, Toni, Lisa and Pascal, Armin and Daisy, my climbing community, and all my friends for the amazing adventures and for the balance you brought to my PhD life. Finally, Tyll provided love, food, and 60% of the care-work[1]. Remarkably, he has been by my side through my bachelor's thesis, my master's thesis, and my dissertation. Thanks for sticking with me through it all!

[1] According to the care work calculator by Die Zeit, WZB, and infas.

Contents

List of Figures

List of Tables

1 Introduction

1.1 Why Education Matters

Education plays a crucial role in fostering economic growth, social equity, and individual well-being. The economics of education provides a lens through which to understand the complex interplay between educational investments, human capital formation, labor market outcomes, and economic development. In the 1950s and 1960s, Schultz, Mincer, and Becker laid the foundation for much of the research in the modern economics of education. Becker (1962) showed that human capital explains at least one third of the variation in labor market earnings. Mincer (1974) pioneered the positive linear relationship between log earnings and years of schooling plus (quadratic) years of potential labor market experience. Schultz, 1961 highlighted the importance of investing into human capital.

The term *human capital* faced some resistance (see Goldin and Katz, 2020 for a detailed discussion). In his presidential address to the American Economic Association in 1961, Schultz emphasized that humans should not be equated with property and assets. Moreover, economists had yet to establish that education directly caused economic gain. The correlation between education and factors such as ability or parental income could lead to significant biases. Later, Hanushek and Woessmann, 2012a provided evidence on the causal effect of better schools on long-term economic growth.

Human capital theory is one of the most notable achievements in the field of economics (Deming, 2023). The proportion of adults worldwide with secondary schooling increased from 13 percent in 1950 to 51 percent in 2010 (Lee and Lee, 2016). People now dedicate more time to education than in the past, reflecting a widespread acknowledgment of its value. This recognition is embodied in the United Nations Sustainable Development Goals (SDGs), particularly SDG 4, which aims to ensure inclusive and equitable quality education for all (UNESCO, 2021).

A key issue is how to best measure human capital. Education is often measured by (years of) schooling, but this approach is misleading as it fails to account for variations in educational quality and actual learning (Hanushek and Kimko, 2000; Lee and Barro, 2001; Filmer et al., 2020). One solution is to consider student achievement, which captures the cognitive skill component of human capital. International large-scale student assessments, such as the Programme for International Student Assessment (PISA) and the Trends in International Mathematics and Science Study (TIMSS), provide good coverage for OECD countries, offering valuable insights into educational quality and skills. Initiatives like PISA for Development aim to extend this coverage, highlighting the low learning levels prevalent in low-income countries (Pritchett, 2013; Pritchett and Viarengo, 2023). However, no globally comparable skill measure exists. This means that it is currently unclear how far we are from achieving the goal of inclusive and equitable quality education for all. *Chapter 2* of this dissertation, which is joint work with Eric A. Hanushek and Ludger Woessmann, addresses this gap by constructing a new database on the share of children not reaching basic skills. We find that two-thirds of the world's youth are not achieving basic skills, and that the associated lost economic output is $700 trillion over the remaining century.

Skills, rather than schooling alone, provide a more sobering assessment of global education. UNESCO, UNICEF, and the World Bank have jointly declared a global learning crisis (UNESCO, UNICEF and Weltbank, 2021). This learning crisis was further amplified by the COVID-19 pandemic. During the peak of the pandemic in April 2020, school closures affected 1.6 billion children worldwide (UNESCO, UNICEF and Weltbank, 2021). In Germany, only 7 percent of schools provided daily online instruction during the initial closures (Werner and Woessmann, 2023). In 2022, students in Germany performed even worse on the PISA test than they did in 2000, when their poor performance first led to what became known as the "PISA shock". Azevedo et al. (2021) estimate that learning losses resulting from 5 months of school shutdowns could lead to future earnings losses totaling approximately $10 trillion globally.

The pandemic is not the only factor amplifying the learning crisis. One alarming trend is the rise in state-based conflicts (Rustad, 2024), which means that children receive less education and suffer long-run losses (Ichino and Winter-Ebmer, 2004). Before the pandemic, natural disasters were the most frequent reason for prolonged unplanned school closures in the US before the pandemic (Jahan et al., 2022). The rising risk of natural disasters and extreme weather due to climate change poses significant challenges globally. *Chapter 3* of this dissertation investigates the effects of natural disasters on student achievement. I find persistent negative effects of natural disasters on student achievement for up to five years after a natural disaster hits a county. Natural disasters adversely affect even younger children who were not yet in school at the time of the disaster. The *education production function* can help us understand how natural disasters affect student outcomes by illustrating the various inputs and processes involved in education, as outlined in *Chapter 3*.

Generally, the *education production function* shifts the focus from education as an input for outcomes like earnings to education as an outcome with various input factors (Hanushek, 2020). School inputs, such as school organization (including class sizes, facilities, and administrative spending) and teacher characteristics, are among several factors that affect learning outcomes (Hanushek, 2020). However, there are other factors, such as families, peers, individual characteristics of the child, and external factors, that determine educational success. More specifically, family background takes socio-demographic factors, such as parental education, income, and family size, into account. The more family background affects educational performance, the greater the inequality in opportunities. Typically, the family background is a much stronger predictor for educational success than, for example, resource endowment of schools (Woessmann, 2005). Therefore, it is essential for both scholars and policymakers to consider the 'home production' aspects of education.

In the tradition of the education production function, recent contributions in the economics of education study the formation of non-academic outcomes (Almlund et al., 2011; Koch et al., 2015; Arold et al., 2022; Schoner et al., 2024). *Chapter 4* of this dissertation contributes to this literature by shedding light on how political attitudes and actions are shaped during adolescence. Civic engagement is crucial for a functioning democracy and for strengthening community bonds, which contributes to a more equitable and responsive society. We leverage the distinctive German context, shaped by numerous reforms that altered the presence and intensity of civic education, leading to substantial variations across states and school tracks. Our analysis indicates that introducing civic education as a new subject positively impacts civic engagement. However, increasing the average instructional hours has negligible effects on average. *Chapter 4* is joint work with Sven Resnjanskij, Larissa Zierow, Marcel Helbig, and Norbert Sendzik.

Overall, the three chapters of this dissertation enhance our understanding of three specific challenges in the economics of education: global universal basic skills, the effects of natural disasters on learning, and the development of civic engagement. All three challanges are crucial for promoting inclusive economic growth, strengthening societal cohesion, and enhancing individual well-being.

The remainder of the introduction of this dissertation is structured as follows. In section 1.2, I describe the data compiled for this dissertation. *Chapter 3* and *Chapter 4* rely on a difference-in-differences estimation which I introduce in section 1.3. Section 1.4 gives an overview of the different dissertation chapters. Section 1.5 concludes by discussing the relevance and policy implications of this dissertation.

1.2 Data

One key contribution of this dissertation is the development and use of new data sources. For *Chapter 2*, we constructed a new, global data base on the share of children below basic skills to address the lack of a globally comparable skill measure. In *Chapter 3*, I combine the Stanford Educational Data Archive with FEMA disaster declarations in the US to study the effect of natural disasters on student achievement. In *Chapter 4*, we collect a unique new data set on the weekly hours of civic education in German secondary schools and combine it with individual data from the German Socio-Economic Panel (SOEP) to study long-term effects of civic education on civic engagement.

1.2.1 Achievement of Basic Skills Around the Globe

To measure the share of students not reaching basic skill levels globally, *Chapter 3* utilizes various internationally comparable student achievement tests in math and science. Each assessment evaluates representative samples of students, and we use the most recent data available for each country. The core of our analysis is a new method for linking scores across the different international tests that allows us to construct country-by-country estimates of deficits in basic skills. To understand the differing uncertainty, we group countries into five layers based on the reliability of the underlying achievement data.

The most reliable assessments in Layer 1 come from 90 countries that participated in a test of the Programme for International Student Assessment (PISA). PISA measures the math, science, and reading achievement of 15-year-old students in school in participating countries every three years. The Trends in International Mathematics and Science Study (TIMSS) assesses mathematics and science knowledge of students in grades 4 and 8 every four years. TIMSS allows us to add 14 countries from lower income groups that have not participated in PISA. We assign these countries to Layer 2.

In addition to the globally oriented achievement tests PISA and TIMSS, Layer 3 includes 20 countries from regional tests in Latin America and Sub-Saharan Africa. These include the Tercer Estudio Regional Comparativo y Explicativo (TERCE), which evaluates students in Latin America and the Caribbean; the Segundo Estudio Regional Comparativo y Explicativo (SERCE), which also focuses on educational quality in Latin America and the Caribbean; the Programme d’Analyse des Systèmes Educatifs de la CONFEMEN (PASEC), which measures educational outcomes in French-speaking West and Central Africa; and the Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ), which assesses reading and math in Southern and Eastern Africa.

India and China did not participate in any recent international test with nationally representative samples. However, sub-national regions of both countries participated in a PISA cycle with samples drawn to be representative for the participating regions. We combine this regionally representative information on the PISA scale with national achievement information of the respective regions relative to the countries’ other regions to derive achievement estimates for India and China. Layer 4 incorporates this sub-national data from India and China.

Layer 5 imputes data for 33 countries without comparable tests using GDP, school enrollment, and regional achievement data.

While international tests provide data on children attending school, more than a third of the world’s children are not enrolled in secondary education, and their skills remain largely unmeasured. PISA-D and the Programme for the International Assessment of Adult Competencies (PIAAC) provide information on the skill levels of out-of-school children relative to their peers in school within specific countries.

Our final dataset includes all countries with populations of at least one million or representing at least 0.01% of world GDP, excluding North Korea, Somalia, South Sudan, Syria, Venezuela, and Yemen due to unreliable data. This results in a sample of 159 countries, covering 98.1% of the world population and 99.4% of global GDP.

1.2.2 Stanford Education Data Archive

Another source enhancing the ability to analyze education outcomes across space is the Stanford Education Data Archive (SEDA). SEDA provides comparable school district and county-level data on student achievement in the US from the 2008/09 to 2017/18 school years. It addresses a critical limitation of the EDFacts data system, which typically lacks comparability across states due to varying test designs and performance benchmarks. SEDA standardizes these state-specific benchmarks onto a common scale using the National Assessment of Educational Progress (NAEP). It offers data on test scores for students in grades 3 to 8, achievement gaps by gender and socio-economic status, as well as demographic and socio-economic information. Using the county-level estimates, I can match student achievement to natural disaster declarations for *Chapter 3*.

1.2.3 FEMA Disaster Declarations

For *Chapter 3*, I combine the Stanford Educational Data Archive with natural disaster declarations. The FEMA disaster declaration is an official statement issued by the Federal Emergency Management Agency (FEMA) in the United States, recognizing that a disaster or emergency situation has occurred. The OpenFEMA Dataset of the Department of Homeland Security contains all major disaster declarations since 1964. Each disaster declaration includes the date the disaster was declared, the area affected, the type of incident, and the assistance program.

1.2.4 A Novel Measure of Civic Education

For *Chapter 4*, we digitized historical legal records from various archives in Germany and compiled a novel database on the hours of civic education taught in all three major school tracks, across all West German states, from grade 5 to grade 10. This database, based on coding available legal records documenting curricular changes, captures weekly compulsory hours of civic education in lower secondary school. This longitudinal dataset details both the intensive (number of hours) and extensive margins (any hours at all) of civic education taught from 1976 onward. Our dataset includes approximately 400 schedules and legal regulations, which quantify state education policy regarding instructional time allocated to various subjects, reflecting the political priorities of different eras. These "Stundentafeln" or school hour schedules present a comprehensive measure for assessing the quantity of civic education provided for different birth cohorts, states, and school types with substantial variation across these dimensions.

1.2.5 Socio-Economic Panel

The German Socio-Economic Panel (SOEP) is a wide-ranging representative longitudinal study of private households in Germany. It collects data on various aspects of life, including demographics, employment, income, education, health, and subjective well-being. In *Chapter 3*, we derive information on civic engagement among individuals in Germany across different states, cohorts, and school tracks from the SOEP.

1.3 Difference-in-Differences Models

If we are interested in the causal effect of an educational policy or event, the underlying question is: what would have happened to the treated group in the absence of this event or reform? In the real world, we cannot observe the *counterfactual* outcome. Thus, an important task in economic research is to find a control group. A control group allows researchers to compare outcomes and isolate the effect of an intervention, ensuring that observed changes are due to the intervention and not other factors. One option is to design an experiment or randomized control trial (RCT) and randomly assign people to the treatment and the control group. Many important contributions in the economics of education rely on such randomization (Banerjee et al., 2007; Chetty et al., 2011; Kinne, 2023; Resnjanskij et al., 2024, to just name a very few). In some scenarios, this approach may be too costly, raise ethical concerns, or be impractical because the events of interest occurred in the past.

An alternative are quasi-experiments, such as the difference-in-differences design. The basic intuition is that we compare the difference in the outcomes before and after treatment of the treatment and control group to each other. Instead of a perfect randomization to treatment, the main identifying assumption is parallel trends. That is, the treatment and the control group would have developed the same in the absence of treatment. Ashenfelter and Card, 1984 coined the term difference-in-differences for this approach. Since then, DiD approaches have become more flexible because the events and policies economists want to study are often more complex than the simple two-period-two-groups design.

Recent advances in the DiD literature focus on the scenario where the treatment affects different groups at varying points in time. In such staggered designs, the estimates may be biased if treatment effects are heterogeneous across groups and over time (De Chaisemartin and d'Haultfoeuille, 2020; Callaway and Sant'Anna, 2021; Goodman-Bacon, 2021; Sun and Abraham, 2021; Wooldridge, 2021; Athey and Imbens, 2022; Borusyak et al., 2024).

In *Chapter 3* of this dissertation, I analyze the effect of natural disasters on students in the United States using county-level disasters and student outcomes. Some countries might adapt more quickly to natural disasters than others, which would bias the results of a traditional DiD. Sun and Abraham, 2021 propose a solution for the event-study context. Examining the event-study estimates allows us to analyze the temporal evolution of treatment effects on students, providing insights into both the immediate and long-term impacts of natural disasters. The Sun and Abraham, 2021 method follows a two step procedure: First, we estimate the average treatment effect on the treated for each cohort separately, where cohorts are groups treated at the same time. Second, we take the weighted average over all estimates and multiply it by the sample share of each cohort in the period. This estimator is consistent under parallel trends and limited anticipation.

In *Chapter 4* of this dissertation, we study how civic education in school affects civic engagement later in life. The treatment is the average number of hours of civic education that increases and decreases over different cohorts, states, and secondary school tracks in Germany. These treatment characteristics further complicate the estimation. Borusyak et al., 2024, Callaway and Sant'Anna, 2021, Goodman-Bacon, 2021, and Sun and Abraham, 2021 assume that the treatment is binary and absorbing, i.e., that groups do not switch out of treatment once treated. Some estimators allow for switching in and out of treatment but assume a binary treatment (Wooldridge, 2021). De Chaisemartin and d'Haultfoeuille, 2024 propose a solution for non-absorbing, continuous treatments. The estimator compares the effect of moving from a treatment dose d to d' to groups that still experience treatment dose d. With this estimator, we identify effects from differences in adult outcomes between cohorts within the same school track in states that experienced alterations in civic education hours, relative to cohorts in the same school track in unaffected states.

1.4 Chapter Overview

This section provides an overview of the three essays that comprise this dissertation, each of which can be read independently. Education is a key development goal, offering a wide range of individual and societal benefits. In *Chapter 2*, we assess the number of children worldwide who still lack access to high-quality education and fail to reach basic skill levels. *Chapter 3* demonstrates that natural disasters pose a significant threat to student learning outcomes. *Chapter 4* goes beyond student achievement to explore whether civic education can foster civic engagement.

Chapter 2 provides new approaches for estimating the lack of basic skills that allow mapping achievement across countries of the world onto a common (PISA) scale. We define *basic skills* as the skills needed to participate effectively in modern economies. We base our measure on students reaching at least Level 1 on the PISA scale, the minimum proficiency standard. Half of the world's population lives in countries that fully participate in some psychometrically-validated testing, such as PISA or TIMSS. This coverage extends to 85 percent of the global population when including countries with regional test participation. We impute skill estimates for countries without test information based on observed test performance in comparable countries. While international tests provide data on children attending school, more than a third of the world's children are not enrolled in secondary education, and their skills remain largely unmeasured. We use information on the skill levels of out-of-school children relative to their peers in school within specific countries. Our comprehensive dataset includes the share of children (not) reaching basic skill levels for 159 countries that cover over 98 percent of the world population and over 99 percent of world GDP. The main findings reveal profound global challenges in achieving universal basic skills. A significant majority of the world's youth, over two-thirds, do not acquire basic skills. This deficiency is widespread across 101 countries, where more than half of children fail to reach basic proficiency, with rates exceeding 90 percent in 36 countries. Even in high-income nations, a quarter of children lack fundamental skills. Despite attending secondary school, 63 percent of global youth do not achieve basic skills. Our growth projections suggest that the present value of lost world economic output due to missing the goal of global universal basic skills amounts to over $700 trillion over the remaining century, or 12% of discounted GDP.

In *Chapter 3* of this dissertation, I investigate the effects of natural disasters on student achievement and explore the underlying mechanisms. Specifically, focusing on student achievement allows for a nuanced examination of the cognitive skill component of human capital, capturing variations in skills, family inputs, school resources, and institutional characteristics. Estimating a staggered two-way fixed effects model, I find persistent negative effects of natural disasters on student achievement for up to five years after a natural disaster hits a county. Natural disasters adversely affect even younger children who were not yet in school at the time of the disaster. School closures cannot explain these long-run effects. Possible explanations include a reduction in the health stock of children, which impedes cognitive and social-emotional development, potential shifts in student composition, a reduction in school quality, and lower parental investments in education. Counties with above-average per-pupil expenditure recover more quickly from natural disasters. Exploring different treatment intensities shows that not only very large disasters drive the results. Although point estimates are consistently higher for large disasters, below-average disasters have a significant negative effect on student achievement.

In the midst of a rising support for populist parties in several European countries (Noury and Roland, 2020) and a resurgence of racism and antisemitism (Huneke, 2024), civic education at school is often highlighted as a potential remedy (Manning and Edwards, 2014b; Hamm et al., 2023). In *Chapter 4*, we investigate the relationship between civic education in schools and civic engagement in adult life, drawing from a comprehensive dataset spanning over four decades. Employing the new continuous difference-in-differences methodology developed by De Chaisemartin and d'Haultfoeuille (2024), we identify effects on adult outcomes by comparing state-school-track groups that experienced changes in civic education hours between cohorts with those that did not, starting from the same baseline. Our analysis shows that introducing civic education as a subject (at the extensive margin) has a positive impact on civic engagement. However, findings regarding the intensive margin, namely the impact of increased average instructional hours, are less definitive, showing negligible effects on average. Our heterogeneity analysis underscores that individuals from lower socio-economic backgrounds tend to derive greater benefits from civic education at the extensive margin.

1.5 General Discussion and Conclusion

This dissertation examines three critical challenges faced by contemporary societies striving to achieve inclusive economic growth, social equity, and individual well-being. Through the perspective of the economics of education, these challenges are:

1. **Providing Universal Basic Skills**: Ensuring that all individuals have access to fundamental educational resources and skills is essential for fostering economic opportunities and social equity.
2. **Climate Change and Natural Disasters**: Climate change and natural disasters pose significant threats to economic stability and societal well-being.
3. **Enhancing Civic Engagement**: Civic engagement is crucial for a functioning democracy and for strengthening community bonds, which contributes to a more equitable and responsive society.

Chapter 2 suggests that the world is severely lacking in its ability to achieve the objective of universal quality education, and this leaves many in the world short of the basic skills needed to participate in modern economies. Approximately two-thirds of youth globally lack basic skills, with the highest deficits in Sub-Saharan Africa (94%), South Asia (89%), MENA (68%), and Latin America (65%). Even in North America and Europe, about a quarter of youths fall short of basic skill levels.

The developing world faces challenges in both school access and quality. Over one-third (36%) of secondary-school-aged youth globally do not attend school, and among those enrolled, 63% fail to acquire basic skills. This suggests that merely attending low-quality schools does not address the problem of inadequate skills. While developed countries have largely addressed attainment issues, they still struggle with quality, leaving some students behind.

The vast education deficits in the global South are made even more important by changes in the global economy: with integrated economies, people are no longer just competing with workers in adjoining cities or states, as most products can be produced anywhere in the world. Moreover, the lack of skills of the potential labor force has immense consequences for global economic development. According to our projections based on historical patterns of long-run growth, the world would gain $732 trillion in added GDP over the remaining century if it were to reach global universal basic skills. These results underscore the urgent need for policymakers worldwide to prioritize and significantly enhance efforts toward ensuring quality education for all children. Our analysis offers a global view of basic skill distribution but is limited by uncertainty, especially in regions that lack regular participation in international testing. Establishing a regular, internationally standardized test, similar to PISA, for these regions would greatly enhance global development efforts. Such assessments, aligned with international standards and focused on basic skills, would offer policymakers valuable insights and potentially yield greater long-term benefits than current development aid.

Chapter 3 reveals persistent negative impacts on student achievement following natural disasters, with students experiencing setbacks for up to five years post-event. The results are particularly concerning given the increasing frequency and severity of natural disasters due to climate change. For the affected regions, the decrease in student achievement implies a depletion in the human capital reservoir, resulting in long-term economic damage. In terms of policy implication, *Chapter 3* also shows that counties with higher per-pupil expenditure demonstrate a swifter recovery compared to their lower-spending counterparts, indicating that augmenting per-pupil spending can enhance community resilience against human capital erosion caused by natural disasters. Future inquiries should delve into the specific types of educational investments pivotal in shielding against disaster-induced damage.

Chapter 4 highlights the positive impact of introducing civic education as a subject, revealing a statistically significant positive effect on civic engagement at the extensive margin. When considering an increase in the number of hours dedicated to civic education, we find that the average effects are negligible. While the political debate often emphasizes the need for stronger integration of civic education, we find that simply increasing the hours taught is no panacea. However, our study faces certain limitations. While our dataset offers detailed insights into the quantitative aspect of civic education provision, it does not capture the qualitative dimensions of pedagogical content. Similar to the discussion on general education, improving the quality of civic education could be an important leverage point, that is outside the scope of this dissertation. Additionally, relying on survey-based data from the German Socio-Economic Panel imposes constraints on the breadth of outcomes. For example, administrative election participation data and data on students' knowledge of civic education topics could deliver further important insights on the effects and mechanisms of the reform.

2 Global Universal Basic Skills: Current Deficits and Implications for World Development*

2.1 Introduction

Ensuring that all of the world's youth have at least basic skills is a prime development goal by itself, but reaching that goal also is immensely important for inclusive and sustainable world development. The United Nations' Sustainable Development Goal (SDG) 4 calls for ensuring quality education for all (e.g., UNESCO, 2021). Learning levels appear very low in some low-income countries (e.g., Pritchett, 2013; Pritchett and Viarengo, 2023), but the limited country coverage of international skill data makes it unclear globally how many children currently fail to reach basic skill levels. This paper addresses the two intertwined questions underlying the development goals: How close are we to reaching the foundational goal of basic skills for all? And what would it mean for world development to reach global universal basic skills? We supplement available data from international and regional student assessments with analyses of skills in untested countries to develop world estimates of the share of children not achieving basic skills in each country and then show the economic costs of these deficits.

While the term can take on different meanings, we conceptually define "basic skills" as the skills needed to participate effectively in modern economies. We anchor our measure at mastering at least the most basic level of the Programme for International Student Assessment (PISA), i.e., PISA Level 1 skills.[1] An important methodological contribution is to use the underlying individual-level distributions to develop reliable cross-country skill measures from the disparate available tests. Test-based measurement of skills provides comparable estimates of skill deficits covering the half of the world's population living in countries that fully participate in such psychometrically-validated testing, and this expands to 85 percent of the world population by including countries with partial test participation. Extending estimates to the entire world requires imputation based on the observed test performance in comparable countries.

---

* This chapter is co-authored with Eric A. Hanushek and Ludger Wößmann. It is based on the paper 'Global Universal Basic Skills: Current Deficits and Implications for World Development' published in the Journal of Development Economics 166 (2024): 103205. The project was supported by the Smith Richardson Foundation.

[1] Our focus on reaching basic skill levels is similar in spirit to the World Bank's learning poverty (LP) metric (Azevedo et al. 2021). There are some significant differences in the details that are related to the fact that the LP indicator is aimed at identifying lack of basic skills at very early ages, whereas we are interested in measuring skill deficits of people closer to labor-force participation ages. In addition, our analysis concentrates on constructing comparable scales.

We separate the estimation of world skill levels into five layers of decreasing reliability that indicate different degrees of certainty and precision in the comparability of available test information. Layer 1 includes countries that have participated in any wave of PISA or PISA for Development (PISA-D) – a total of 90 countries. Layer 2 adds countries that have participated in the Trends in International Mathematics and Science Study (TIMSS) but not in PISA – 14 additional countries. Layer 3 incorporates countries that have participated in regional tests – TERCE and SERCE in Latin America and SACMEQ and PASEC in Sub-Saharan Africa – but not in PISA or TIMSS, an additional 20 countries. Layer 4 adds the two countries – India and China – that have some sub-national test information but that have not fully participated in the international assessments.

Even though the different tests were not designed with cross-test comparisons in mind, we show that it is possible to transform student-level achievement on all tests into a PISA-equivalent score while introducing minimal constraints on the underlying score distributions. Our method equates the scales of the different tests by using the student-level distributional information found in the group of countries that participate in multiple test regimes. We rescale the performance of countries participating only in TIMSS (Layer 2) or in one of the regional tests (Layer 3) onto the PISA scale using the underlying distributional information from countries that jointly participate in these and in PISA. From the resultant individual-level database on achievement distributions, we can find the observed share of students (not) reaching basic skill levels.

The full achievement distributions provide common support at the student level which is fundamental to our harmonization of scores across tests. This is particularly relevant for countries in Sub-Saharan Africa that perform outside the observed range of average achievement on the broad international tests. Previous transformation methods based on linear extrapolation from country mean scores tend to overestimate these countries' true achievement levels.

Estimating achievement of basic skills in India and China (Layer 4) which did not participate representatively in the international tests adds an additional level of complexity. Both participated in PISA with selected provinces or states, and we use additional within-country achievement information to provide estimates of national achievement on the PISA scale.

For countries that never participated in any of the international tests (Layer 5), we resort to imputations of achievement based on cross-country regressions of achievement on educational enrollment, per-capita GDP, and indicators of world regions and income groups. While obviously not ideal, imputation provides a way to get world estimates of skill deficits.

Finally, the international tests provide data on children in school, but over a third of the world's children are out of secondary schools, and their skills are generally not measured. We use information from PISA-D and from the Programme for the International Assessment of Adult Competencies (PIAAC) to estimate the skill levels of children who are not in school (relative to children in school in the specific country).

The 126 countries with direct assessments of students (Layers 1-4) represent 85 percent of the world population and 96 percent of world GDP. The imputations for the 33 countries in Layer 5 allows us to provide estimates of achievement deficits in all 159 countries that have a population of at least one million or a GDP that is at least 0.01 percent of world GDP. These 159 countries cover over 98 percent of the world population and over 99 percent of world GDP.

Our estimates show that the world has a long way to go to reach global universal basic skills. The world distribution of basic skills can be summarized in six stylized facts:

1. At least two-thirds of the world's youth do not obtain basic skills.

2. The share of children not reaching basic skills exceeds half in 101 countries and rises above 90 percent in 36 of these countries.

3. Even in high-income countries, a quarter of children lacks basic skills.

4. Skill deficits reach 94 percent in Sub-Saharan Africa and 89 percent in South Asia but also hit 68 percent in Middle East and North Africa and 65 percent in Latin America.

5. While skill gaps are most apparent for the third of global youth not attending secondary school, 63 percent of the world's secondary-school students fail to reach basic skills.

6. Half of the world's youth live in the 35 countries that fail to participate fully in international tests (which includes India and China) and thus lack regular and reliable foundational performance information.

To address the uncertainty in the underlying skill data, we perform a range of sensitivity analyses to provide bounds on the baseline results. The analyses address the reliability of the different layers, the imputations performed, and the skills of out-of-school children.

We use our skill measures to quantify the economic gains that the world could reap from reaching the goal that every child achieves at least a basic skill level. Using estimates of the association between skills and long-run growth rates from existing empirical growth models with worker skills, we project country by country the future path of GDP with improved skills.

The discounted added world GDP amounts to $732 trillion compared to the status quo GDP trajectory over the remaining century. This economic gain from reaching global universal basic skills is over five times the current annual world GDP, or 11.6 percent of the discounted future GDP over the same horizon. This amount documents the lost economic output due to missing the goal of global universal basic skills. Importantly, the gain from lifting all students who are currently in school to at least basic skill levels turns out to be more than twice as large as the gain from enrolling the children currently not attending school at current quality levels.[2]

Our method for combining tests using the full student distributions extends the literature on international skill measurement (e.g., Das and Zajonc, 2010; Hanushek and Woessmann, 2012a; Angrist et al., 2013; Altinok et al., 2014; Lim et al., 2018; Sandefur, 2018; Patel and Sandefur, 2020; Angrist et al., 2021; Fuente and Doménech, 2024) and provides consistent estimates for the whole world. Our global perspective on human capital and economic growth contributes to the empirical literature on skills and growth (Barro, 1991; Mankiw et al., 1992; Hall and Jones, 1999; Bils and Klenow, 2000; Hanushek and Kimko, 2000; Krueger and Lindahl, 2001; Hanushek and Woessmann, 2008; Ciccone and Papaioannou, 2009; Hanushek and Woessmann, 2012a,b; Barro and Lee, 2015; Hanushek and Woessmann, 2015a, 2016; Lee and Lee, 2020; Angrist et al., 2021). The projection model follows previous applications for OECD countries (Hanushek and Woessmann, 2011, 2015b, 2020a) and US states (Hanushek et al., 2017a,b).

In what follows, we describe the data (section 2.2), the scaling method (section 2.3), and the results on basic skills (section 2.4) and on economic projections (section 2.5).

2.2 Data: Five Layers of Information from Student Achievement Tests

To measure the share of students not reaching basic skill levels in each country, we draw on various student achievement tests that have been designed to provide internationally comparable achievement information in math and science.[3] Each assessment tests representative samples of students in the participating countries, and we use the most recent assessment available for each country.[4] The different tests use different sets of questions and

[2] To be clear, the measures of basic skills and of educational outcomes are not a measure of school quality, because they reflect a wide range of educational inputs including families, poverty, nutrition, and more general societal institutions which may also work through factors such as delayed enrollment or grade repetition.

[3] Achievement in math and science may be more readily compared across countries than in reading because of language differences (Hanushek and Woessmann, 2012a). Nonetheless, at both the individual and country level, math, science, and reading scores are highly correlated. For example, in PISA 2018 the country-level correlation of reading achievement is 0.834 with math achievement and 0.884 with science achievement. Furthermore, adding the Progress in International Reading Literacy Study (PIRLS), which tests reading in fourth grade, would not expand the number of included countries. Thus, we rely on just math and science scores.

[4] We do not include assessments where tests are specifically adapted by participating countries or where participating populations are not drawn to be representative, such as many countries in the Early Grade Reading Assessment (EGRA) (included in Angrist et al., 2021). EGRA is useful for providing information to participating school systems, but it is not designed to provide nationally representative information that is comparable across countries (Dubeck and Gove, 2015)

have different target populations. We assume that each of them measures the underlying distribution of math and science skills in the participating countries.[5]

Our analytic sample includes all countries in the world that have a population of at least one million and all countries that represent at least 0.01 percent of world GDP but excluding North Korea, Somalia, South Sudan, Syria, Venezuela, and Yemen that lack reliable GDP and population data. This leaves us with an analysis sample of 159 countries that covers 98.1 percent of the world population and 99.4 percent of world GDP.

The information from the different tests are transformed onto a common international scale with varying degrees of certainty and precision. To understand the differing uncertainty, we group countries into five layers based on the reliability of underlying achievement data.

Layer 1: Countries participating in PISA. The most reliable assessments come from countries that have participated in a test of the Programme for International Student Assessment (PISA). Set up by the Organisation for Economic Co-operation and Development (OECD) in 2000, PISA measures the math, science, and reading achievement of 15-year-old students in school in participating countries every three years (OECD, 2019). This layer where achievement is directly measured on the PISA scale includes 90 countries, covering 37 percent of the world population and 66 percent of world GDP (Table 2.1). Unfortunately, no low-income country has ever participated in PISA.

There are three subsets of countries in Layer 1. First, the largest group of countries is those participating in the most recent international PISA cycle, 2018.[6] Representative achievement data from this cycle are available for 75 countries – 47 high-income countries, 24 upper-middle-income countries, and 4 lower-middle-income countries. The countries participating in PISA 2018 cover one third of the world population and nearly two thirds of world GDP.

Second, an additional eight countries, including another four lower-middle-income countries, participated in a previous PISA cycle but not in 2018: five in 2015, one in 2012, and two in 2009. As the different PISA cycles measure achievement on a psychometrically linked scale, their achievement scores are directly comparable with the most recent PISA cycle.

[5] Throughout this analysis, we assume that the different testing regimes produce unbiased measures of country-level skills. While various measurement errors can influence individual scores, we assume that these are averaged out in the large country-level samples. Some questions have been raised about systematic country-level differences arising from test-taking effort as opposed to underlying skills (e.g., Gneezy et al., 2019; Zamarro et al., 2019; Hanushek et al., 2022), but the evidence is not consistent across other studies (Baumert and Demmrich, 2001). It is also unclear to what extent effort differences are part of skill differences. We do not believe that these effects have a major influence on the level and pattern of skill deficits as we define them, but we are unable to analyze any possible biases directly.

[6] The Covid-19 pandemic postponed the 2021 PISA cycle to 2022; data will be released at the end of 2023.

Third, another seven countries participated in PISA for Development (PISA-D). The PISA-D initiative was launched by the OECD together with partner organizations to develop the PISA data-collection instruments for participation by interested low- and middle-income countries (OECD, 2018a,b). Seven countries administered the PISA-D assessment in 2017 – Cambodia, Ecuador, Guatemala, Honduras, Paraguay, Senegal, and Zambia.

Layer 2: Countries participating in TIMSS. A second important source of internationally comparable achievement information is the Trends in International Mathematics and Science Study (TIMSS). Emerging from prior occasional international testing, the International Association for the Evaluation of Educational Achievement (IEA) established TIMSS in 1995 and implemented it on a four-year cycle through 2019. TIMSS tests the math and science achievement of students in fourth and eighth grade (Mullis et al., 2020). TIMSS allows us to add a number of countries from lower income groups that have not participated in PISA. We assume that PISA and TIMSS are consistently measuring the same skills so that we can merge countries with just TIMSS scores into the overall skill distribution of Layer 1 countries. We add seven countries that participated in the most recent TIMSS eighth-grade assessment in 2019 plus another six countries that participated in a prior eighth-grade assessment (two each in 2015, 2011, and 2007). While we generally use the eighth-grade results, we rely on fourth-grade results for one country, Pakistan, which participated only in the fourth-grade assessment in TIMSS 2019. Together, the TIMSS assessments add fourteen countries to our analysis, including ten middle-income countries, representing seven percent of the world population (Table 2.1).

Layer 3: Countries participating in regional tests – TERCE, SERCE, SACMEQ, and PASEC. In addition to the globally oriented achievement tests PISA and TIMSS, there are a series of regional achievement tests in Latin America and Sub-Saharan Africa. In Latin America, the Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación (LLECE) conducts regional tests of student achievement in math and reading in grades three and six. While most of the participants of the LLECE tests also participated in either PISA or TIMSS, we use the sixth-grade math test to obtain information on student achievement in Nicaragua in 2013 from the Tercer Estudio Regional Comparativo y Explicativo (TERCE) and in Cuba in 2006 from the Segundo Estudio Regional Comparativo Explicativo (SERCE).

Two regional tests provide achievement information for many Sub-Saharan African countries that did not participate in the global tests. The Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ) provides a testing cycle of math and reading achievement of sixth-grade students in multiple countries in Southern and Eastern Africa (Bietenbeck et al., 2018). We draw on the results of the most recent wave with released micro data, the SACMEQ III test conducted between 2006 and 2011, to extend our analysis by nine countries that did not participate in PISA or TIMSS.[7]

The Conférence des ministres de l'Éducation des Etats et gouvernements de la Francophonie (CONFEMEN) has established a testing cycle of math and reading achievement of sixth-grade students in francophone Sub-Saharan Africa, the Programme d'analyse des systèmes éducatifs de la CONFEMEN (PASEC). The PASEC 2014 cycle provides us with achievement information for nine francophone countries in Sub-Saharan Africa that did not participate in PISA or TIMSS.

Together, the regional tests in Latin America and Sub-Saharan Africa provide us with achievement information for an additional twenty countries (beyond PISA and TIMSS), eight of which are in the low-income group and ten in the lower-middle-income group. These countries represent five percent of the world population.

Layer 4: Countries with sub-national regions participating in PISA. The two countries with the largest populations in the world, India and China (together accounting for 36 percent of the world population), did not participate in any recent international test with nationally representative samples.[8] However, sub-national regions of both countries participated in a PISA cycle with samples drawn to be representative for the participating regions. In 2010, the two Indian states Tamil Nadu and Himachal Pradesh took the test of the PISA 2009 cycle, and Shanghai participated in PISA 2012.[9] We combine this regionally representative information on the PISA scale with national achievement information of the respective regions relative to the countries' other regions to derive achievement estimates for India and China (see section 2.3.3).

Layer 5: Countries not participating in comparable international achievement tests. Layers 1-4 provide achievement information for 126 of the 159 countries in our analysis sample, corresponding to 84.8 percent of the world population and 95.7 percent of world GDP. The remaining 33 countries (covering 13.3 percent of the world population and 3.6 percent of world GDP) did not participate in any internationally comparable achievement test. In our analysis, we impute achievement in these countries based on data on per-capita GDP, secondary-school enrollment, and achievement data from countries in the same world region and income group.

[7] Unfortunately, the micro data of the more recent SACMEQ IV test (2012-2014) have not been made available at the time of our analysis.

[8] In 1971, India participated in the first international science study of the IEA (Comber and Keeves, 1973).

[9] We rely on the 2012 testing in Shanghai because of the availability of other data that permit estimating scores for the entire country. Four additional of the most developed Chinese cities and provinces – Beijing, Guangdong, Jiangsu, and Zhejiang – participated in varying combinations in PISA 2015 and/or 2018.

2.3 Methods: Depicting Skills on a Common Global Scale

We begin by defining basic skill levels (section 2.3.1). We then describe our core method for transforming the various international test distributions onto the PISA scale (section 2.3.2). India and China require special approaches described in section 2.3.3. Section 2.3.4 describes the imputation of achievement in countries without international test participation, and section 2.3.5 develops the estimation of skill levels of children not attending secondary school.

2.3.1 Defining Basic Skills

The modern world economy is internationally competitive with strong production linkages across country borders. Workers are not just competing with others in their own country for employment and wages but also with workers in other countries. The location of production and participation in it depends importantly on the skill levels of a nation's people and the way that a country's development builds upon the aggregate skills of its populations.

There is no currently accepted standard for the minimal skills required to be internationally competitive in a modern information-intensive economy. Consistent with our focus on long-run economic growth, we think of development as minimally requiring individuals to have the skills that would allow them to be successful in economies that look like those of today's high-income countries. We adopt the pragmatic definition that basic skills correspond to the PISA Level 1 skills (fully attained), the lowest of the six performance levels defined on the PISA scale.[10]

The OECD, 2019 describes the conceptual differences in what students should know at different proficiency levels for math as follows:

> "At Level 1, students can answer questions involving familiar contexts where all relevant information is present and the questions are clearly defined. They are able to identify information and carry out routine procedures according to direct instructions in explicit situations. They can perform actions that are almost always obvious and follow immediately from the given stimuli."

> "At Level 2, students can interpret and recognize situations in contexts that require no more than direct inference. They can extract relevant information from a single source and make use of a single representational mode. Students at this level can employ basic algorithms, formulae, procedures or conventions to solve problems involving whole numbers. They are capable of making literal interpretations of results."

[10] The same standard is used by Filmer et al., 2006 to develop Millennium Learning Goals.

The border line between Levels 1 and 2 is 420 points on the PISA math scale and 410 points on the PISA science scale where the OECD mean is 500 with a standard deviation of 100 (OECD, 2019).

This definition of basic skill levels may be thought of as a modern definition of functional literacy. Without the necessary skills to compete and thrive in the modern world economy, many people are unable to contribute to and participate in development gains. Literacy was once defined in terms of the ability to read simple words. But in today's interconnected societies, it is far more. It is the capacity to understand, use, and reflect critically on written information, to reason mathematically and use mathematical concepts, procedures, and tools to explain and predict situations, and to think scientifically and draw evidence-based conclusions. For development, citizens around the world will need the basic skills that industrial employers seek and that the formal labor market rewards. While some developing countries may today appear to have economies unprepared to employ even basic skills fully, past analyses (described in section 2.5.1) suggest that even subsistence agriculture can benefit from basic education and that the natural evolution of economies involves expansion of technologies that employ the available skills.

2.3.2 Transforming the Other Achievement Tests onto the PISA Scale (Layers 2 and 3)

The core of our analysis is a new method for linking scores across the different international tests that allows us to construct country-by-country estimates of deficits in basic skills. The scales that the different test regimes use to document achievement are not directly comparable to one another, even if their arbitrary choice of a common mean and variance makes them appear to be consistent with each other. [11]

Our transformation builds on the fact that there is a subset of countries – which we call linking countries – that take the PISA test along with TIMSS or one of the different regional tests. We interpret the distribution of scores from representative samples for the two distinct test regimes within each linking country as being two different samples of performance from a common underlying skill distribution. If the student-level scores follow a normal distribution, the mean and standard deviation from the student-level data provide the conversion parameters needed to equate each of the tests to the PISA scale.[12]

Consider first the TIMSS conversion to the PISA scale (Layer 2). If scores are normally distributed and we know the true mean and standard deviation of scores for TIMSS and PISA, i.e., $N(\mu_{TIMSS}, \sigma_{TIMSS})$ and $N(\mu_{PISA}, \sigma_{PISA})$, we can convert any individual TIMSS score,

[11] For example, both TIMSS and PISA were scaled at their introduction to have a mean of 500 and a standard deviation of 100. However, these statistics were established by calculations across very different sets of countries, making them inherently incomparable.

[12] Normality is a general feature of the item response theory (IRT) that is used to scale achievement on the different underlying tests.

t_i, into the corresponding PISA score, p_i, by:

$$p_i = \frac{(t_i - \mu_{TIMSS})}{\sigma_{TIMSS}}\sigma_{PISA} + \mu_{PISA} \tag{2.1}$$

That is, we first standardize achievement on TIMSS to have mean zero and standard deviation one and then assign it the standard deviation and mean of the PISA scale.

We can estimate the necessary conversion parameters from the pooled TIMSS and pooled PISA score distributions for the set of common countries, C. From the pooled individual data for the common countries, we estimate the means and standard deviations for the two distributions: i.e., $N(m^C_{TIMSS}, s^C_{TIMSS})$ and $N(m^C_{PISA}, s^C_{PISA})$. This then allows us to express achievement of students in countries that participated in TIMSS but not in PISA on the common PISA scale:[13]

$$p_i = \frac{(t_i - m^C_{TIMSS})}{s^C_{TIMSS}}s^C_{PISA} + m^C_{PISA} \tag{2.2}$$

There are 32 countries that participated both in the TIMSS 2019 eighth-grade test and PISA 2018, and these allow us to perform the re-scaling procedure from Equation 2.2.[14] (See Appendix Table A2.1 for a list of the linking countries). The correlation of average achievement scores across the two tests is 0.908 in math and 0.895 in science, providing confidence in the underlying assumption that the two tests refer to a common underlying skill distribution.[15]

Focusing on the scores in the set of countries jointly taking TIMSS and PISA assessments is important to ensure that our parameters come from the same normal distribution. Individual countries can (and do) have different means and standard deviations, so we would not want to calculate the necessary sample parameters for TIMSS and for PISA from different sets of countries.

Figure 2.1 summarizes the elements of the transformation procedure for the case of TIMSS. Panel A shows the distribution of student-level achievement on the TIMSS test for three groups of countries: all TIMSS participants, the group of countries participating in both TIMSS and PISA (linking countries), and the group of countries whose TIMSS score we would like to transform onto the PISA scale (to-be-rescaled countries). The distribution in the linking countries is quite similar to the TIMSS countries overall, whereas the distribution in the to-be-rescaled

[13] Hanushek and Woessmann, 2015b used this approach to combine the PISA 2012 and TIMSS 2011 tests. It has not been extended to any of the Layer 3 and 4 countries or used for a global analysis yet.

[14] Because the different TIMSS cycles are expressed on the same psychometrically linked scale, we can also use the same re-scaling parameters to transform performance from the prior TIMSS eighth-grade assessments onto the PISA scale. We use the 44 countries that participated both in the TIMSS 2019 fourth-grade test and in PISA 2018 to transform the score of Pakistan on the TIMSS 2019 fourth-grade test.

[15] The high correlation at the country level, as previously noted by Loveless, 2017, occurs despite the fact that TIMSS and PISA are based on different conceptual underpinnings with TIMSS being curricular based and PISA being more applied to real-world problems.

countries is shifted to the left (reflecting the addition of lower-income countries that comes from TIMSS). Quite obviously, all three distributions have a normal shape.[16]

Panel B shows performance of students on PISA for the linking countries along with the rescaled distribution of scores for the countries included in TIMSS but not in PISA. The large number of linking countries, the underlying normal distributions, and the substantial overlap of the group distributions provide confidence in the reliability of the transformation.

We repeat the same procedure with the Latin American regional tests (Layer 3) using the set of common countries C in each case to transform the regional tests onto the PISA scale. Ten countries participated both in TERCE and PISA 2018 and six countries in SERCE and PISA 2018. The large number of linking countries and the underlying normal distributions yield similarly reliable transformations even though the overall performance levels of the linking countries are significantly below those of the entire set of PISA countries (see Appendix Figures A2.1 and A2.2). Importantly, despite the large mean differences for the Latin American countries, there is substantial support in the overall PISA distribution of individual scores.[17]

Greater uncertainty arises, however, when there is a small number of linking countries that take both PISA and the regional tests. For SACMEQ and PASEC, there is only a single country that provides overlap between each of the respective regional tests and PISA: Zambia participated in SACMEQ III and PISA-D and Senegal participated in PASEC 2014 and PISA-D.[18] In each case, we use the linking country's mean achievement and its standard deviation in Equation 2.2 to transform the achievement on the SACMEQ and PASEC tests onto the PISA scale. However, estimation errors in the conversion parameters introduce additional uncertainty when there is only one linking country. Further, attributes of individual countries and the sampling of students may yield test distributions that diverge from normality, introducing other possible complications in the conversion of scores. Nonetheless, by going to the individual student test distributions (with their broad range of observed scores), we are best able to extrapolate to the range of national differences in performance.

[16] With the large student samples of the combined linking countries, standard tests for normality are uninformative.

[17] Hanushek and Woessmann, 2012b have combined the Latin American regional tests prior to TERCE with global tests, but their transformation was not based on individual-level data and distributions, but rather on country-level standard deviations and stronger assumptions on cross-country distributions.

[18] Conceptually, it would also be possible to project across multiple tests such as regional tests to TIMSS scale and then TIMSS to PISA scale if some countries participated both in the regional test and in TIMSS. In practice, however, this is only the case for one country, South Africa, which participated in SACMEQ and in TIMSS 2019, and the fact that it participated with ninth-grade students in the eighth-grade TIMSS test further complicates linkage. Still, if we use South Africa's TIMSS achievement to link the SACMEQ test to the PISA scale, results are very similar to the baseline linkage through Zambia's PISA-D achievement: the difference in the share of students in Sub-Saharan Africa estimated to fall below basic skills is less than one percentage point, suggesting highly robust estimates with the alternative linkage.

Figure 2.2 provides a summary of the transformation procedure for the case of PASEC.[19] Panel A shows the distributions of all PASEC participants, the linking country (Senegal), and the to-be-rescaled countries on the original PASEC test. Senegal performs somewhat higher than the other PASEC countries, but there is obviously ample common support across the different student test populations. Panel B shows the respective rescaled distributions on the PISA scale along with the PISA performance of Senegal. Given the relatively low achievement of the linking country on PISA, the rescaled achievement of the other PASEC countries falls far to the left on the PISA scale. Still, using the student-level micro data, ample common support allows for a valid transformation because achievement overlaps for substantial shares of students.

Our approach is externally validated by an alternative approach that uses joint test items for psychometric linkage. The UNESCO Institute for Statistics, 2022 uses joint test items to establish a concordance ("Rosetta Stone") between PASEC and TIMSS.[20] In three PASEC countries, students took both the PASEC assessment and a subset of less challenging item blocks from the TIMSS fourth-grade math test. The Rosetta Stone crosswalk allows a direct transformation of PASEC scores onto the TIMSS scale. To this we can compare a crosswalk from PASEC to TIMSS using our approach: we first transform PASEC to the PISA scale using Senegal as the linking country and then transform these scores to the TIMSS scale using the set of common countries in PISA and fourth-grade TIMSS. When plotting the results of the two methods against each other for the ten PASEC countries, the countries fall virtually on the 45-degree line (see Appendix Figure A2.4). This result strongly supports the validity of our linkage method even for the case of a single linking country.

We use the micro databases from all underlying rescaled tests to calculate the share of students not reaching basic skills – i.e., scoring below 420 (410) points on the PISA math (science) scale – in each country. From the full individual-level distribution of skills for tested students within each country, we can directly estimate the portion of the population that lacks basic math and science skills while incorporating any country-specific skewness arising from, say, special attention to the bottom of the distribution through intense compensatory programs or to the top of the distribution through limited promotion opportunities.[21]

[19] Appendix Figure A2.3 shows the respective distributions for SACMEQ.

[20] See Appendix A2.1 for another important study that uses the psychometric "Rosetta Stone" approach, Patel and Sandefur, 2020. SACMEQ and TIMSS also use a set of common test items, but analysis in Sandefur, 2018 suggests that psychometric linkage in this case may be unreliable. A second alternative transformation that could work for the TIMSS-PISA re-scaling would be to regress the TIMSS mean score of linking countries on the PISA mean score, providing aggregate estimates of the linking parameters. This clearly fails for a single linking country and is dubious for very small samples of linking countries. Importantly, it requires significant out-of-sample prediction.

[21] As we are interested in the full distribution of scores, we use all of the plausible values of latent achievement provided in each test. The international assessments provide a series of plausible values for the score of each student to account for the fact that they use matrix testing procedures where each student takes just a subset of the overall assessment item pool.

A variety of alternative approaches for estimation across test regimes has been suggested (e.g., Angrist et al., 2013; Altinok et al., 2014; Patel and Sandefur, 2020; Angrist et al., 2021). These approaches face similar problems to those faced here, but their focus is very different. While entirely concerned with aggregate student outcomes, none uses the core information about the underlying individual skill distribution of students. These approaches to harmonizing the test data necessarily require extrapolating scores far from the observed means in the PISA and TIMSS tests. Appendix A2.1 provides an overview of these alternatives and compares them to our test linking analysis.

2.3.3 Achievement in India and China (Layer 4) on a Global Scale

The two most-populace countries in the world – India and China – have not participated in any of the recent international tests on a national basis, even though sub-national territories within each have participated in PISA. The challenge is going from the regional data to the nation and in so doing developing estimates of national basic skill deficits. Our approach is to combine the sub-national PISA performance information with broader within-country performance information to derive estimates of national achievement distributions expressed on the PISA scale.

Two Indian states (Tamil Nadu and Himachal Pradesh) participated in PISA in 2009. These states on average scored 347.9, ranking them 72 out of 74 countries and regions. In order to adjust these results to reflect the nation, we rely on available independent testing in 2009 for 18 states (out of the 28 states and 8 union territories) including Tamil Nadu and one union territory: Educational Initiatives, 2012 developed common tests in math and language (given in 13 different languages) for grades 4, 6, and 8 and administered these assessments to over 100,000 public-school students in both urban and rural settings. On average, Tamil Nadu students scored 0.019 standard deviations above the national math mean (expressed in standard deviations of the Tamil Nadu student population). We use this adjustment factor to shift the observed PISA distribution for Tamil Nadu to obtain an estimate of the national distribution.

Based on a re-centered distribution of Tamil Nadu scores, we estimate that 85.1 percent of Indian students fall below basic skill levels. While there is some variation in potential estimates from alternative sources, this very large estimated skill deficit proves to be entirely consistent with other ways to judge test performance in India.[22]

[22] Estimating the achievement of two other Indian states – Rajasthan and Orissa – on the international TIMSS scale, Das and Zajonc, 2010 similarly find very low achievement levels. In the Annual Status of Education Report 2018 (ASER, 2019), which assesses students only in rural areas, only 44 percent of the students in the national sample in Standard VIII (14- to 16-year-olds) can perform the most basic task of doing division (i.e., solving a three-digit by one-digit numerical division problem correctly). An additional possible source of information is the Indian Human Development Survey (IHDS) that assesses the math achievement of 8- to 11-year-olds for the first and second child in each household, but it only provides an ordinal variable.

Deriving estimates of skill deficits for China relies on the linking methods of Layers 2 and 3 except linkage comes from the province of Shanghai which has participated in PISA 2012.[23] The China Education Panel Study (CEPS) is a nationally representative survey of seventh and ninth graders in the academic year 2013-2014. The available student-level data contain results from a standardized cognitive ability test that measures students' abilities in language, graphics, calculation, and logic.[24] Importantly, in addition to the nationally representative sample, the CEPS also contains a random sample of students in Shanghai. Focusing on ninth grade where students are approximately 15 years old, the sample encompasses 5,574 students in the nationally representative sample and 587 students in Shanghai. While the cognitive ability test is not the same as the PISA achievement test, under the assumption of similar regional distributions of ability and achievement it provides a way to directly estimate skills across the country. The ability to link the CEPS with PISA provides the means to translate country-wide scores to the PISA scale using our prior transformation method and permits direct estimation of the share of students in China who do not achieve the basic skill level. Only 3.3 percent of Shanghai students fall below the basic skill level in PISA. Results of our linkage analysis suggest that 13.9 percent of the Chinese national student population perform below the basic skill threshold.

2.3.4 Imputation of Achievement in Countries without Test Participation (Layer 5)

The largest data problems come from the 33 countries that never participated in international achievement tests. We impute achievement deficits based on available data on their educational participation (net enrollment) and economic development (GDP per capita) along with the basic skill information for other countries in the same world region and same income group. Imputations introduce unavoidable uncertainty in the estimates of both skills and their impact on growth which we address with sensitivity analyses.[25]

Our imputation comes from estimating the relationship between the proportion of the students below basic skills in country j (ρ^j) for all of the Layer 1-3 countries:

$$\rho^j = \alpha_0 + \alpha_1 E_j^N + \alpha_2 GDP_j + \nu_j + \mu_j + \varepsilon_j \tag{2.3}$$

[23] There is no clear way to compare the results of the four additional wealthy provinces that subsequently also participated in PISA to the rest of China. A number of commentators have concluded that these provinces are not representative of China as a whole (see, for example, Loveless, 2014; Schneider, 2019, and Gruijters, 2020).
[24] The CEPS also contains scores of mid-term exams in Chinese and math collected from transcripts, but these are not standardized or comparable across schools but designed by each school itself.
[25] Using imputed achievement based upon GDP would create obvious circularity in subsequent analyses of the level of GDP, but our economic analysis focuses on growth rates that are not directly affected by such circularity. Our economic projections are based on a growth model that has controlled for the level of GDP.

where E_j^N is net enrollment in secondary school, GDP_j is gross domestic product per capita, ν and μ are indicators for world regions and income groups, respectively, and ε is an error term.[26]

However, net enrollment rates in secondary school are missing for 39 of the Layer 1-3 countries. Therefore, we first impute net enrollment rates based on the more widely available gross enrollment rates, E_j^G, per-capita GDP, and the region and income-group fixed effects. As gross enrollment rates can be greater than 100 percent,[27] we estimate a nonlinear imputation model:

$$E_j^N = \beta_0 + \beta_1 E_j^G + \beta_2 I[E_j^G > 1] + \beta_3 I[E_j^G > 1] * E_j^G + \beta_4 GDP_j + \nu_j + \mu_j + \varepsilon_j \quad (2.4)$$

where $I[E_j^G > 1]$ is an indicator for a gross enrollment greater than 1. This estimation allows for a relationship between net and gross enrollment that is kinked at 100 percent gross enrollment. With an R^2 of 0.955, this prediction model provides a very good fit to the data on net and gross enrollment rates.[28] (Appendix Table A2.2 shows the different imputation regressions).

Based on the estimates from Equation 2.4, we substitute the imputed net enrollment rate E_j^N into the estimation of Equation 2.3. Again, the model fits the data quite well ($R^2 = 0.860$), providing credence to the imputation procedure. We can then impute math values of ρ^j for the Layer 5 countries that have not participated in any of the international assessments using the estimated parameters from Equation 2.4. Missing science values of ρ^j are then imputed by a linear regression of science on math ρ^j ($R^2 = 0.949$).[29][30]

[26] The regional groupings follow World Bank classifications, except that we subdivide the Europe and Central Asia region (where Kazakhstan, the Kyrgyz Republic, the Russian Federation, Tajikistan, Turkmenistan, and Uzbekistan form the Central Asia region and the remaining countries in the World Bank's Europe and Central Asia region form the Europe region).

[27] In contrast to net enrollment rates, gross enrollment rates can exceed 100 percent because of early enrollment or grade repetition so that the school population exceeds the number of children in the grade-appropriate age span.

[28] Data for net and gross enrollment in secondary school come from the World Bank's World Development Indicators (WDI) and refer to the most recent data point in the period 2015-2019 available for each country. Missing values of the imputation variables on the right-hand side of this regression are imputed by multiple imputation.

[29] As SACMEQ and PASEC do not test science achievement, the latter imputation also applies for estimating ρ^j for science in the SACMEQ and PASEC countries.

[30] The imputation regressions are based on the full sample of countries in Layers 1-3. An alternative is to restrict the sample to only low- and middle-income countries which may be more similar to the Layer 5 countries, with the downside that the sample size for the estimation of the imputation parameters is cut into half (50 instead of 100 countries for the math imputation). Results based on imputations that do not use high-income countries turn out to be very similar, with the share of youths below basic skills in Layer 5 countries estimated to be slightly higher than in our baseline model (92.6% instead of 90.3%). This increases the world estimate from 67.2% to 67.9%, indicating that our baseline estimate may slightly underestimate the true extent of skill deficits.

2.3.5 Skill Levels of Children Who Are not in School

The PISA proficiency levels that we use to define mastery of basic skills are set for 15-year-olds, i.e., at the secondary school level. Our analysis of net enrollment rates in secondary school indicates wide variation across countries but that 36 percent of youths globally on average are no longer enrolled in secondary school. Understanding the skill levels of children who drop out of school before the secondary level is not straightforward. Most prior analyses stop at simply counting the numbers of children with low school attainment and do not attempt to go further in assessing their skill levels. Two data sources do nonetheless provide some, albeit imperfect, information about the achievement of out-of-school youth compared to in-school youth.

The PISA for Development assessment includes a unique out-of-school component that tests the achievement of representative samples of 15-year-old children who are no longer in school (OECD, 2020b,c). Because of the low tested achievement levels of the out-of-school children, their achievement is not reported by specific PISA scores but only by categories of proficiency level.[31] In the five countries that participated in the out-of-school assessment (Guatemala, Honduras, Panama, Paraguay, and Senegal), the median achievement of the out-of-school children is 295 PISA points, or over two standard deviations below the OECD mean. This corresponds to the 33^{rd} percentile of the achievement distribution of the youths currently in school in these five countries. With the very low student achievement in these countries, this equates to the 9^{th} percentile of PISA achievement in non-OECD countries. Thus, there is considerable uncertainty when generalizing from these results to other countries.

A second data source provides information on more developed countries (which also have noticeable numbers of youth out of school). The OECD's Survey of Adult Skills, the Programme for the International Assessment of Adult Competencies (PIAAC), contains achievement data for adults in 33 (mostly developed) countries (OECD, 2016). PIAAC samples a representative cross-section of adults, assesses their schooling levels, and gives them a battery of achievement tests. We pool the data across countries and restrict the sample to the age group of 16- to 24-year-olds in order to focus on the current conditions. When we compare the achievement of those who have dropped out of upper secondary school to the achievement of those who did not drop out, we find that dropouts on average across the PIAAC countries achieve at the 14^{th} percentile of the achievement distribution of those completing school.

Although there remains considerable uncertainty, these two sets of calculations give a rough range of the relative achievement of out-of-school youths compared to in-school youths. The data are insufficient, however, to consider country-specific variations in relative skills of dropouts. We therefore follow the assumption in Hanushek and Woessmann, 2015b that youths outside school perform on average at the 25^{th} percentile of those currently in school

[31] The PISA-D test battery includes a greater proportion of items at the low end of the regular PISA test, but it does not include a sufficient number of more fundamental concepts and items that would provide a more complete picture of the distribution of skills for the very bottom category.

in their respective country.[32] Errors in this assumption are not too important for the more developed countries, where the dropout rates are relatively small. For developing countries with larger dropout rates, such errors might be significant when estimating mean achievement. But they are less important in our context of understanding the population that lacks basic skills. The low scores of the in-school population in most of these countries suggests that any reasonable estimate of the skills of dropouts will imply that *nearly all of the dropouts* do not fully attain Level 1 on PISA and thereby lack basic skills.

2.4 Achievement of Basic Skills around the Globe

Our results indicate that the world is far away from ensuring that all children reach at least basic skill levels. In this section, we report our main results on the share of children in the world who fail to reach basic skill levels (section 2.4.1), as well as sensitivity analyses (section 2.4.2).

2.4.1 Main Results

Results suggest that 63.1 percent of the world's in-school children do not reach basic skill levels along with very large differences by country income groups (column 1 of Table 2.2). While the share of students below basic skills is 23.9 percent in high-income countries, it increases to 38.3 percent in upper-middle-income countries, 81.3 percent in lower-middle-income countries, and 90.5 in low-income countries.[33]

The share of students not reaching basic skill levels is highest in Sub-Saharan Africa (89.3 percent) and South Asia (85.0 percent). Yet it also reaches 63.9 percent in the Middle East and North Africa (MENA) region and 61.2 percent in Latin America. In contrast, it is 40.0 percent in Central Asia, 25.9 percent in Europe,[34] 31.1 percent East Asia and Pacific, and 22.2 percent in North America.

While not surprising, the variations across countries underscore the uneven development challenges. In 30 countries, the share of unprepared youth exceeds 90 percent, and in 45 countries it exceeds 80 percent (see Appendix Figure A2.5 and column 2 of Appendix Table A2.4 for country results). In 93 countries, the share of students not reaching basic skill levels is estimated to be more than half. On the other hand, the share is below 20 percent in 25 countries. Just 5.5 percent of students score below the basic skill level in Macao (China), and this share is below 10 percent in two additional countries (Estonia and Singapore).[35]

[32] In our analysis, we assume that the distribution of the achievement of out-of-school youths, centered on the 25^{th} percentile, is normal with a standard deviation equivalent to the in-school youths in the respective country.
[33] When aggregating countries into country groups and world estimates, we weight them by their share in the number of children aged 0-14 years (WDI data for 2019).
[34] The share is 22.5 percent in the subgroup of 27 European Union countries.
[35] While not the focus of our analysis, it is useful to consider the more common assessment of aggregate performance levels of countries. Our scale transformations allow us to express country mean performance on

These results refer only to those children who are currently in school. Importantly, 35.5 percent of children of secondary-school age are no longer enrolled in school globally (column 2 of Table 2.2). This large out-of-school population is well-known and has rightfully been the subject of a wide variety of previous policy initiatives. The school attendance pattern also shows a strong income-group gradient (from 6.9 percent in high-income countries to 69.3 percent in low-income countries). In our baseline estimates, we assume that out-of-school children have a normal distribution with a mean at the 25^{th} percentile of the in-school distribution in the respective country (column 2 of Appendix Table A2.3).

When we include skill deficits of the out-of-school population, we find that two thirds of the world's youth (67.1 percent) are short of reaching basic skill levels (column 3 of Table 2.2). The share is as high as 95.6 percent in the group of low-income countries, but even in high-income countries it reaches one quarter. Across world regions, the share ranges from 23.9 percent in North America and 28.4 percent in Europe (24.3 percent in the European Union subgroup) to 89.2 percent in South Asia and 94.1 percent in Sub-Saharan Africa.

The large international variation becomes very apparent in Figure 3 which puts the country shares of children who do not reach basic skill levels on a world map. In many countries in Sub-Saharan Africa, the share is estimated to be close to 100 percent. There are 36 countries in which more than 90 percent of children do not reach basic skills (see column 6 of Appendix Table A2.4 for details). In 101 countries, the share is estimated to be more than half of children. Two countries have shares below 10 percent, and 19 countries have shares below 20 percent.

2.4.2 Sensitivity Analyses

Our baseline estimates do not directly consider the uncertainty in the identification of low skills. The different sources imply varying confidence in the country-by-country estimates of those below basic skills. The rigorous, scientifically validated, and readily-linked testing regimes for PISA and TIMSS imply high confidence in the estimates for Layers 1 and 2. Unfortunately, these tests have limited penetration into developing countries, including none for low-income countries. The regional tests in Latin America and Africa have rigorous testing regimes and add more developing countries including low-income countries. But this gain is potentially offset by the earlier grade of testing, different timing, and more fragile linking to the Layer 1 testing, implying more uncertainty for Layer 3. The largest uncertainty of the world estimates comes, however, from Layers 4 and 5.

---

the PISA scale. Overall, the estimated achievement of the global student population is 381.7 PISA points, or more than one standard deviation below the OECD mean (column 1 of Appendix Table A2.3). Again, there is a clear gradient across income groups, with low-income countries achieving two standard deviations below the OECD mean. The highest-achieving country is Singapore at 560 points on the PISA scale in math and science on average. In 23 countries, average achievement exceeds 500 PISA points. In another 29 countries, it is estimated to fall below 300 points (details are shown in Appendix Figure A2.6 and provided in column 3 of Appendix Table A2.4).

One simple sensitivity check is to restrict the entire analysis just to those countries with skill information from the upper layers. Thus, column 2 of Table 2.3 restricts the analysis to only the 104 countries in Layers 1 and 2, and column 3 to only the 124 countries in Layers 1, 2, and 3. Note that these analyses are no longer representative of the world as a whole, but only for the countries that participated with nationally representative samples in the respective international tests. Given the selectivity of test participation, this likely leads to a substantial underestimation of the deficit of basic skills at a global scale.

Perhaps surprisingly, just considering PISA and TIMSS participants – i.e., dropping most of the poorly achieving countries in the world – still implies that the share of children achieving below the basic skill level is 56.6 percent (column 2). Adding Layer 3 countries (but still dropping all Layer 4 and 5 countries) increases this share to 62.6 percent (column 3), which is not far below our global estimate. That is, the result that a majority of children worldwide does not reach basic skills is not an artefact of uncertainty in Layers 4 and 5. Importantly, when comparing across the first three columns of Table 2.3, the estimates within each of the world regions – where columns 2 and 3 are based on much fewer observations in the poorer regions but are now based on higher-quality data – in fact remain in the same range. (The only exceptions are the estimated proportions of children without basic skills in the upper-middle-income group and the East Asia & Pacific region that in fact go up because China, which is estimated to have a relatively low share of unprepared students, is no longer considered). This again suggests that the larger estimate of skill deficits at the global level comes from a proper consideration of the status of all children across the different world regions, rather than from uncertainty of the lower-layer estimates.

We highlight uncertainty in India and China (Layer 4) simply because of their size and the fact that they only have partial test information. The available subregional information coupled with other information about the relative positions of the subregions in the country provides solid data for full-country estimates. Still, we probe sensitivity to alternative bounds, but especially for China the available data that allows for bounding is unfortunately quite limited.

For India, we can take the actual estimates of the two states that participated in PISA – Tamil Nadu and Himachal Pradesh – as bounds on the skill deficit in India. This bounds the Indian estimates of those lacking basic skills between 88.6 and 90.1 percent (see Appendix Table A2.5, columns 2 and 3). As such, they have limited impact on the global average, moving it from 67.2 percent at the lower bound to 67.5 percent at the upper bound.

Bounding for China is more problematic because of the substantial economic differences between the eastern provinces (that participated in PISA) and the more rural central and western provinces (that did not).[36] Our baseline approach is to use the distributional linking approach from Layers 2 and 3 to go from Shanghai results to the whole country using the national CEPS data. But the basic-skill threshold is at the far left of the Shanghai distribution,

[36] The rural population of China is still 65 percent of the total population, and even larger in terms of school-age population; see (in Chinese) National Bureau of Statistics of the People's Republic of China, 2010, 2015, 2016.

which introduces some uncertainty, and prior evidence suggests that learning deficits may be much higher in rural areas.[37] We consider two alternative bounds that use the achievement of rural students in two other Southeast Asian countries – high-achieving Vietnam and low-achieving Cambodia – for Chinese students who live in rural areas (while using our baseline China estimate only for urban students).[38] The bounding estimates suggest a share of children in China below basic skills of 19.0 percent based on rural Vietnam (close to our baseline estimate of 18.0 percent)[39] and of 68.9 percent based on rural Cambodia (see Appendix Table A2.5, columns 4 and 5). This bounding is wide and thus not very informative, and given the size of China, the alternatives would bound the global estimate of skill deficits between 67.6 percent and 74.1 percent.

Finally, while the sensitivity analyses so far are based on the observed achievement of students, the consideration of out-of-school children introduces additional uncertainty in our analysis. The baseline analysis assumes that out-of-school children on average perform at the 25^{th} percentile of the distribution of in-school children in their country. To see how sensitive the estimates are to alternative assumptions, we perform calculations that assume that out-of-school children instead perform at the 15^{th} and 35^{th} percentiles, respectively, of the in-school distribution (columns 4 and 5 of Table 2.3). It turns out that the world estimates of the share of children falling below basic skill levels are not very sensitive to these alternative assumptions, ranging from 65.9 percent to 68.5 percent.

2.5 The Economic Gains from Global Universal Basic Skills

The overall motivation for this analysis is understanding how global development could be altered by improved schooling policies that aided those currently without internationally competitive skills. There is little doubt that increasing the quantity and quality of education in

[37] A few previous studies have provided data for rural students that indicate sizable discrepancies between urban and rural areas. For example, Wang et al., 2018 use two waves (2010 and 2014) of the China Family Panel Survey to consider math and Chinese-language tests across a national sample. They conclude that rural 10-to-15-year-olds are at least two years behind the urban students (roughly equivalent to one half to two-thirds of a standard deviation). For younger children, Emmers et al., 2021 find dramatically increased risk of cognitive, language, and social-emotional delays for rural children, and Zhao et al., 2019 find that a third of rural children have IQ scores one standard deviation or more below the international mean. Gao et al., 2021 compares reading achievement of fourth-grade students in three Western provinces to other countries of the world and places the three provinces last among 44 countries participating in the international PIRLS test.

[38] We derive estimates of rural performance by considering only students going to school located in communities with less than 15,000 people, as reported by school principals in the PISA background questionnaire. This yields a share of rural students achieving below the basic skill level of 15.6 percent in Vietnam and of 95.8 percent in Cambodia. Note, however, that this relatively high in-school performance in Vietnam comes from a very selected sample, as almost half of the secondary-school-age children are not in school.

[39] In an alternative conversion analysis based on the China Family Panel Studies (CFPS), we obtain an even lower estimate, but the conversion there is highly unreliable because the CFPS test has very little support at the level of the basic-skill threshold (see Appendix A2.2).

a country would improve the economic outcomes for the affected youth. But our motivation is more the impact on the aggregate economic outcomes of countries – which we see as the engine for addressing the broad development goals identified in the SDGs. Here we provide direct estimates of the country-by-country economic gains that accrue from moving to universal basic skills. For policy purposes, these benefit estimates would have to be set against the cost of any policy achieving the goal, but that unfortunately is currently subject to considerable uncertainty. We describe the underlying framework (section 2.5.1), the policy reform scenarios (section 2.5.2), and the simulation model (section 2.5.3), followed by baseline results (section 2.5.4) and sensitivity analyses (section 2.5.5).

2.5.1 Skills and Growth

The projections build on prior work of how the skills of the population relate to economic growth (Hanushek and Woessmann, 2015a). A key element of the projections is that neither educational improvement nor economic growth occurs instantaneously. We take pains to include the time path of improvement, recognizing that school reform takes time and that transforming a country's entire labor force takes even longer.

We project the economic gains that individual countries and the world could reap if they focused on the improvement of basic skills. The projections rely on the empirical growth models estimated in Hanushek and Woessmann, 2012a that were developed in the spirit of endogenous growth models.[40] The "knowledge capital" of nations emphasized there is measured by international tests of student achievement in math and science expressed on the same PISA scale that we use above. Furthermore, that study documents a series of econometric analyses that are consistent with an interpretation of the estimated growth coefficients as a causal effect of skills on growth. The estimates suggest that a one standard deviation increase in test scores (i.e., 100 score points on the PISA scale) is associated with an increase in the average annual growth rate in real GDP per capita by 0.0198 in the long run.[41]

Growth projections are, of course, subject to considerable uncertainty, particularly at the low end of current skills. The estimates of long-run growth implicitly assume that modern industry develops within each country over time as the skills of the population improve. This assumption matches what has been seen in the past, with the East Asian experiences in South Korea, urban China, and other places being prime examples. But the development impact

[40] Alternative estimates based on an augmented neoclassical growth model have been provided in Hanushek and Woessmann, 2011, 2015b and show roughly one-fifth lower long-run economic impact.

[41] In the empirical growth model, we also experimented with specifications that consider the ends of the skill distribution (e.g., population shares reaching basic and very high skill levels), but these growth models yield relatively imprecise estimates because the ends of the distribution are thin in many countries and because there is not enough variation across countries in the specific shape of the distribution (Hanushek and Woessmann, 2015a). Thus, most of the observed variation in the low- and high-performing shares is joint rather than separate. While the point estimates provide similar results, we revert to the growth model estimated in mean achievement which provides more precise estimates.

of skill improvement can actually be seen at earlier points when better education improved the performance of small farmers in low-income countries. The seminal paper by Welch, 1970 on the value of education for decision-making under uncertainty tested the underlying economic hypotheses using data on U.S. farmers. The importance of cognitive skills for agricultural efficiency in low-income countries is found in several analyses of Asian agricultural development in the early periods of economic development (Jamison and Lau, 1982; Jamison and Moock, 1984). It is developed and tested rigorously in Foster and Rosenzweig, 1996.

There is also concern that recent changes in modern economies, often categorized by a move toward artificial intelligence, might change the historic skills-growth relationship. While occupational patterns have changed significantly in the U.S. (Autor, 2019), there is no evidence of a decline in skilled employment(Acemoglu et al., 2022). A portion of this literature discusses the disappearance of routine tasks (e.g., Autor et al., 2003). The focus on the task content of different occupations should not be confused with our definition of basic skills that included the ability to “carry out routine procedures.” The basic skills identified in the Level 1 performance in PISA are required in a broad set of occupations and are the building block for more advanced skills, making it unlikely that their returns will fall dramatically with current technological developments.[42]

2.5.2 Three Reform Scenarios

The policy scenarios that we consider have dual objectives. They lift the skills of those currently left behind, thus dealing directly with more income-equalizing objectives, while they add to the country’s knowledge capital, thus dealing with overall economic development objectives. The reform scenarios are assumed to follow a linear improvement path taking R years to be completely accomplished ($R = 15$ in the baseline model). This implies that the education levels of each of the first R cohorts of students following the initiation of reform will have different (and improving) skill levels. To describe the three alternative scenarios, it is useful to look at the achievement level ($\overline{A_\tau}$) of each new cohort during the reform period:

$$\overline{A_\tau} = (1 - \theta_\tau)A_\tau^S + \theta_\tau A_\tau^{NS} \text{ for } \tau = 1, 2, ...R \tag{2.5}$$

where A_τ^S and A_τ^{NS} are the achievement of youth in school and not in school, respectively, and θ_τ is the proportion of youth not in secondary school.

Scenario I: Current students achieve at least basic skills. In Scenario I, all children who are currently in school reach at least basic skills. That is, all students (ρ_0) who currently perform below PISA Level 1 are lifted to the Level 1 threshold. By contrast, the achievement of those students who are already above the threshold does not change; neither does the

[42] Basic skills are regularly shown to be the foundation for further human capital development (e.g., Hoyos et al., 2021). Moreover, it is often suggested that improving human capital is particularly important for developing countries in the face of rapid technological change (e.g., World Bank, 2019).

achievement of out-of-school children change. This is conceptually a school reform that implements a minimum quality standard in all schools.[43]

To calculate how much this scenario would change countries' average achievement, we use the achievement micro data. Replacing the achievement of each student who scores below the Level 1 threshold by achievement at Level 1 (denoted A^*), we can calculate each country's average student achievement after this reform. In this case, the path of achievement for the in-school population is given by:

$$A_\tau^S = (1 - \rho_0)A_0^S + (A^* - A_\rho^S)\frac{\tau}{R}\rho_0 \tag{2.6}$$

where A_0^S is the average achievement of students initially above Level 1 and A_ρ^S is the average achievement of those initially below Level 1.[44] Over the course of the reform period, R, the skills of all in-school youth are brought to the minimal skill level (A^*), but there is no change in the skills of the out-of-school population. The aggregate skills of each cohort are thus just the weighted sum indicated in Equation 2.5 with the in-school component given by Equation 2.6.

Scenario II: Full participation at current quality. Scenario II focuses on the achievement of those children who are currently out of school. The average achievement of out-of-school children is lifted to the average achievement of in-school children in the respective country. That is, by the end of the reform period, the country achieves full school participation at current quality levels, leaving all students who are already enrolled unaffected:

$$\overline{A_\tau} = (1 - \theta_\tau)A_0^S + \theta_\tau A_0^{NS} \text{ where } \theta_\tau = \theta_0(1 - \frac{\tau}{R}) \tag{2.7}$$

In a sense, Scenario II applies the opposite approach from Scenario I, extending access to schools without changing their quality. In our setting, this scenario amounts to lifting the average achievement of out-of-school children from the 25^{th} percentile to the mean of the respective country distribution.

Scenario III: All children achieve at least basic skills. Scenario III is a combination of the previous two scenarios where all children achieve at least basic skills. That is, there is full participation in secondary school with every student attaining at least the basic skill level. The achievement of each new cohort of entrants over the reform period is given by the sum of two components. The first component is the Scenario I improvement, which is weighted by the share of in-school children. The second component is the difference between Level

[43] These estimates are best thought of as a lower bound on improvements from any actual school reform. It is difficult to consider such a precisely targeted reform that does not also lift the achievement of additional students above the basic skills threshold.

[44] When $A_0^S < A^*$, the first term substitutes the average achievement of those initially above A^*.

1 achievement (A^*) and the average achievement of those out-of-school children who fall below Level 1. Note that the improvement for out-of-school youths here is different from Scenario II, as they improve to the basic-skill level rather than to the mean level of current students in the country. These reform scenarios anticipate improvements in the new cohorts up to the end of the reform period. After that, future cohorts would continue with the final level of skills, $\overline{A_R}$.

2.5.3 The Simulation Model

The skills of each cohort are of course not the same as the skills of the workforce in the country. The workforce begins with people educated from before the reform period. They will ultimately be replaced by more skilled people through retirement of the existing workers, but that replacement continues past the period of school reform. Thus, for example, if working life is assumed to be 40 years, each cohort of new, higher-achieving students is only a fraction of the total labor force, i.e., 2.5 percent each year.[45]

We calculate the skills of the workforce each year by replacing the oldest workers with the skills of each new cohort (i.e., $\overline{A_\tau}$) weighted as $1/W$, where W is the length of work life. In calculating the knowledge capital of the reforming country, we consider four separate phases:

1. *School reform* ($\tau = 1, \ldots, R$): During the reform period R, workers with the initial skill level in the economy are being replaced by progressively more skilled workers.

2. *Main replacement* ($\tau = R+1, \ldots, W$): Workers of the original skill level will be replaced by the new higher-skilled workers for the next ($W - R$) years.

3. *Quality consolidation* ($\tau = W + 1, \ldots, W + R$): For the next R years, some of the variable quality workers educated during the reform period are replaced by the higher-skill workers.

4. *Completely higher skilled* ($\tau = W + R + 1, \ldots$): The workforce is constant at higher skills.

To estimate the economic effects of this upskilling of the labor force, we use the estimated impact of aggregate skills, or knowledge capital, on growth rates (γ) found in Hanushek and Woessmann, 2012a. We assume that GDP without the reform grows at a constant rate of potential GDP, i.e., $g_\tau^{no\ reform} = p$. For each year of the simulations, we calculate the growth of GDP with the reform as:

$$g_\tau^{reform} = p + \gamma \overline{A_\tau} \tag{2.8}$$

[45] This is a simplification that permits consistent estimates across countries, but the reality in each country will depend both on the labor force and retirement institutions as well as on the population pyramid.

GDP without and with the reform then evolves as:

$$GDP_{\tau}^{\Delta} = (1 + g_{\tau-1}^{\Delta})GDP_{\tau-1}^{\Delta} \text{ where } \Delta \in (\text{reform, no reform}) \tag{2.9}$$

The total value V of the reform is given by the sum of the discounted values of the annual differences between the GDP with reform and the GDP without reform:

$$V = \sum_{\tau=1}^{S}(GDP_{\tau}^{reform} - GDP_{\tau}^{no\ reform}) * (1 + d)^{-\tau} \tag{2.10}$$

where S is the end of the simulation period and d is the discount rate.

Importantly, these simulations assume that all countries can develop simultaneously. They also assume that the economies of developing countries will evolve with the improvement of schools so that they effectively use the higher quality labor force.

The parameters for the baseline version of our simulation model are given in Table 2.4. In the simulations, we consider future returns over an 80-year period (S), roughly until the end of the century. In developed countries, this time horizon is roughly equivalent to the expected lifetime of a child born at the beginning of the reform. The discount rate in the baseline model is 3 percent.[46] The status quo growth rate of 1.5 percent reflects movement of the global production frontier as seen in the long-run growth rates for the OECD.[47] The starting value of GDP for each country is taken from the World Development Indicators (WDI) of the World Bank for 2019 (purchasing power parity (PPP), current prices).

2.5.4 Baseline Results

The results of our simulation model suggest that reaching the development goal of global universal basic skills would lead to very large economic gains. Further, the largest gains come from addressing school quality issues. The net present value of reform Scenario I, where all current students achieve at least basic skills, amounts to $364 trillion of added world GDP over the remainder of the century (Table 2.5). This is equivalent to 2.7 times current annual world GDP, or 5.8 percent of the discounted future GDP stream over the same period. At the end of the projection period in 2100, world GDP would be 23.8 percent higher than without the reform.

The value of Scenario II – full school participation at current quality levels – is about half the value of Scenario I. It amounts to $175 trillion, or 2.8 percent of discounted future GDP. This

[46] This is a standard value of the social discount rate used in long-term projections (e.g., Börsch-Supan, 2000). Deriving a practical value of the social discount rate in cost-benefit analysis of intergenerational projects from an optimal growth model, Moore et al., 2004 suggest a discount rate of 3.5 (2.5) percent for the first (next) 50 years.

[47] This rate would correspond to the steady-state growth rate in many macroeconomic models. Clearly the short-run growth of many emerging countries, as seen in India and China, is much higher, reflecting the combination of steady-state growth and catch-up growth that comes from moving toward the frontier.

is the case even though over one-third of the world's youth are not completing secondary school.

The big gain comes from the combination of the two, where both in-school and out-of-school children are lifted at least to basic skill levels. Fully achieving global universal basic skills in Scenario III would raise future world GDP by $732 trillion – over 5.4 times current annual GDP, or 11.6 percent of discounted future GDP. By the end of the century, global GDP would be 56.0 percent higher than under status quo trajectories.[48]

Table 2.6 breaks these estimates down by world regions (see Appendix Table A2.6 for results by country). On average, over two-thirds of the youth in low-income countries do not complete secondary schooling. As a result, in low-income countries the value of Scenario II is nearly as large as the value of Scenario I, although even there, quality improvements of schools for current students reap higher value than expansion at current quality levels. The importance of improved school quality is nonetheless overwhelming as Scenario III – which puts the out-of-school children into schools that provide basic skills – has a present value that is 35 times current GDP for these low-income countries.

Interestingly, the largest economic gains (in absolute terms) come from the lower-middle-income country group, partly because of its size. Over half of the improved world GDP from universal basic skills accrues to the 41 lower-middle-income countries. Across the world regions, the largest absolute gains accrue in South Asia and Sub-Saharan Africa.

2.5.5 Sensitivity Analyses

To provide a sense of the sensitivity of the simulation results, we can bound the various parameters of the underlying simulation model.[49] Bounding the reform duration between 10 and 20 years, the working life between 35 and 45 years, and the growth coefficient between plus and minus one standard error all leave the projected value of Scenario III between $640 and $850 trillion (see Appendix Table A2.7 for details). Because of the time-delayed impact of reform on growth, the one parameter that makes the biggest difference for the results is the discount rate: With a higher discount rate of 4 percent, the reform value is $417 trillion, whereas it is $1,323 trillion with a lower discount rate of 2 percent.

The discussion in section 2.3 highlighted some of the uncertainty in the estimates of skill deficits. There is particular uncertainty about the achievement of those children not attending school, who are assumed to achieve at the 25^{th} percentile of the respective country distribution

[48] If, in countries whose current mean student achievement is above the basic-skill threshold, out-of-school youths are assumed to improve to the mean of the country's in-school students (as in Scenario II) rather than to the basic-skill level, the increase in future world GDP grows to $771 trillion, or 572 percent of current annual GDP (this is the Scenario III reported in Hanushek and Woessmann, 2015b).

[49] Appendix A2.3 provides additional details on the various sensitivity analyses.

of in-school students. If we alternatively assume performance at the 35^{th} or 15^{th} percentile, the reform value is $657 or $838 trillion, respectively (see Appendix Table A2.8).

We can bound the uncertainty in the skill estimates of current students by assuming that their achievement increase is 10 percent lower or higher, respectively, than in our baseline model. With these bounds, estimated reform values range from $681 to $785 trillion. To jointly consider uncertainty in the estimates of in-school and out-of-school children (as well as in the enrollment measures), a similar ±10 percent bound on the achievement gain for all children provides a range of economic gains from $636 to $836 trillion.

Inherently, however, the uncertainty is not evenly distributed but increases across the five layers of decreasing reliability of test information. We therefore allow the breadth of the bounds of the estimates to increase with the layers, assuming ±5 percent of the baseline achievement increase for Layer 1, ±10 percent for Layer 2, ±15 percent for Layer 3, ±20 percent for Layer 4, and ±25 percent for Layer 5. With these bounds, the estimates of the world reform value range from $604 to $884 trillion. However, this analysis assumes systematic errors for all countries at the lower and upper side, respectively, at the same time. In the more likely case where errors are random within each layer, in expectation the reform value would be equivalent to our baseline estimate, as country errors on either side cancel out in the world estimate.

Overall, the sensitivity analyses (except for the obvious relevance of the choice of discount rates in long-term projections) indicate that the economic gains from achieving universal basic skills may only be 10 percent of discounted world GDP, instead of the 12 percent in our baseline.

2.6 Conclusions

All member states of the United Nations endorsed the Sustainable Development Goals in 2015. An essential element of these 17 goals was the call to ensure inclusive and equitable quality education for all. Because of the fundamental importance of education for economic development and, by implication, for meeting the other 16 SDGs, education is actually the cornerstone to the entire effort.[50] Yet our results suggest that the world is incredibly short of meeting the goal of universal quality education, and this leaves many in the world short of the basic skills needed to participate in modern economies.

The PISA and TIMSS tests provide a starting point for identifying the world distribution of skills, but only few of the poor countries in the world choose to participate in these tests. If we expand coverage to a global scale, including the addition of regional test information, it becomes evident that a majority of children in the world does not reach basic skill levels.

[50] Education can also improve sustainability by enhanced adaptive capacity to climate change, changed environmental behavior, and facilitated adoption of clean technologies (Lutz et al., 2014).

The disparities in skills are profound. According to our estimates, at least two-thirds of the world's youth have skill levels below the basic competitive level. The largest shares of children who do not reach basic skills are in Sub-Saharan Africa (94 percent), South Asia (89 percent), the MENA region (68 percent), and Latin America (65 percent). But even in North America and Europe, about a quarter of youths do not reach basic skill levels. The skewed international distribution is quite evident at the country level: as many as 36 countries have more than 90 percent of their children not reaching basic skills, standing in sharp contrast to the 19 countries that have shares below 20 percent.[51]

The developing world faces the dual problem of access to and quality of schools.[52] Over one third (36 percent) of the global youth of secondary-school age do not attend school. Still, even among enrolled students, 63 percent of the world's students do not reach basic skills. These findings suggest that attendance at low-quality schools will not solve the problem of missing basic skills. Solving the school quality problem, of course, is not simple. The more developed countries have generally resolved school attainment issues, but they have not completely solved the quality challenges as significant shares of their students are still left behind.

The vast education deficits in the global South are made even more important by changes in the global economy: with integrated economies, people are no longer just competing with workers in adjoining cities or states, as most products can be produced anyplace in the world. A larger question is the potentially limited value of increased skills in economies that are dominated by subsistence agriculture, limited manufacturing production, and generally undeveloped markets. But the history of development in East and South Asia offers an indication of the development path associated with increasing skills. First, past research has shown that farming, even at a low level, can benefit as more educated farmers make better crop and planting decisions. Second, on a broader scale, economies have transformed through production in manufacturing with increasing value-added sectors and through the movement toward more information-based activities. Thus, while not certain, it seems natural to conclude that industry develops in ways that match the available skills of the potential labor force.

This means that the large shares of the world's children that do not reach basic skill levels have immense consequences for global economic development. According to our projections based on historical patterns of long-run growth, the world would gain $732 trillion in added GDP over the remaining century if it were to reach global universal basic skills. This is equivalent to

[51] These estimates likely underestimate concerns about the future because of differential population growth. According to UN population projections, Sub-Saharan Africa is expected to account for more than half of the world's population growth until 2050. By that time, the share of the world's children (aged below 18) who live in Sub-Saharan Africa will have grown from 23.7 percent to 35.5 percent (United Nations, 2022).

[52] Obviously, the educational outcomes observed around the world are not solely the result of school quality. While school quality undoubtedly contributes to the existing deficits, many other family and societal factors also contribute (e.g., Woessmann, 2016b). We emphasize school quality because changing it is the most direct way in which public policy can reduce the learning deficits.

over five times current annual world GDP and to 12 percent of discounted future GDP over the remainder of the century. Perhaps more relevantly, the total Official Development Assistance – most of which does not go toward skill development – was just $161 billion in 2020.

Our analysis provides a global picture of the distribution of basic skills around the world, but it comes with uncertainty, particularly for the large part of the world that does not regularly participate in international testing. The neediest countries in the world do not routinely participate in either national or international tests. As a result, they have no information about the current level of skill development (as seen from the vantage point of the international economy). Nor do they have information about whether their schools are improving or not as measured in terms of international skill levels. Echoing the conclusions by the World Bank, 2018, it would be a great service to world development if there were a regular, internationally standardized test of representative samples of students in all countries of the global South. Just like what PISA has done for richer countries, such a globally comparative test would provide policy makers with much better information to focus their energies and to devise suitable policies. Ideally, the test would be both linked to the PISA scale and geared towards measuring basic levels, so that the tested content is relevant in countries that struggle to reach international levels.[53] Developing and funding assessment instruments benchmarked to international educational standards are likely to have much more long-run payoff than much of the current development aid.

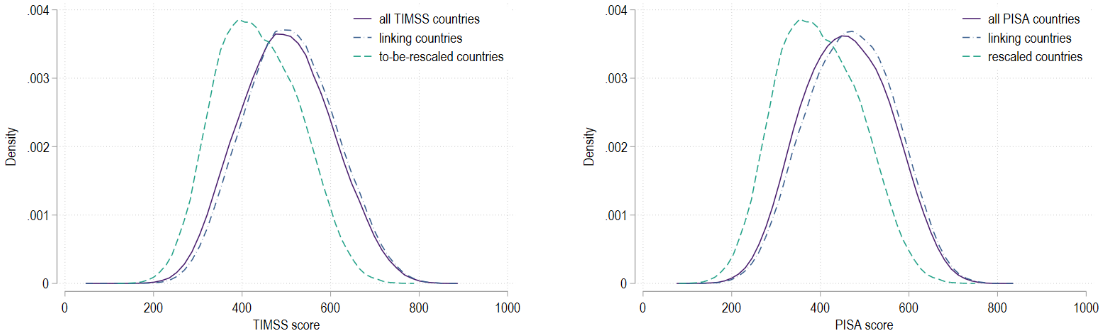
Finally, the previous picture considers just the pre-pandemic world. The pandemic has significantly changed the educational outcomes of the current cohorts of students. Their losses as a result of school closures and reluctance to return to the classroom will not disappear by restoring schools to their January 2020 performance (Hanushek and Woessmann, 2020b).

Even worse, there is mounting evidence that the learning losses from the pandemic have been disproportionately severe for poor children – both those in developed economies and those more generally in developing economies. Not only were schools generally closed for longer periods in developing economies but options to replace traditional in-person classes were also more limited. The need to recover from the setbacks of the pandemic places extra demands on the reform mandates described here.

[53] When implemented with adaptive testing methods, the questionnaire items can be chosen to be relevant and meaningful for each participating child. Thus, a test can realistically cover a wide range of performance levels.

Figures and Tables

Figure 2.1 : Conversion of TIMSS achievement onto the PISA scale

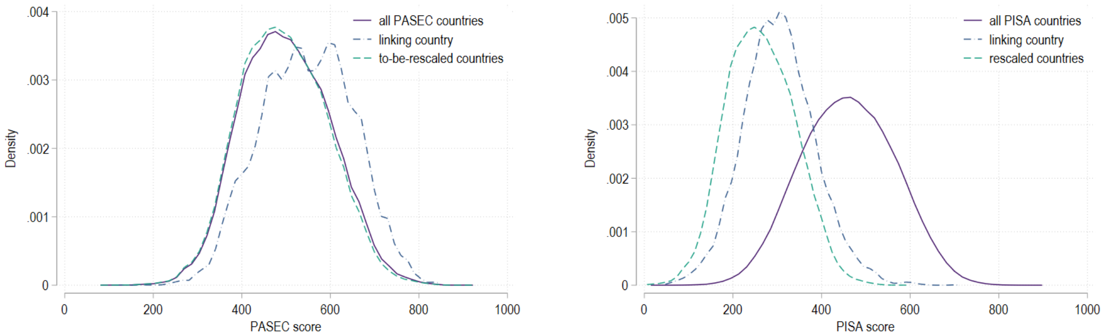


(a) Achievement on TIMSS scale

(b) TIMSS achievement transformed to PISA scale

Notes: Gaussian kernel densities, bandwidth 10.

Figure 2.2 : Conversion of PASEC achievement onto the PISA scale

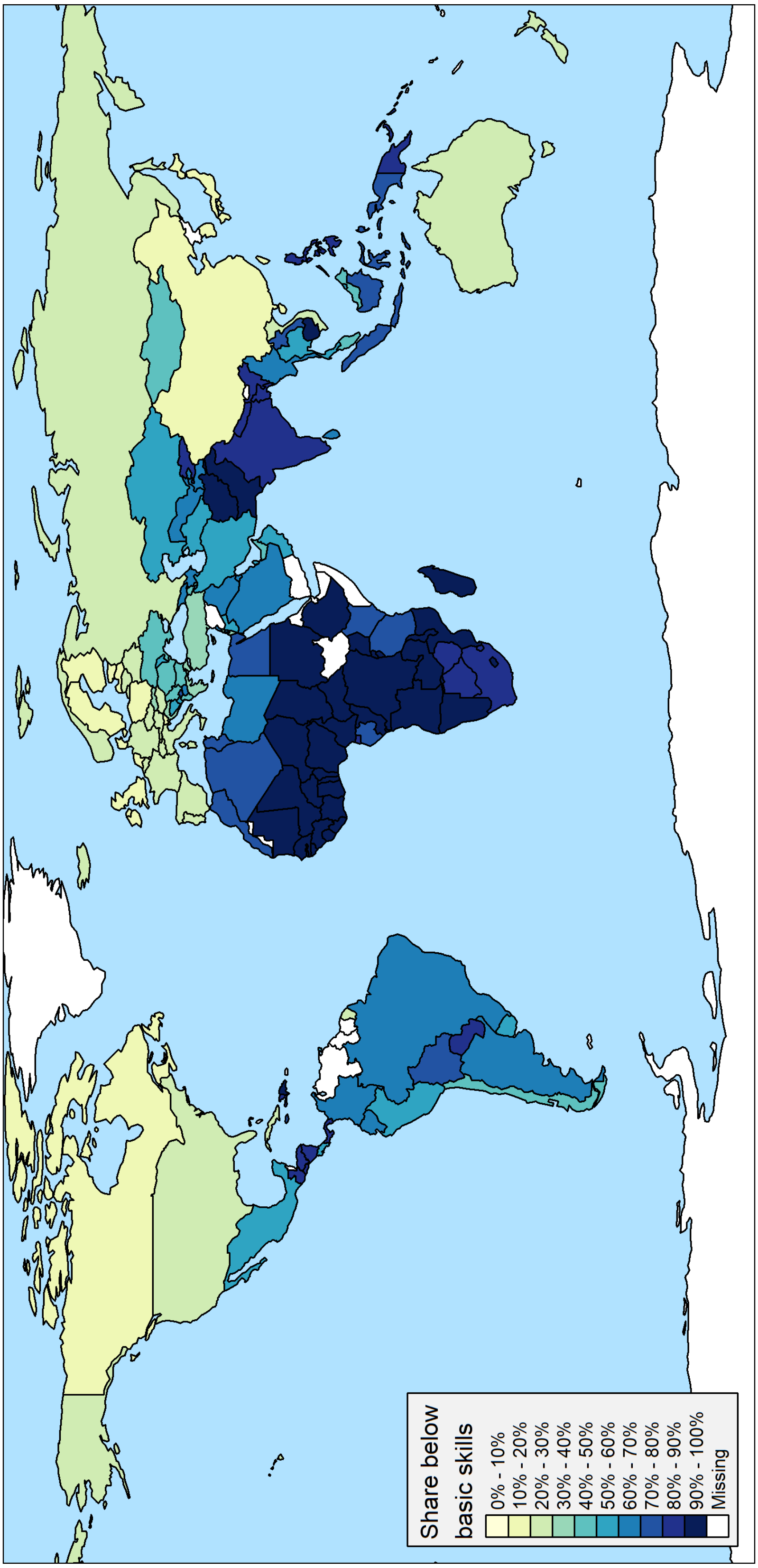


(a) Achievement on PASEC scale

(b) PASEC achievement transformed to PISA scale

Notes: Gaussian kernel densities, bandwidth 10.

Figure 2.3 : World map of lack of basic skills: Share of children who do not reach basic skill levels



Notes: Estimated share of children (incl. those currently out of school) who do not reach at least basic skill levels in math and science (equivalent to PISA Level 1). See section 2.3 for methodological details.

Table 2.1 : Available skill data at different layers of certainty

	Number of countries					Share of world population		Share of world GDP	
	Total	By income group							
Layers	(1)	Low (2)	Lower-middle (3)	Upper-middle (4)	High (5)	Percent (6)	Cumulative (7)	Percent (8)	Cumulative (9)
1. PISA participants	**90**	**0**	**12**	**28**	**50**	**0.369**	**0.369**	**0.665**	**0.665**
1a) PISA 2018	75	0	4	24	47	0.333	0.333	0.647	0.647
1b) Previous PISA rounds	8	0	4	1	3	0.022	0.356	0.013	0.66
1c) PISA for Development	7	0	4	3	0	0.013	0.369	0.005	0.665
2. TIMSS participants	**14**	**0**	**6**	**4**	**4**	**0.068**	**0.437**	**0.038**	**0.702**
2a) TIMSS 2019	7	0	1	2	4	0.033	0.402	0.027	0.692
2b) Previous TIMSS rounds	6	0	4	2	0	0.007	0.408	0.003	0.695
2c) TIMSS Grade 4	1	0	1	0	0	0.028	0.437	0.008	0.702
3. Participants in regional tests	**20**	**8**	**10**	**2**	**0**	**0.051**	**0.487**	**0.010**	**0.713**
3a) TERCE/SERCE	2	0	1	1	0	0.002	0.439	0.002	0.705
3b) SACMEQ	9	3	5	1	0	0.029	0.468	0.005	0.709
3c) PASEC	9	5	4	0	0	0.019	0.487	0.003	0.713
4. Sub-territorial PISA participation	**2**	**0**	**1**	**1**	**0**	**0.361**	**0.848**	**0.245**	**0.958**
4a) India	1	0	1	0	0	0.178	0.665	0.071	0.784
4b) China	1	0	0	1	0	0.183	0.848	0.174	0.958
5. No international participation	**33**	**15**	**12**	**6**	**0**	**0.133**	**0.981**	**0.036**	**0.994**
Total	159	23	41	41	54	0.981		0.994	

Notes: Col. 1-5: Number of countries falling in the respective category of data availability on student achievement. Col. 2-5: Country grouping follows World Bank classification. Col. 6: Share of world population. Col. 7: Cumulative share of col. 6. Col. 8: Share of world GDP. Col. 9: Cumulative share of col. 8. Country sample: All 159 countries with a population of at least one million or a GDP that is at least 0.01 percent of world GDP.

Table 2.2 : Basic skill deficits on a global scale

	Share of students below basic skills (1)	Share of children not enrolled in secondary school (2)	Share of all children below basic skills (3)
World	0.631	0.355	0.672
By income group			
Low-income countries	0.905	0.693	0.956
Lower-middle-income countries	0.813	0.440	0.858
Upper-middle-income countries	0.383	0.189	0.423
High-income countries	0.239	0.069	0.255
By region			
Sub-Saharan Africa	0.893	0.665	0.941
South Asia	0.850	0.402	0.892
Middle East & North Africa	0.639	0.195	0.679
Latin America & Caribbean	0.612	0.210	0.652
Central Asia	0.400	0.094	0.421
East Asia & Pacific	0.311	0.219	0.354
Europe	0.259	0.102	0.284
North America	0.222	0.069	0.239

Note: Col. 1: Estimated share of current students who do not reach at least basic skill levels in math and science (equivalent to PISA Level 1). Col. 2: One minus net secondary enrollment rate (from WDI and own imputations). Col. 3: Estimated share of children (incl. those currently out of school) who do not reach at least basic skill levels in math and science. See section 2.3 for methodological details. Country groups follow World Bank classification.

Table 2.3 : Sensitivity of skill estimates: Restriction to higher layers of reliability and bounding of out-of-school children

	Baseline	Only countries in		Assumption on out-of-school children	
		Layers 1 and 2	Layers 1, 2, and 3	35th percentile	15th percentile
	(1)	(2)	(3)	(4)	(5)
World	0.672	0.566	0.626	0.659	0.685
By income group					
Low-income countries	0.956	n.a.	0.967	0.947	0.963
Lower-middle-income countries	0.858	0.792	0.810	0.844	0.871
Upper-middle-income countries	0.423	0.596	0.595	0.408	0.441
High-income countries	0.255	0.255	0.255	0.247	0.264
By region					
Sub-Saharan Africa	0.941	0.906	0.911	0.929	0.952
South Asia	0.892	0.952	0.952	0.880	0.900
Middle East & North Africa	0.679	0.677	0.677	0.665	0.692
Latin America & Caribbean	0.652	0.647	0.645	0.638	0.666
Central Asia	0.421	0.325	0.325	0.413	0.431
East Asia & Pacific	0.354	0.547	0.547	0.337	0.376
Europe	0.284	0.284	0.284	0.272	0.298
North America	0.239	0.239	0.239	0.230	0.249

Note: Estimated share of children (incl. those currently out of school) who do not reach at least basic skill levels in math and science (equivalent to PISA Level 1). Col. 1: baseline results (see col. 3 of Table 2.2). Col. 2 and 3: sample restricted to only countries in Layers 1 and 2 and to only countries in Layers 1, 2, and 3, respectively. Col. 4 and 5: assume that out-of-school children on average achieve at the 35th and 15th percentile, respectively, of the in-school children in the respective country. See section 2.3 for methodological details. Country groups follow World Bank classification.

Table 2.4 : Parameters of the simulation model

Parameter	Definition	Baseline value
R	Reform period (years)	15
W	Length of work life	40
S	Simulation period (years)	80
d	Discount rate	0.03
p	Status quo growth rate	0.015
γ	Growth coefficient	0.0198
A*	Math basic skills (Level 1)	420
	Science basic skills (Level 1)	410

Note: Growth coefficient: Additional annual growth for a one standard deviation increase in test scores. See section 2.5.3 for details.

Table 2.5 : World estimates of economic gains from achieving global universal basic skills

	Scenario I: Current students achieve at least basic skills (1)	Scenario II: Full participation at current quality (2)	Scenario III: All children achieve at least basic skills (3)
Value of reform (bn USD)	363,567	175,348	732,127
In % of current GDP	270%	130%	543%
In % of discounted future GDP	5.80%	2.80%	11.60%
GDP increase in year 2100	23.80%	10.60%	56.00%

Note: Discounted value of future increases in GDP until 2100 due to the reform scenario, expressed in billion USD, as a percentage of current GDP, and as a percentage of discounted future GDP. "GDP increase in year 2100" indicates by how much GDP in 2100 is higher due to the reform (in percent). Basic skills: Achieving at least at the equivalent of PISA Level 1. See section 2.5.2 for details on the reform scenarios and section 2.5.3 for details on the simulation model.

Table 2.6 : Economic gains from achieving universal basic skills: By country groups

	Scenario I: Current students achieve at least basic skills		Scenario II: Full participation at current quality		Scenario III: All children achieve at least basic skills	
	Value of reform (bn USD) (1)	In % of current GDP (2)	Value of reform (bn USD) (3)	In % of current GDP (4)	Value of reform (bn USD) (5)	In % current GDP (6)
World	363,567	270%	175,348	130%	732,127	543%
By income group						
Low-income countries	6,598	554%	5,652	475%	41,349	3475%
Lower-middle-income countries	142,345	715%	58,416	294%	383,560	1928%
Upper-middle-income countries	112,519	229%	73,167	149%	182,249	371%
High-income countries	102,105	160%	38,113	60%	124,969	196%
By region						
Sub-Saharan Africa	28,378	642%	18,681	422%	125,798	2844%
South Asia	97,894	821%	35,084	294%	259,524	2176%
Middle East & North Africa	47,724	634%	8,743	116%	66,681	886%
Latin America & Caribbean	49,635	491%	14,126	140%	76,926	761%
Central Asia	10,848	204%	3,355	63%	13,241	248%
East Asia & Pacific	51,965	121%	60,939	142%	92,672	216%
Europe	42,946	151%	20,555	72%	55,498	195%
North America	34,177	147%	13,865	60%	41,787	179%

Note: Discounted value of future increases in GDP until 2100 due to the reform scenario, expressed in billion USD and as a percentage of current GDP. Basic skills: Achieving at least at the equivalent of PISA Level 1. See section 2.5.2 for details on the reform scenarios and section 2.5.3 for details on the simulation model. Country groups follow World Bank classification.

Appendix

Appendix A2.1 Comparison to other Data Linkages

Ours is not the only attempt to combine information from different international student achievement tests. In particular, two important recent contributions to this literature – Angrist et al., 2021 and Patel and Sandefur, 2020 – use alternative methods to link some of the same underlying tests. Importantly, the focus of our study is somewhat different from these studies, as our main goal is to estimate the share of children who do not reach basic skills in each country and in the world rather than to estimate countries' mean student achievement. Nevertheless, because we also estimate country mean scores, it is revealing to describe how our method relates to these two papers and to indicate in which dimensions results differ or are similar.

For pairs of international tests that have more than one country participating in both tests, Angrist et al., 2021 use aggregate country data rather than the underlying micro data. They estimate a regression of country mean scores in TIMSS or PIRLS on the countries' mean score in any other test. They then predict achievement of the remaining countries in the other test to the TIMSS/PIRLS scale using the estimated regression parameters. For pairs of tests that have only one overlapping country, they use a method they call linear linking that uses the within-country mean and standard deviation in a way that appears to differ from ours.

Panel A of Figure A2.7 plots our mean country scores against the scores from Angrist et al., 2021, using the latest secondary-school math and science scores from their dataset (which stops in 2017). Two features stand out. First, for the countries in our Layers 1 (PISA) and 2 (TIMSS), the two methods yield broadly similar patterns. This suggests that in cases where there are many overlapping countries and projections that broadly fall within the mean levels observed in the set of countries participating in both tests, their linear prediction yields similar results to our method. Second, for most countries in Layer 3 (the regional tests), the Angrist et al., 2021 method tends to overestimate countries' mean achievement compared to our method, particularly on the included African test (SACMEQ). The difference is substantial: For example, in five of the nine African countries, the difference between the two methods exceeds 50 PISA points, or the equivalent of more than one and a half years of learning according to standard estimates.[1] This suggests that methods that do not draw on the full student distributions which provide common support across tests in the student-level micro data can lead to quite

[1] The countries are Kenya (difference 88.2 PISA points, compared to a standard deviation of 100), Namibia (79.8), Eswatini (67.7), Lesotho (56.2), and Uganda (52.4). For the calculation, we express achievement estimated by the Angrist et al., 2021 method (which is expressed on the TIMSS fourth-grade scale) on the PISA scale using conversion Equation 2.2.

different results in particular when projecting outside the range of observed country mean scores.[2]

In their analysis, Angrist et al., 2021 use additional test information derived from several years (2000-2017), primary as well as secondary school, and reading in addition to math and science.[3] Panel B of Figure A2.7 plots our score against their headline figure that uses this broader set of information. The overall pattern is very similar, with strong overlap for Layer 1 and 2 countries but substantial difference for Layer 3 countries. In particular, compared to our method, achievement tends to be particularly overestimated for the PASEC countries, most of which fall outside the common support of observed country mean achievement on the PISA and TIMSS tests.[4]

A second important recent paper, Patel and Sandefur, 2020, uses psychometric linkage to transform achievement on the regional PASEC and TERCE tests onto the international TIMSS and PIRLS scales in primary school.[5] A sample of students in the Indian state of Bihar is given a subset of publicly available questionnaire items from each of the four international tests in order to create a “Rosetta Stone” that allows direct linkage of scores across the tests. Conceptually, this approach is highly appealing. In practice, however, implementing the method runs into severe limitations. In particular, for most of the linkages the number of available linking items is very limited. For example, there are only four multiple-choice linking items for TERCE and twelve for PASEC math, whereas a rule of thumb suggests requiring at least 30 items for reasonable linkage (Patel and Sandefur, 2020). In addition, the choice of the sample of Bihar test-takers may reduce informational content for the linkage. Bihar is among the lowest-achieving states on the Indian ASER test, and India itself tends to perform relatively poorly on the international scale (see section 2.3.3). This implies that very low shares of students in the linkage study get any of the test items correct, which hampers international linkage.

Interestingly, the pattern comparing results based on our method to the Patel and Sandefur, 2020 method provides a very similar pattern to the previous comparison (see Panel C of Figure A2.7): While the overall pattern is broadly consistent for PISA and TIMSS participants, the Patel and Sandefur, 2020 method tends to vastly overestimate the achievement of PASEC

[2] Differences also reflect that we can draw on the more recently available PISA-D data which provide country linkage for participants on the African regional tests directly on the PISA scale.

[3] Where no data are available from PISA, TIMSS, PIRLS, or regional tests, Angrist et al., 2021 additionally include scores from the Early Grade Reading Assessment (EGRA), a short oral test of basic literacy. We do not use the EGRA data because many participating countries adapt questionnaire items to local conditions and do not draw representative samples. While EGRA provides useful information to participating school systems, it is not designed to be internationally comparable (Dubeck and Gove, 2015) and thus has major shortcomings when the aim is to provide cross-country skill comparisons.

[4] Angrist et al., 2021 include only the PASEC reading scores in their final analysis “since PASEC is the least reliable linking function, in particular for math scores” (their Supplementary information, p. 20).

[5] Another study using psychometric methods to link international test data on reading in primary school is Steinmann et al., 2014.

participants (by the equivalent of 51-89 PISA points once transformed to the PISA math scale). Again, the figure indicates that achievement in these countries falls below the common support of the international tests on our method. While conceptually appealing, the psychometric "Rosetta Stone" approach to linkage may have to await implementation on a broader scale, including both expanded test-specific questions and a more internationally representative group of test-takers.

Overall, the patterns indicate that the various methods produce rather stable results in the range of country mean achievement observed in the broad international tests, whereas differences are particularly salient for countries achieving outside that range, where missing common support at the country mean level makes use of the micro student distributions particularly valuable. Importantly, the least reliable estimates come from the countries most central to many of the development discussions.

Appendix A2.2 Alternative Transformation for China

A potential alternative approach to obtain estimates of skill deficits for China exploits the fact that four of the highest income cities and provinces of China – Beijing, Shanghai, Jiangsu, and Zhejiang (often labeled BSJZ) – have participated in PISA 2018. We can re-weight the observed PISA score distribution of the combined four provinces using data from the 2014 wave of the China Family Panel Studies (CFPS). The CFPS contains nationally representative data for 25 of the 31 provincial-level administrative divisions in China. The child questionnaire includes children aged 10 to 15 and assesses their cognitive ability in math by a crude ability test that is not psychometrically scaled. Each child can score between 0 and 24, although there is little variation in scores near the bottom of the distribution. The score corresponds to the question number of the most difficult problem that the student answered correctly.

To estimate the national PISA score distribution, we can re-weight the BSJZ distribution according to the national distribution of CFPS scores. From the percentile distribution of CFPS scores in BSJZ, we calculate the corresponding PISA scores for each point of the CFPS test distribution, i.e., $\overline{PISA}_{\kappa}^{BSJZ}$ for $\kappa = 1, 2, \ldots, 24$. We then find the corresponding proportion of students nationally that score at each point of the CFPS distribution (ω_{κ}^{China} where $\sum \omega = 1$).

We can also use the re-weighted CFPS distribution to estimate national mean achievement:

$$\overline{PISA}^{China} = \sum_{\kappa=1}^{24} \omega_{\kappa}^{cHINA} \times \overline{PISA}_{\kappa}^{BSJZ} \tag{A.1}$$

This re-weighting yields an estimate of the national average PISA score of 553.1, down from 591.4 for the four tested provinces. As discussed in section 2.4.2, these estimates are quite inconsistent with other studies on student learning in rural China.

Correspondingly, the estimation implies that just 3.2 percent of the Chinese national student population perform below the basic skill threshold. However, the conversion is highly unreliable at the relevant basic-skill threshold because only few students in the BSJZ provinces fall below the PISA basic skill level (2.2 percent) and because the CFPS test has very little support at this level. We therefore prefer the alternative conversion in section 2.3.3 based on the China Education Panel Study (CEPS) and the performance of Shanghai students in PISA 2012.

Appendix A2.3 Further Sensitivity Analyses of Economic Gains

Sensitivity analyses of the economic results for achieving universal basic skills (Scenario III) with respect to alternative parameter choices of the simulation model are shown in Appendix Table A2.7. While there are obvious interactions among the parameter choices, we isolate the independent effects through a series of individual parameter modifications.

Faster reform implementation or shorter work lives imply a quicker transformation of a country's knowledge capital. In contrast to the 15-year reform period in our baseline model, a slower 20-year reform leads to $654 trillion additional GDP, whereas a faster 10-year reform leads to a total gain of $819 trillion (columns 1-2). There is a $661 trillion improvement with a 45-year working life but a $812 trillion improvement with a 35-year working life, which allows for faster churning (columns 3-4).

The specific growth payoff for higher achievement (γ) has an obvious direct impact on the results. To account for the imprecision of the empirical estimation of the growth coefficient, we perform projections with growth coefficients that are lower or higher, respectively, by one standard error of the coefficient estimate (Hanushek and Woessmann, 2012a). The projected value of the reform ranges from $640 to $848 trillion between these two estimates (columns 5-6).

Because of the time-delayed impact of reform on growth, the economic gains start low and increase across the simulation period. As a result, the one parameter that makes the biggest difference for the results of the long-term projections is the rate at which future gains are discounted. With a higher discount rate of 4 percent, the reform value is $417 trillion, whereas it is $1,323 trillion with a lower discount rate of 2 percent (columns 7-8).

The discussion in section 2.3 highlighted some of the uncertainty in the estimates of skill deficits. We can provide some indication of the impact this uncertainty has on our estimates of the economic gains following school improvements (see Appendix Table A2.8). There is particular uncertainty about the performance level of those children not attending school. Our baseline model assumes that they achieve at the 25th percentile of the respective country distribution of in-school students. If we alternatively assume performance at the 35th or 15th percentile, the reform value is $657 or $838 trillion, respectively (columns 1-2).

The skills of children currently in school are also measured with uncertainty. To see how sensitive the simulation results are to measurement error in the skill estimates of current students, we provide bounds that assume that students' achievement increase is 10 percent lower or higher, respectively, than in our baseline model. With these bounds, estimated reform values range from \$681 to \$785 trillion (columns 3-4). To jointly consider uncertainty in the estimates of in-school and out-of-school children (as well as in the enrollment measures), we provide a similar ±10 percent bound on the achievement gain for all children. Estimates of the economic value range from \$636 to \$836 trillion for these bounds (columns 5-6).

Inherently, however, the uncertainty is not evenly distributed across the varying layers of our analysis of deficiencies in basic skills. While the Layer 1 estimates are quite certain, uncertainty about the achievement gaps increases across the other layers – being clearly the highest for the Layer 5 countries. Layer 4 also is a concern emanating from the size of India and China, implying that modest errors in the achievement estimates can have large impacts on the overall economic projections. In a final sensitivity analysis, we therefore allow the breadth of the lower and upper bounds of the estimates to increase with the layers, assuming ±5 percent of the baseline achievement increase in Layer 1 countries (where we are relatively certain about the PISA performance), ±10 percent for Layer 2 countries (TIMSS participants), ±15 percent for Layer 3 countries (participants in regional tests), ±20 percent for Layer 4 countries (India and China), and ±25 percent for Layer 5 countries (where non-participation in international tests implies high uncertainty of our imputations). With these bounds increasing with layers, the estimates of the world reform value range from \$604 to \$884 trillion (columns 7-8). Because the level of reliability tends to follow income levels (and participation in international tests), the range is substantially wider for low- than for high-income countries. Even with a steeper increase of uncertainty by layer – assuming ±5 percent for Layer 1, ±10 percent for Layer 2, ±20 percent for Layer 3, ±30 percent for Layer 4, and ±40 percent for Layer 5 – the range of the global estimate is \$560 to \$956 trillion (not shown).

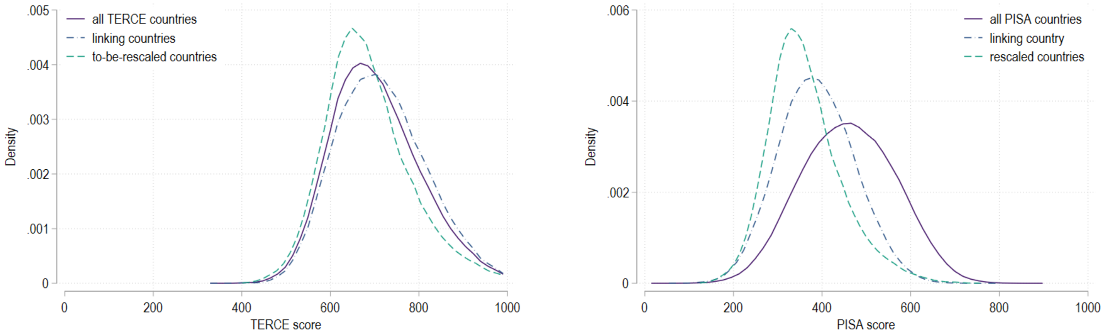
Note that these bounding analyses assume systematic errors for all countries at the lower and upper side, respectively, at the same time. In the more likely case where errors are random within each layer, in expectation the reform value would be equivalent to our baseline estimate, as country errors on either side cancel out in the world estimate.

Interestingly, for the fully imputed achievement in Layer 5, there is little uncertainty about the proportion lacking basic skills, as it is almost complete. The uncertainty comes from determining their current overall skill levels and thus the gains that would come from bringing everybody up to Level 1 performance. The limited size of the Layer 5 economies means that uncertainty there has relatively little impact on the aggregate world estimates, even though it yields substantially uncertain results for the individual countries.

Overall, the sensitivity analyses (except for the obvious relevance of the choice of discount rates in long-term projections) indicate that the economic gains from achieving universal basic skills may only be 10 percent of discounted world GDP, instead of the 12 percent in our baseline.

Appendix A2.4 Appendix Figures and Tables

Figure A2.1 : Conversion of TERCE achievement onto the PISA scale

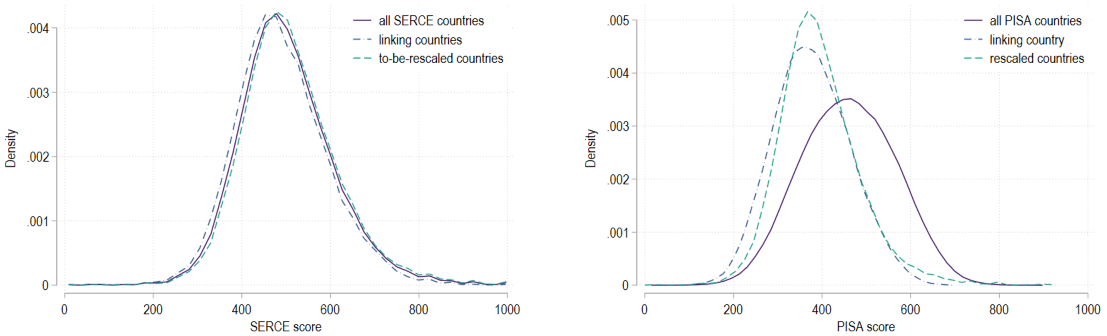


(a) Achievement on TERCE scale

(b) TERCE achievement transformed to PISA scale

Notes: Gaussian kernel densities, bandwidth 10.

Figure A2.2 : Conversion of SERCE achievement onto the PISA scale

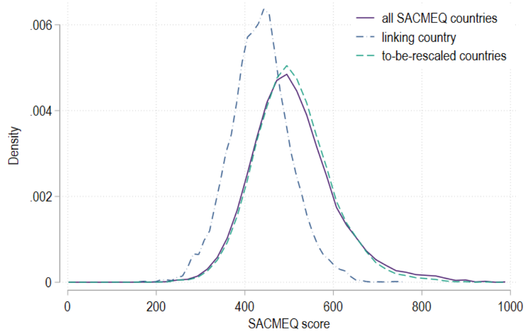


(a) Achievement on SERCE scale

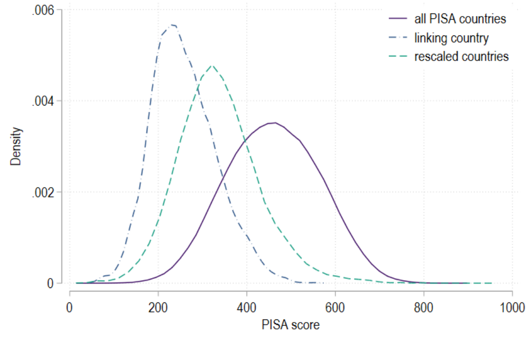
(b) SERCE achievement transformed to PISA scale

Notes: Gaussian kernel densities, bandwidth 10.

Figure A2.3 : Conversion of SACMEQ achievement onto the PISA scale



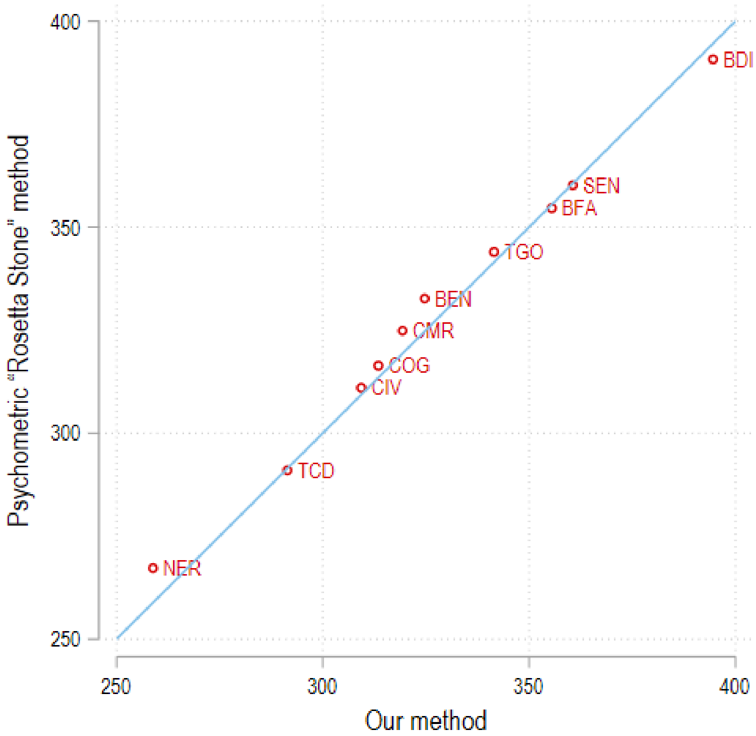
(a) Achievement on SACMEQ scale



(b) SACMEQ achievement transformed to PISA scale

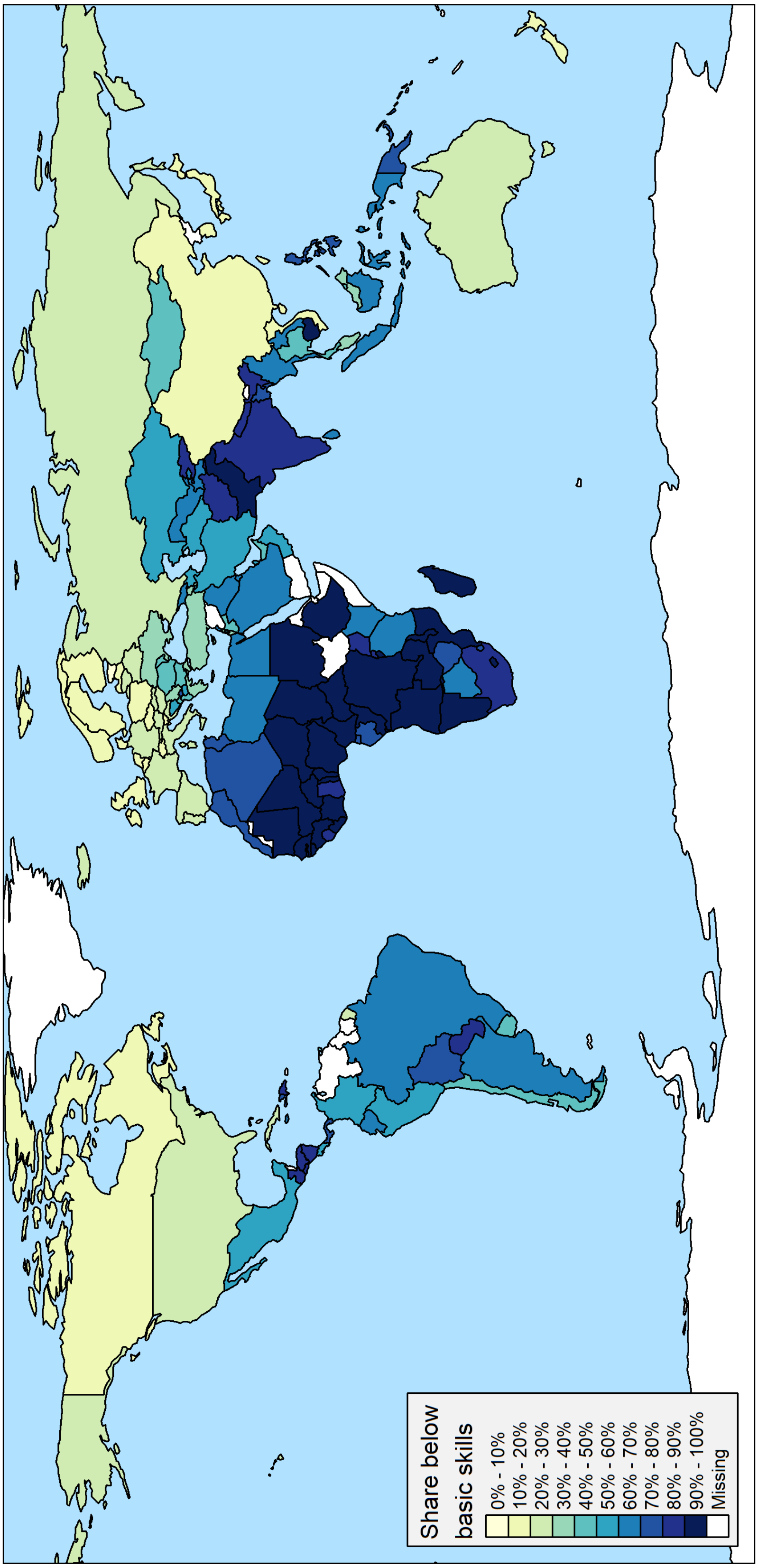
Notes: Gaussian kernel densities, bandwidth 10.

Figure A2.4 : PASEC-TIMSS linkage: Comparison of the psychometric approach and our method



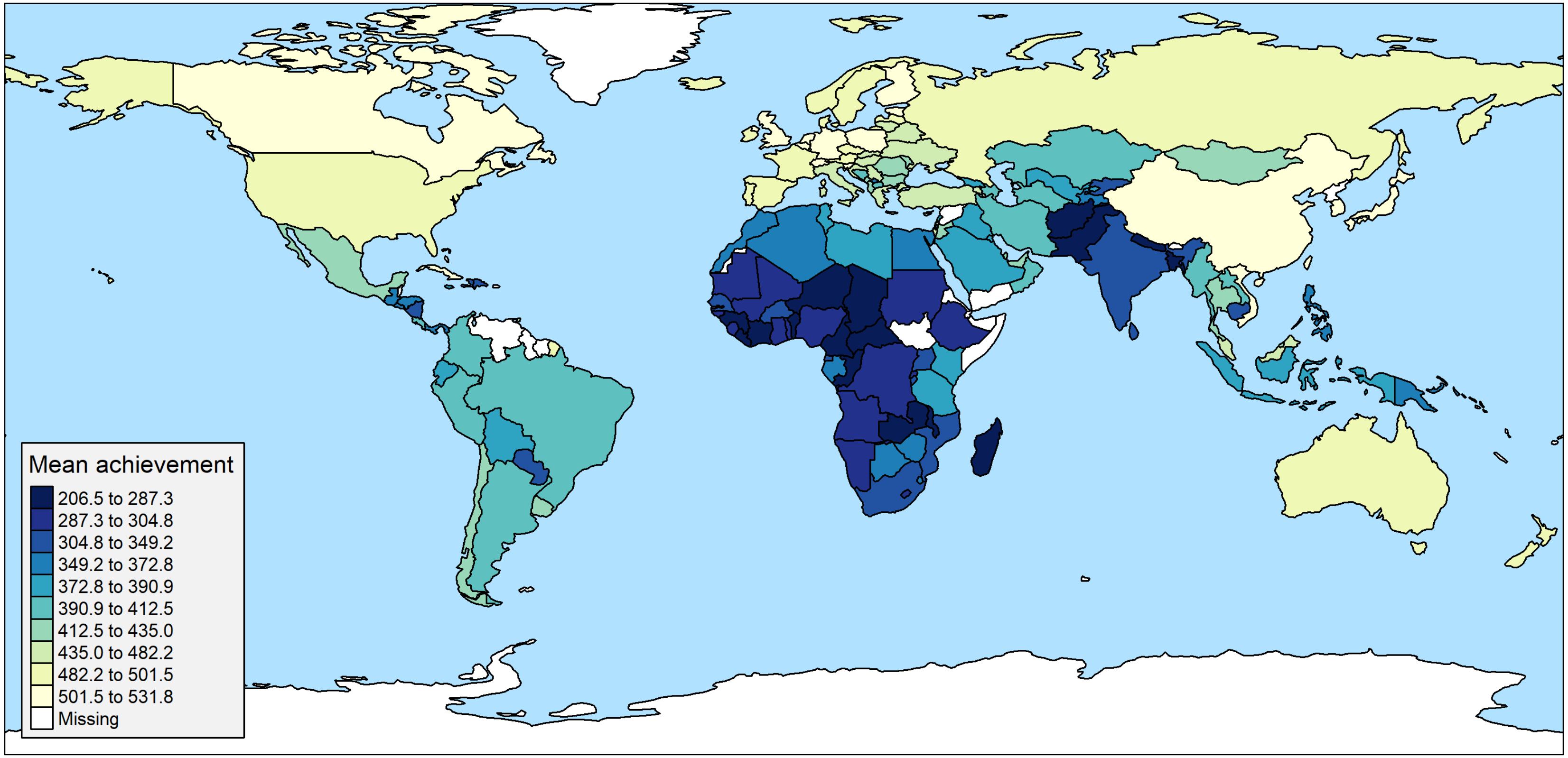
Notes: PASEC math results transformed onto fourth-grade TIMSS scale based on psychometric "Rosetta Stone" method using the concordance in UNESCO Institute for Statistics, 2022. Horizontal axis: PASEC math results transformed onto fourth-grade TIMSS scale based on our method, first transforming from PASEC into PISA and then from PISA into TIMSS using the respective linking countries.

Figure A2.5 : Share of students who do not reach basic skill levels



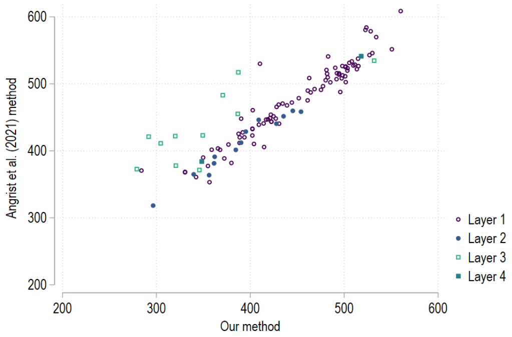
Notes: Estimated share of current students who do not reach at least basic skill levels in math and science (equivalent to PISA Level 1) in each country. See section 2.3 for methodological details.

Figure A2.6 : Mean achievement of students on a global scale

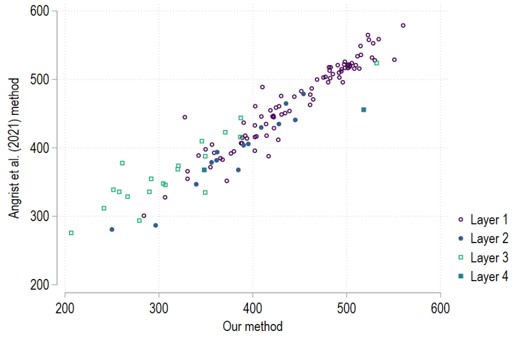


Notes: Estimated mean achievement in math and science, expressed on the PISA scale. Categories refer to deciles of the country distribution. See section 2.3 for methodological details.

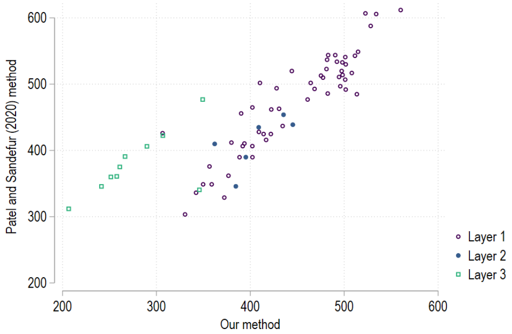
Figure A2.7 : Comparison to estimates based on alternative methods



(a) Angrist et al. (2021), latest secondary-school test, math and science



(b) Angrist et al. (2021), all tests, grades, and subjects



(c) Patel and Sandefur (2020), median math score

Notes: Data source: Angrist et al., 2021, Patel and Sandefur, 2020, and own calculations. See Appendix A2.1 for methodological details.

Table A2.1 : Linking countries for scale transformations

TIMSS 2019 and PISA 2018 Australia, Chile, Finland, France, Georgia, Hong Kong, Hungary, Ireland, Israel, Italy, Japan, Jordan, Kazakhstan, Korea, Rep., Lebanon, Lithuania, Malaysia, Morocco, New Zealand, Norway, Portugal, Qatar, Romania, Russian Federation, Saudi Arabia, Singapore, Sweden, Taiwan, Turkey, United Arab Emirates, United Kingdom, United States
TERCE and PISA 2018 Argentina, Brazil, Chile, Colombia, Costa Rica, Dominican Republic, Mexico, Panama, Peru, Uruguay
SERCE and PISA 2018 Argentina, Colombia, Dominican Republic, Panama, Peru, Uruguay
SACMEQ and PISA-D Zambia
PASEC and PISA-D Senegal

Note: See section 2.3.2 for details.

Table A2.2 : Regressions for Layer 5 imputations

		Share of students below basic skills	
	Net secondary enrollment	Math	Science
	(1)	(2)	(3)
Gross secondary enrollment	0.731***		
	-0.053		
GDP per capita / 10,000	-0.406	-0.023***	0.004
	(0.382)	(0.008)	(0.004)
$I[E_j^G > 1]$	65.96***		
	-7.733		
$I[E_j^G > 1]$ × Gross secondary enrollment	-0.666***		
	(0.076)		
Net secondary enrollment / 100		-0.440***	0.065
		(0.15)	(0.071)
Share of children below basic skills in math			0.971***
			(0.045)
Fixed effects for world regions	Yes	Yes	Yes
Fixed effects for income groups	Yes	Yes	Yes
Observations	120	100	106
R^2	0.955	0.860	0.949

Note: Country-level least squares regressions. Sample: Countries in Layers 1-3. Dependent variables indicated in column headers. Fixed effects for world regions and income groups follow World Bank classification. Standard errors in parentheses. Significance level: *** 1 percent.

Table A2.3 : Estimates of mean achievement and achievement at the 25th percentile of the country distributions

	Mean achievement (1)	Achievement at 25^{th} percentile (2)
World	381.7	323.0
By income group		
Low-income countries	289.4	238.6
Lower-middle-income countries	334.8	278.2
Upper-middle-income countries	452.6	389.7
High-income countries	488.9	423.9
By region		
Sub-Saharan Africa	302.8	251.8
South Asia	323.2	263.8
Middle East & North Africa	381.5	319.2
Latin America & Caribbean	394.5	337.3
Central Asia	438.3	382.5
East Asia & Pacific	474.2	411.2
Europe	479.7	416.2
North America	492.5	425.9

Note: Col. 1: Estimated mean achievement in math and science, expressed on the PISA scale. Col. 2: Estimated achievement at the 25^{th} percentile of the country distribution. See section 2.3 for methodological details. Country groups follow World Bank classification.

Table A2.4 : Student achievement on a global scale: Country data (1/6)

	Layer	Currently enrolled students			Enrollment	Estimated % of	Per-capita	GDP	Population
		% < basic skills	Mean achiev.	25^{th} perc. achiev.		children < basic skills	GDP		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Afghanistan	5	0.831	254.8	161.2	0.501	0.905	2,152.4	81.9	38.0
Albania	1a	0.448	427.0	373.2	0.866	0.482	14,013.0	40.0	2.9
Algeria	1b	0.759	367.7	320.4	0.750	0.796	12,009.2	517.0	43.1
Angola	5	0.914	303.2	251.5	0.362	0.958	6,952.4	221.3	31.8
Argentina	1a	0.613	391.8	330.8	0.908	0.633	22,999.3	1,033.6	44.9
Armenia	2b	0.391	435.2	378.6	0.805	0.445	14,231.2	42.1	3.0
Australia	1a	0.207	497.2	430.2	0.923	0.225	51,748.4	1,312.6	25.4
Austria	1a	0.215	494.4	426.7	0.870	0.246	58,076.3	515.7	8.9
Azerbaijan	1b	0.577	402.1	354.2	0.885	0.601	15,052.8	150.9	10.0
Bahrain	2a	0.370	445.1	375.1	0.902	0.398	47,228.1	77.5	1.6
Bangladesh	5	0.787	279.3	179.7	0.665	0.845	4,954.8	807.9	163.0
Belarus	1a	0.268	471.6	409.6	0.956	0.280	20,094.5	189.3	9.4
Belgium	1a	0.199	503.4	434.1	0.949	0.210	54,269.5	623.5	11.5
Benin	3c	0.961	266.5	215.4	0.466	0.980	3,426.3	40.4	11.8
Bolivia	5	0.711	373.7	320.0	0.766	0.752	9,093.4	104.7	11.5
Bosnia and Herzegovina	1a	0.573	402.4	346.4	0.898	0.597	15,728.2	51.9	3.3
Botswana	2b	0.698	361.9	300.8	0.453	0.802	17,039.3	39.3	2.3
Brazil	1a	0.618	393.6	329.7	0.817	0.657	15,388.2	3,247.7	211.0
Brunei Darussalam	1a	0.469	430.5	362.3	0.826	0.511	64,724.1	28.0	0.4
Bulgaria	1a	0.456	430.1	361.2	0.891	0.484	24,523.8	171.1	7.0
Burkina Faso	3c	0.931	306.7	257.2	0.310	0.966	2,267.5	46.1	20.3
Burundi	3c	0.903	349.1	305.8	0.275	0.932	783.5	9.0	11.5
Cambodia	1c	0.925	327.5	287.4	0.580	0.948	4,574.4	75.4	16.5
Cameroon	3c	0.962	260.9	206.9	0.460	0.981	3,901.1	100.9	25.9
Canada	1a	0.149	515.0	451.0	0.998	0.149	49,309.5	1,853.7	37.6
Central African Republic	5	0.959	269.5	225.6	0.127	0.992	985.1	4.7	4.7
Chad	3c	0.961	241.3	197.4	0.189	0.992	1,646.4	26.3	15.9
Chile	1a	0.437	430.5	371.5	0.887	0.466	25,395.5	481.3	19.0

Note: (continued on next page)

continued (2/6)

	Layer	Currently enrolled students			Enrollment	Estimated % of	Per-capita	GDP	Population
		% < basic skills	Mean achiev.	25^{th} perc. achiev.		children < basic skills	GDP		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
China	4b	0.139	517.9	447.4	0.815	0.180	16,653.3	23,443.7	1,407.7
Colombia	1a	0.580	402.1	344.7	0.775	0.630	15,688.6	789.8	50.3
Congo, Dem. Rep.	5	0.906	291.5	242.6	0.344	0.960	1,144.4	99.3	86.8
Congo, Rep.	3c	0.971	257.7	212.0	0.320	0.990	3,987.5	21.5	5.4
Costa Rica	1a	0.540	409.0	357.9	0.824	0.582	22,511.3	113.6	5.0
Cote d'Ivoire	3c	0.976	251.4	206.3	0.402	0.990	5,433.0	139.7	25.7
Croatia	1a	0.283	468.3	406.6	0.924	0.303	30,576.6	124.3	4.1
Cuba	3a	0.221	531.8	431.0	0.842	0.259	21,120.6	239.4	11.3
Cyprus	2a	0.310	453.8	398.0	0.953	0.323	42,384.2	37.4	1.2
Czech Republic	1a	0.196	498.1	432.6	0.905	0.218	42,847.0	457.3	10.7
Denmark	1a	0.167	501.0	442.3	0.909	0.186	58,701.0	341.3	5.8
Dominican Republic	1a	0.878	330.4	280.9	0.706	0.905	19,191.6	206.1	10.7
Ecuador	1c	0.642	388.4	337.9	0.847	0.673	11,851.5	205.9	17.4
Egypt, Arab Rep.	2a	0.675	361.2	289.8	0.828	0.711	12,260.7	1,230.8	100.4
El Salvador	2b	0.806	355.9	310.4	0.618	0.854	9,147.3	59.0	6.5
Equatorial Guinea	5	0.771	338.9	280.8	0.481	0.860	19,285.0	26.2	1.4
Estonia	1a	0.095	526.8	468.8	0.944	0.105	37,850.1	50.2	1.3
Eswatini	3b	0.794	370.5	317.8	0.417	0.855	9,018.9	10.4	1.1
Ethiopia	5	0.918	288.6	240.5	0.308	0.968	2,315.3	259.5	112.1
Finland	1a	0.140	514.6	454.4	0.961	0.147	50,321.5	277.9	5.5
France	1a	0.209	494.2	429.0	0.947	0.222	49,072.4	3,300.1	67.2
Gabon	5	0.718	351.5	290.3	0.631	0.793	15,577.9	33.8	2.2
Gambia	5	0.910	290.7	242.1	0.329	0.964	2,319.0	5.4	2.3
Georgia	1a	0.628	390.1	331.1	0.959	0.637	15,623.2	58.1	3.7
Germany	1a	0.204	501.5	431.6	0.853	0.238	55,652.9	4,624.4	83.1
Ghana	2b	0.881	296.4	227.1	0.572	0.920	5,774.3	175.6	30.4
Greece	1a	0.339	451.5	391.1	0.933	0.357	30,356.3	325.5	10.7
Guatemala	1c	0.833	349.6	305.4	0.438	0.897	9,019.3	149.8	16.6

Note: (continued on next page)

continued (3/6)

	Layer	Currently enrolled students			Enrollment	Estimated % of	Per-capita	GDP	Population
		% < basic skills	Mean achiev.	25^{th} perc. achiev.		children < basic skills	GDP		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Guinea	5	0.959	269.3	225.6	0.114	0.993	2,675.6	34.2	12.8
Guinea-Bissau	5	0.909	290.9	242.2	0.333	0.963	2,021.3	3.9	1.9
Haiti	5	0.759	347.6	300.1	0.607	0.831	3,203.3	36.1	11.3
Honduras	1c	0.802	356.4	309.0	0.438	0.875	5,978.8	58.3	9.7
Hong Kong SAR, China	1a	0.104	533.9	475.5	0.961	0.110	62,106.1	466.3	7.5
Hungary	1a	0.250	481.0	415.1	0.893	0.276	33,514.9	327.5	9.8
Iceland	1a	0.229	485.1	421.7	0.913	0.250	58,290.1	21.0	0.4
India	4a	0.851	348.2	303.8	0.636	0.888	6,997.9	9,562.0	1,366.4
Indonesia	1a	0.660	387.4	336.8	0.787	0.700	12,311.5	3,331.8	270.6
Iran, Islamic Rep.	2a	0.529	408.8	346.8	0.814	0.574	12,913.2	1,070.7	82.9
Iraq	5	0.643	378.4	315.5	0.777	0.692	11,398.0	448.1	39.3
Ireland	1a	0.164	497.9	441.4	0.987	0.167	87,379.7	431.2	4.9
Israel	1a	0.337	462.6	384.6	0.986	0.340	40,004.0	362.2	9.1
Italy	1a	0.249	477.3	415.1	0.947	0.263	44,334.2	2,648.0	59.7
Jamaica	5	0.649	387.2	331.3	0.740	0.701	10,190.5	30.0	2.9
Japan	1a	0.112	528.1	467.3	0.912	0.127	42,616.6	5,381.0	126.3
Jordan	1a	0.499	414.5	356.6	0.626	0.591	10,497.3	106.0	10.1
Kazakhstan	1a	0.548	410.1	355.4	0.998	0.549	27,466.2	508.5	18.5
Kenya	3b	0.687	387.1	324.0	0.443	0.772	4,641.1	244.0	52.6
Korea, Rep.	1a	0.146	522.5	456.2	0.980	0.150	43,044.7	2,225.8	51.7
Kosovo	1a	0.766	365.4	316.7	0.786	0.799	11,797.1	21.1	1.8
Kuwait	2a	0.621	384.5	318.8	0.865	0.650	51,962.0	218.6	4.2
Kyrgyz Republic	1b	0.843	330.4	273.4	0.844	0.860	5,480.7	35.4	6.5
Lao PDR	5	0.636	392.6	344.7	0.600	0.715	8,220.2	58.9	7.2
Latvia	1a	0.180	491.7	435.2	0.938	0.194	31,883.3	61.0	1.9
Lebanon	1a	0.611	388.6	316.2	0.855	0.644	15,179.6	104.1	6.9
Lesotho	3b	0.925	304.4	258.0	0.414	0.958	2,693.2	5.7	2.1
Liberia	5	0.960	272.8	228.2	0.157	0.991	1,531.9	7.6	4.9

Note: (continued on next page)

continued (4/6)

	Layer	Currently enrolled students			Enrollment	Estimated % of	Per-capita	GDP	Population
		% < basic skills	Mean achiev.	25^{th} perc. achiev.		children < basic skills	GDP		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Libya	5	0.635	381.0	318.0	0.775	0.685	15,815.9	107.2	6.8
Liechtenstein	1b	0.122	529.8	466.8	0.859	0.145	39,010.4	1.5	0.0
Lithuania	1a	0.240	481.6	418.2	0.984	0.244	38,540.8	107.7	2.8
Luxembourg	1a	0.271	480.1	408.3	0.836	0.313	117,341.9	72.8	0.6
Macao SAR, China	1a	0.055	550.6	497.2	0.864	0.069	132,654.9	85.0	0.6
Madagascar	5	0.923	287.2	239.4	0.298	0.971	1,687.1	45.5	27.0
Malawi	3b	0.949	279.0	232.4	0.342	0.978	1,602.1	29.8	18.6
Malaysia	1a	0.391	438.9	383.7	0.722	0.464	29,623.4	946.5	31.9
Mali	5	0.921	287.8	239.9	0.299	0.970	2,419.9	47.6	19.7
Malta	1a	0.319	464.2	390.5	0.930	0.339	45,937.7	23.2	0.5
Mauritania	5	0.938	297.0	246.6	0.310	0.972	5,570.0	25.2	4.5
Mauritius	1b	0.485	418.6	354.7	0.843	0.526	23,836.9	30.2	1.3
Mexico	1a	0.516	414.0	361.3	0.812	0.562	19,863.0	2,534.0	127.6
Moldova	1a	0.466	424.5	359.7	0.780	0.523	13,577.4	36.2	2.7
Mongolia	2b	0.416	427.6	379.4	0.746	0.484	13,014.1	42.0	3.2
Montenegro	1a	0.473	422.4	364.6	0.891	0.501	23,097.3	14.4	0.6
Morocco	1a	0.726	372.2	321.1	0.645	0.789	7,865.9	291.5	36.5
Mozambique	3b	0.912	320.6	270.5	0.193	0.959	1,336.0	40.6	30.4
Myanmar	5	0.626	394.6	345.9	0.641	0.698	4,940.2	267.0	54.0
Namibia	3b	0.927	291.6	235.7	0.682	0.945	10,227.6	25.5	2.5
Nepal	5	0.807	274.0	175.5	0.619	0.868	4,119.9	117.9	28.6
Netherlands	1a	0.179	511.3	440.4	0.932	0.194	59,004.3	1,023.4	17.3
New Zealand	1a	0.200	501.5	433.4	0.969	0.207	45,437.9	226.2	5.0
Nicaragua	3a	0.855	345.6	305.9	0.724	0.882	5,682.2	37.2	6.5
Niger	3c	0.960	206.5	165.6	0.201	0.992	1,276.2	29.7	23.3
Nigeria	5	0.938	296.9	246.5	0.310	0.971	5,352.7	1,075.7	201.0
North Macedonia	1a	0.553	403.7	339.7	0.757	0.611	17,565.2	36.5	2.1
Norway	1a	0.199	495.7	432.3	0.956	0.209	66,799.2	357.2	5.3

Note: (continued on next page)

continued (5/6)

	Layer	Currently enrolled students			Enrollment	Estimated % of	Per-capita	GDP	Population
		% < basic skills	Mean achiev.	25^{th} perc. achiev.		children < basic skills	GDP		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Oman	2a	0.557	395.1	324.4	0.962	0.567	32,607.0	162.2	5.0
Pakistan	2c	0.904	249.8	156.9	0.374	0.952	4,896.4	1,060.4	216.6
Panama	1a	0.763	358.7	302.7	0.638	0.818	32,769.9	139.2	4.2
Papua New Guinea	5	0.754	362.3	320.9	0.324	0.872	4,474.7	39.3	8.8
Paraguay	1c	0.840	342.1	294.6	0.789	0.863	13,149.0	92.6	7.0
Peru	1a	0.575	402.0	343.9	0.893	0.600	13,397.3	435.6	32.5
Philippines	1a	0.794	354.8	301.7	0.656	0.841	9,291.7	1,004.6	108.1
Poland	1a	0.143	513.3	451.3	0.941	0.155	33,797.8	1,283.1	38.0
Portugal	1a	0.215	492.1	426.8	0.947	0.227	36,172.1	372.1	10.3
Qatar	1a	0.511	416.7	344.6	0.761	0.570	93,771.1	265.6	2.8
Romania	1a	0.453	427.8	363.6	0.828	0.498	31,901.4	618.0	19.4
Russian Federation	1a	0.215	482.8	425.0	0.907	0.237	29,967.1	4,398.1	144.4
Rwanda	5	0.898	293.7	244.4	0.359	0.956	2,321.7	29.3	12.6
Saudi Arabia	1a	0.676	379.7	325.1	0.964	0.683	48,948.2	1,677.4	34.3
Senegal	1c	0.941	306.4	261.0	0.377	0.971	3,503.6	57.1	16.3
Serbia	1a	0.391	444.1	377.2	0.921	0.412	18,842.5	130.9	6.9
Sierra Leone	5	0.875	299.3	248.7	0.418	0.939	1,777.3	13.9	7.8
Singapore	1a	0.081	560.0	497.2	0.998	0.081	102,573.4	585.0	5.7
Slovak Republic	1a	0.273	475.1	408.0	0.848	0.312	31,966.6	174.4	5.5
Slovenia	1a	0.155	508.0	447.3	0.957	0.164	40,670.9	84.9	2.1
South Africa	2a	0.807	339.6	276.0	0.719	0.843	14,289.8	836.8	58.6
Spain	1a	0.231	482.3	421.1	0.969	0.238	41,696.3	1,965.3	47.1
Sri Lanka	5	0.671	309.7	204.0	0.910	0.694	13,622.9	297.0	21.8
Sudan	5	0.901	293.2	244.3	0.341	0.959	4,350.1	186.2	42.8
Sweden	1a	0.190	500.9	435.7	0.991	0.192	54,598.8	561.2	10.3
Switzerland	1a	0.186	505.3	436.9	0.853	0.219	72,033.9	617.7	8.6
Taiwan, China	1a	0.146	523.4	457.3	0.975	0.151	55,078.2	1,300.2	23.6
Tajikistan	5	0.635	368.6	319.7	0.804	0.687	3,732.9	34.8	9.3

Note: (continued on next page)

continued (6/6)

	Layer	Currently enrolled students			Enrollment	Estimated % of	Per-capita	GDP	Population
		% < basic skills	Mean achiev.	25^{th} perc. achiev.		children < basic skills	GDP		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Tanzania	3b	0.662	386.6	330.5	0.265	0.787	2,773.2	156.1	58.0
Thailand	1a	0.487	422.2	362.5	0.773	0.542	19,233.9	1,339.2	69.6
Timor-Leste	5	0.634	392.5	344.1	0.627	0.709	3,780.0	4.9	1.3
Togo	3c	0.915	289.8	232.8	0.410	0.958	2,211.6	17.9	8.1
Trinidad and Tobago	1b	0.491	420.9	351.9	0.882	0.521	26,920.1	37.6	1.4
Tunisia	1b	0.705	376.6	325.4	0.790	0.741	11,900.0	139.2	11.7
Turkey	1a	0.310	460.9	400.5	0.872	0.342	26,867.5	2,241.5	83.4
Turkmenistan	5	0.533	406.9	350.4	0.888	0.561	16,195.5	96.2	5.9
Uganda	3b	0.889	320.0	272.9	0.126	0.957	2,275.2	100.7	44.3
Ukraine	1a	0.312	461.1	398.0	0.640	0.406	13,346.5	560.9	44.4
United Arab Emirates	1a	0.442	434.3	359.0	0.928	0.461	71,150.5	695.2	9.8
United Kingdom	1a	0.184	503.2	437.7	0.971	0.190	49,041.5	3,277.8	66.8
United States	1a	0.229	490.3	423.4	0.925	0.247	65,279.5	21,433.2	328.3
Uruguay	1a	0.474	421.7	361.4	0.882	0.505	24,006.8	83.1	3.5
Uzbekistan	5	0.613	387.8	334.2	0.909	0.634	7,658.9	257.2	33.6
Vietnam	1b	0.125	509.6	453.1	0.554	0.212	8,381.2	808.5	96.5
West Bank and Gaza	2b	0.579	390.1	321.4	0.872	0.610	6,509.6	30.5	4.7
Zambia	1c	0.960	283.9	238.6	0.348	0.983	3,617.2	64.6	17.9
Zimbabwe	3b	0.778	349.2	284.9	0.404	0.860	3,783.5	55.4	14.6

Note: Col. 1: Layer of reliability of underlying achievement information (see Table 2.1 and section 2.2 for details). Col. 2: Estimated share of current students who do not reach at least basic skill levels in math and science (equivalent to PISA Level 1). Col. 3: Estimated mean achievement in math and science, expressed on the PISA scale. Col. 4: Estimated achievement at the 25th percentile of the country distribution. Col. 5: Net secondary enrollment rate (from WDI and own imputations). Col. 6: Estimated share of children (incl. those currently out of school) who do not reach at least basic skill levels in math and science. Col. 7-9: GDP per capita (2019, PPP, current prices), GDP (billion), and population (million), respectively (from WDI and own imputations). See section 2.3 for methodological details.

Table A2.5 : Sensitivity of skill estimates: Alternative bounds on India and China

	Baseline	India		China	
		Based on Tamil Nadu	Based on Himachal Pradesh	Based on rural Vietnam	Based on rural Cambodia
	(1)	(2)	(3)	(4)	(5)
India	0.888	0.886	0.901		
China	0.18			0.19	0.689
World	0.672	0.672	0.675	0.676	0.741
By income group					
Lower-middle-income countries	0.858	0.858	0.864		
Upper-middle-income countries	0.423			0.472	0.638
By region					
South Asia	0.892	0.89	0.901		
East Asia & Pacific	0.354			0.36	0.633

Note: Estimated share of children (incl. those currently out of school) who do not reach at least basic skill levels in math and science (equivalent to PISA Level 1). Col. 1: baseline results (see col. 3 of Table 2.2). Col. 2 and 3: assume India achieves at level of Tamil Nadu and Himachal Pradesh, respectively. Col. 4 and 5: assume that China baseline value applies only to urban children (35%) and that rural children (65%) achieve at the level of rural Vietnam and rural Cambodia, respectively.

Table A2.6 : Economic gains from achieving universal basic skills: Country results (1/7)

	Scenario I		Scenario II		Scenario III			
	bn USD (1)	% curr. GDP (2)	bn USD (3)	% curr. GDP (4)	bn USD (5)	% curr. GDP (6)	% disc. GDP (7)	GDP 2100 (8)
Afghanistan	1,095	1337%	556	680%	4,559	5568%	119%	750%
Albania	120	301%	37	93%	165	413%	9%	37%
Algeria	3,279	634%	796	154%	5,528	1069%	23%	103%
Angola	1,293	584%	1,017	460%	6,656	3008%	64%	348%
Argentina	6,600	639%	742	72%	8,089	783%	17%	73%
Armenia	113	269%	60	143%	179	425%	9%	38%
Australia	1,734	132%	866	66%	2,163	165%	4%	14%
Austria	658	128%	587	114%	950	184%	4%	16%
Azerbaijan	664	440%	106	70%	851	564%	12%	51%
Bahrain	269	347%	68	88%	341	440%	9%	39%
Bangladesh	13,638	1688%	3,756	465%	33,444	4140%	89%	517%
Belarus	347	183%	65	34%	390	206%	4%	18%
Belgium	826	132%	278	45%	953	153%	3%	13%
Benin	411	1017%	151	373%	1,599	3955%	85%	488%
Bolivia	700	668%	172	164%	1,138	1087%	23%	105%
Bosnia and Herzegovina	253	488%	38	73%	319	614%	13%	56%
Botswana	171	434%	183	467%	656	1671%	36%	172%
Brazil	18,345	565%	4,945	152%	27,760	855%	18%	80%
Brunei Darussalam	95	341%	43	154%	147	524%	11%	47%
Bulgaria	661	386%	166	97%	865	506%	11%	45%
Burkina Faso	218	474%	220	478%	1,352	2933%	63%	338%
Burundi	26	284%	39	435%	162	1789%	38%	187%
Cambodia	581	771%	168	223%	1,424	1888%	40%	199%
Cameroon	1,049	1039%	405	401%	4,271	4231%	91%	531%
Canada	1,744	94%	28	2%	1,755	95%	2%	8%

Note: (continued on next page)

continued (2/7)

	Scenario I		Scenario II		Scenario III			
	bn USD (1)	% curr. GDP (2)	bn USD (3)	% curr. GDP (4)	bn USD (5)	% curr. GDP (6)	% disc. GDP (7)	GDP 2100 (8)
Central African Republic	11	232%	25	543%	205	4383%	94%	555%
Chad	110	418%	131	501%	1,437	5475%	117%	734%
Chile	1,581	328%	414	86%	2,072	430%	9%	38%
China	10,217	44%	40,039	171%	22,957	98%	2%	8%
Colombia	3,428	434%	1,339	169%	5,710	723%	15%	67%
Congo, Dem. Rep.	582	586%	442	446%	3,276	3299%	71%	390%
Congo, Rep.	148	692%	92	431%	965	4497%	96%	573%
Costa Rica	425	374%	132	116%	630	554%	12%	50%
Cote d’Ivoire	1,310	938%	515	368%	6,318	4522%	97%	577%
Croatia	225	181%	74	60%	277	223%	5%	19%
Cuba	324	135%	504	211%	575	240%	5%	21%
Cyprus	88	235%	12	33%	98	263%	6%	23%
Czech Republic	529	116%	364	80%	699	153%	3%	13%
Denmark	322	94%	233	68%	416	122%	3%	10%
Dominican Republic	2,015	977%	394	191%	3,756	1822%	39%	191%
Ecuador	1,121	544%	205	100%	1,555	755%	16%	70%
Egypt, Arab Rep.	11,465	932%	1,978	161%	16,968	1379%	29%	138%
El Salvador	350	593%	136	231%	796	1348%	29%	134%
Equatorial Guinea	155	591%	109	416%	534	2043%	44%	218%
Estonia	23	45%	21	42%	27	55%	1%	5%
Eswatini	39	374%	44	426%	149	1437%	31%	145%
Ethiopia	1,372	529%	1,205	464%	8,925	3439%	74%	411%
Finland	229	83%	83	30%	257	92%	2%	8%
France	4,730	143%	1,460	44%	5,469	166%	4%	14%
Gabon	246	726%	103	305%	564	1667%	36%	172%

Note: (continued on next page)

continued (3/7)

	Scenario I		Scenario II		Scenario III			
	bn USD (1)	% curr. GDP (2)	bn USD (3)	% curr. GDP (4)	bn USD (5)	% curr. GDP (6)	% disc. GDP (7)	GDP 2100 (8)
Gambia	30	559%	25	454%	182	3345%	72%	397%
Georgia	391	674%	18	30%	428	736%	16%	68%
Germany	5,804	126%	6,167	133%	8,730	189%	4%	16%
Ghana	1,968	1120%	717	408%	5,646	3214%	69%	378%
Greece	820	252%	167	51%	969	298%	6%	26%
Guatemala	651	435%	506	338%	2,400	1603%	34%	164%
Guinea	71	207%	188	549%	1,508	4413%	94%	560%
Guinea-Bissau	22	566%	18	452%	129	3332%	71%	395%
Haiti	230	637%	90	249%	548	1518%	32%	154%
Honduras	236	405%	212	364%	886	1521%	33%	154%
Hong Kong SAR, China	288	62%	133	29%	319	69%	1%	6%
Hungary	519	158%	298	91%	696	213%	5%	18%
Iceland	30	144%	15	71%	38	183%	4%	16%
India	64,656	676%	20,453	214%	139,747	1461%	31%	147%
Indonesia	17,010	511%	4,661	140%	26,908	808%	17%	75%
Iran, Islamic Rep.	4,864	454%	1,613	151%	7,454	696%	15%	64%
Iraq	2,799	625%	824	184%	4,671	1042%	22%	100%
Ireland	404	94%	41	10%	420	97%	2%	8%
Israel	1,229	339%	49	13%	1,273	351%	8%	31%
Italy	4,717	178%	1,117	42%	5,429	205%	4%	18%
Jamaica	156	518%	58	192%	276	917%	20%	87%
Japan	3,073	57%	3,671	68%	4,018	75%	2%	6%
Jordan	320	302%	309	291%	769	725%	16%	67%
Kazakhstan	2,489	489%	6	1%	2,498	491%	11%	44%
Kenya	914	375%	1,203	493%	3,301	1353%	29%	135%

Note: (continued on next page)

continued (4/7)

	Scenario I		Scenario II		Scenario III			
	bn USD (1)	% curr. GDP (2)	bn USD (3)	% curr. GDP (4)	bn USD (5)	% curr. GDP (6)	% disc. GDP (7)	GDP 2100 (8)
Korea, Rep.	2,207	99%	372	17%	2,337	105%	2%	9%
Kosovo	146	692%	29	136%	229	1084%	23%	105%
Kuwait	1,583	724%	250	114%	2,134	976%	21%	93%
Kyrgyz Republic	448	1265%	41	115%	623	1760%	38%	183%
Lao PDR	230	391%	151	256%	528	895%	19%	85%
Latvia	59	96%	27	45%	71	116%	2%	10%
Lebanon	724	696%	142	136%	1,012	972%	21%	93%
Lesotho	39	681%	21	373%	161	2808%	60%	320%
Liberia	21	283%	40	532%	318	4209%	90%	528%
Libya	648	605%	200	186%	1,088	1015%	22%	97%
Liechtenstein	1	51%	2	114%	1	80%	2%	7%
Lithuania	174	162%	14	13%	182	169%	4%	14%
Luxembourg	124	170%	112	154%	195	268%	6%	23%
Macao SAR, China	20	24%	79	93%	30	35%	1%	3%
Madagascar	234	515%	213	469%	1,591	3497%	75%	419%
Malawi	193	647%	127	424%	1,098	3678%	79%	446%
Malaysia	1,928	204%	1,915	202%	3,824	404%	9%	36%
Mali	245	515%	223	469%	1,655	3478%	74%	416%
Malta	66	285%	15	66%	79	342%	7%	30%
Mauritania	129	513%	123	488%	820	3253%	70%	383%
Mauritius	122	405%	39	130%	177	586%	13%	53%
Mexico	8,716	344%	3,264	129%	13,421	530%	11%	48%
Moldova	125	346%	68	188%	212	587%	13%	53%
Mongolia	99	235%	67	160%	177	421%	9%	37%
Montenegro	52	359%	12	81%	67	465%	10%	41%

Note: (continued on next page)

continued (5/7)

	Scenario I		Scenario II		Scenario III			
	bn USD (1)	% curr. GDP (2)	bn USD (3)	% curr. GDP (4)	bn USD (5)	% curr. GDP (6)	% disc. GDP (7)	GDP 2100 (8)
Morocco	1,480	508%	703	241%	3,242	1112%	24%	108%
Mozambique	104	257%	234	577%	1,108	2731%	58%	310%
Myanmar	1,097	411%	620	232%	2,295	860%	18%	81%
Namibia	340	1332%	60	236%	712	2793%	605	318%
Nepal	1,870	1586%	626	531%	5,274	4475%	965	570%
Netherlands	1,162	113%	634	62%	1,425	139%	3%	12%
New Zealand	305	135%	61	27%	334	148%	3%	13%
Nicaragua	301	810%	53	143%	522	1404%	30%	141%
Niger	158	532%	135	455%	2,127	7149%	153%	1031%
Nigeria	5,535	515%	5,245	488%	35,024	3256%	70%	384%
North Macedonia	167	459%	75	206%	294	805%	17%	75%
Norway	499	140%	126	35%	562	157%	3%	13%
Oman	1,247	769%	55	34%	1,358	837%	18%	79%
Pakistan	10,360	977%	9,328	880%	68,626	6472%	138%	908%
Panama	886	636%	378	272%	1,987	1428%	31%	144%
Papua New Guinea	109	277%	151	384%	558	1421%	30%	143%
Paraguay	887	958%	120	130%	1,367	1476%	32%	149%
Peru	2,202	506%	347	80%	2,802	643%	14%	59%
Philippines	6,818	679%	2,441	243%	14,463	1440%	31%	145%
Poland	1,029	80%	599	47%	1,238	96%	2%	8%
Portugal	545	146%	165	44%	631	170%	4%	14%
Qatar	1,114	420%	607	228%	1,984	747%	16%	69%
Romania	2,158	349%	886	143%	3,258	527%	11%	47%
Russian Federation	5,425	123%	3,026	69%	7,029	160%	3%	14%
Rwanda	177	604%	129	439%	941	3211%	69%	377%

Note: (continued on next page)

continued (6/7)

	Scenario I		Scenario II		Scenario III			
	bn USD (1)	% curr. GDP (2)	bn USD (3)	% curr. GDP (4)	bn USD (5)	% curr. GDP (6)	% disc. GDP (7)	GDP 2100 (8)
Saudi Arabia	12,608	752%	423	25%	13,619	812%	17%	76%
Senegal	342	599%	222	389%	1,583	2772%	59%	315%
Serbia	408	312%	89	68%	499	381%	8%	34%
Sierra Leone	96	688%	56	406%	412	2968%	64%	343%
Singapore	278	48%	10	2%	280	48%	1%	4%
Slovak Republic	328	188%	231	133%	487	279%	6%	24%
Slovenia	74	88%	28	33%	85	100%	2%	8%
South Africa	8,083	966%	1,984	237%	15,138	1809%	39%	189%
Spain	3,019	154%	476	24%	3,290	167%	4%	14%
Sri Lanka	6,275	2113%	365	123%	7,874	2651%	57%	299%
Sudan	1,064	572%	836	449%	6,056	3252%	70%	383%
Sweden	719	128%	43	8%	739	132%	3%	11%
Switzerland	622	101%	804	130%	964	156%	3%	13%
Taiwan, China	1,311	101%	275	21%	1,407	108%	2%	9%
Tajikistan	237	681%	43	124%	357	1026%	22%	99%
Tanzania	340	218%	921	590%	2,176	1393%	30%	140%
Thailand	4,271	319%	2,381	178%	7,350	549%	12%	50%
Timor-Leste	20	411%	12	239%	43	886%	19%	84%
Togo	133	742%	84	469%	613	3428%	73%	409%
Trinidad and Tobago	164	437%	39	105%	218	581%	12%	53%
Tunisia	845	608%	194	139%	1,321	949%	20%	90%
Turkey	4,155	185%	2,229	99%	5,846	261%	6%	23%
Turkmenistan	482	501%	78	81%	612	636%	14%	58%
Uganda	169	168%	593	589%	2,790	2770%	59%	315%
Ukraine	885	158%	1,720	307%	2,224	397%	8%	35%

Note: (continued on next page)

continued (7/7)

	Scenario I		Scenario II		Scenario III			
	bn USD (1)	% curr. GDP (2)	bn USD (3)	% curr. GDP (4)	bn USD (5)	% curr. GDP (6)	% disc. GDP (7)	GDP 2100 (8)
United Arab Emirates	2,972	428%	482	69%	3,557	512%	11%	46%
United Kingdom	4,022	123%	779	24%	4,374	133%	3%	11%
United States	32,433	151%	13,837	65%	40,032	187%	4%	16%
Uruguay	317	382%	76	91%	418	503%	11%	45%
Uzbekistan	1,767	687%	161	63%	2,122	825%	18%	77%
Vietnam	274	34%	2,769	343%	1,110	137%	3%	12%
West Bank and Gaza	212	696%	35	113%	283	929%	20%	88%
Zambia	432	669%	263	407%	2,286	3539%	76%	425%
Zimbabwe	276	498%	301	543%	1,177	2124%	45%	229%

Note: Scenario I: Current students achieve at least basic skills (equivalent to PISA Level 1). Scenario II: Full participation at current level. Scenario III: All children achieve at least basic skills (equivalent to PISA Level 1). Discounted value of future increases in GDP until 2100 due to the reform scenario, expressed in billion USD, as a percentage of current GDP, and as a percentage of discounted future GDP. “GDP 2100” indicates by how much GDP in 2100 is higher due to the reform (in percent). See section 2.5.2 for details on the reform scenarios and section 2.5.3 for details on the simulation model.

Table A2.7 : Sensitivity of simulation results: Alternative parameter choices

	Reform duration (R)		Working life (W)		Growth coefficient (γ)		Discount rate (d)	
	20 years (1)	10 years (2)	45 years (3)	35 years (4)	0.0176 (5)	0.022 (6)	4% (7)	2% (8)
World	654,209	818,922	660,711	811,978	639,717	847,821	416,732	1,322,616
By income group								
Low-income countries	36,403	46,934	36,715	46,599	35,118	49,505	23,130	75,894
Lower-middle-income countries	340,069	432,351	343,213	429,004	330,203	452,056	216,327	698,792
Upper-middle-income countries	164,445	201,858	166,218	199,961	162,191	206,282	104,895	325,811
High-income countries	113,292	137,779	114,565	136,414	112,205	139,978	72,380	222,119
By region								
Sub-Saharan Africa	111,219	142,181	112,212	141,124	107,710	149,137	70,686	229,952
South Asia	229,620	293,158	231,710	290,929	222,543	307,383	146,062	473,744
Middle East & North Africa	60,025	74,027	60,651	73,353	59,076	75,867	38,257	119,562
Latin America & Caribbean	69,324	85,307	70,063	84,520	68,298	87,306	44,199	137,746
Central Asia	11,986	14,617	12,121	14,476	11,857	14,877	7,655	23,575
East Asia & Pacific	83,744	102,493	84,661	101,517	82,707	104,547	53,449	165,365
Europe	50,353	61,137	50,922	60,522	49,904	62,053	32,174	98,541
North America	37,938	46,002	38,371	45,537	37,622	46,651	24,250	74,131

Note: Scenario III: All children achieve at least basic skills (equivalent to PISA Level 1). Discounted value of future increases in GDP until 2100 due to the reform scenario, expressed in billion USD. See section 2.5.2 for details on the reform scenarios and section 2.5.3 for details on the simulation model. Country groups follow World Bank classification.

Table A2.8 : Sensitivity of simulation results: Measurement error in skill estimates

	Out-of-school children		In-school children		All children		Uncertainty increasing with layer	
	35^{th} perc.	15^{th} perc.	Achievement increase		Achievement increase		Lower bound	Upper bound
			– 10%	+10%	– 10%	+10%		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
World	657,030	838,119	680,518	785,284	635,967	835,800	603,637	883,629
By income group								
Low-income countries	35,577	49,802	39,707	43,049	34,867	48,639	27,868	59,019
Lower-middle-income countries	338,324	444,724	358,273	409,895	328,065	444,853	292,889	492,355
Upper-middle-income countries	165,672	206,453	168,565	196,237	161,364	203,839	164,812	200,322
High-income countries	117,457	137,140	113,973	136,103	111,671	138,469	118,068	131,933
By region								
Sub-Saharan Africa	109,779	148,613	119,884	131,919	106,982	146,678	91,323	169,212
South Asia	228,241	300,995	241,434	278,432	221,066	302,330	193,747	338,490
Middle East & North Africa	62,506	72,469	60,714	72,829	58,764	74,930	59,649	74,102
Latin America & Caribbean	71,236	84,606	70,785	83,217	67,944	86,250	71,764	82,259
Central Asia	12,530	14,395	12,039	14,463	11,799	14,712	11,824	14,740
East Asia & Pacific	82,307	108,652	86,516	98,933	82,292	103,345	83,158	102,403
Europe	51,329	62,248	50,937	60,102	49,672	61,393	52,560	58,453
North America	39,102	46,141	38,209	45,389	37,448	46,162	39,612	43,970

Note: Scenario III: All children achieve at least basic skills (equivalent to PISA Level 1). Discounted value of future increases in GDP until 2100 due to the reform scenario, expressed in billion USD. Uncertainty increasing with layer: achievement increase – / + 5% for Layer 1, 10% for Layer 2, 15% for Layer 3, 20% for Layer 4, and 25% for Layer 5. See section 2.5.2 for details on the reform scenarios and section section 2.5.3 for details on the simulation model. Country groups follow World Bank classification.

3 (Not) Going to School in Times of Climate Change: Natural Disasters and Student Achievement

3.1 Introduction

The rising risk of natural disasters and extreme weather due to climate change poses significant challenges globally. In 2021, the United States experienced a record-breaking streak, with seven consecutive years of over ten billion-dollar disasters (Smith, 2022). Population growth and development patterns increase the damage costs of these events (Smith, 2022). Extending beyond immediate physical damage to broader socio-economic dimensions is crucial to comprehend the multifaceted impacts of natural disasters and long lasting effects of climate change. One critical aspect is the disruption caused to educational systems. Nearly half of the school closures in the US, before the Covid-19 pandemic, were attributed to natural disasters (Jahan et al., 2022). While school closures are perhaps the most visible manifestation of how natural disasters affect learning, it represents only one facet of a complex phenomenon. Natural disasters can affect child health (Currie and Rossin-Slater, 2013), school quality, and infrastructure damage, as well as induce family income instability (Deuchert and Felfe, 2015) and housing displacement with unequal effects for different socio-economic groups (Nguyen and Minh Pham, 2018).

This paper investigates the effects of natural disasters on student achievement and unravels the underlying mechanisms. Focusing on student achievement allows for a nuanced examination of the cognitive skill component of human capital, capturing variations in any skill inputs including ability, child health, family support, school resources, and institutional characteristics (Woessmann, 2016a). I combine county level student achievement, achievement gaps, demographic compositions, and school financial information from the Stanford Education Data Archive (SEDA) (Reardon et al., 2019) with disaster declarations from the Federal Emergency Management Agency (FEMA).

The variation of natural disasters across US counties and school years serves as the foundation for my analysis in a two-way fixed effects framework. First, I study how the effect of natural disasters dynamically evolves over the school years following the event, using the event study approach by Sun and Abraham, 2021. Second, I estimate static two-way fixed effects models with several treatment specifications to study different intensity channels such as the type of natural disaster, the frequency by which natural disasters strike different counties, and disaster size along various dimensions such as damage, fatalities, and duration.

I find persistent negative effects of natural disasters on student achievement for up to five years after a natural disaster hits a county. Natural disasters adversely affect even those cohorts who were not yet enrolled at the time of the disaster, an effect that cannot be explained by school closures. Possible explanations include a decline in children's health, which hinders cognitive and social-emotional development, changes in the student composition of affected areas, and reduced school quality due to infrastructure damage or the loss of skilled teachers. I show suggestive evidence that financial distress in families may be a key mechanism. Specifically, natural disasters increase the share of students receiving free or reduced lunch, which seems to be unrelated to migration patterns. I present new evidence related to the resilience of schools to natural disasters and show that counties with above-average per-pupil expenditure recover more quickly from natural disasters.

Exploring different treatment intensities shows that not only very large disasters drive the results. Although point estimates are consistently higher for large disasters, below-average disasters have a significant negative effect on student achievement. In terms of the type of disaster, volcanic activity and landslides have the most devastating effect. Those disasters occur very rarely and might be harder to predict. Also hurricanes, fires, and earthquakes lead to significantly lower student achievement. Generally, predictability does not alleviate all damage. My results show that more frequent disasters are more harmful, indicating that a higher frequency leaves the county little time to recover and does not allow them to adapt sufficiently to disasters.

I contribute to the literature on the adverse effects of natural disasters during early life which shows that in-utero and post-birth exposure to natural disasters (Currie and Rossin-Slater, 2013) or pollution (Klauber et al., 2024) negatively affect various health outcomes, cognitive skills, and income (Karbownik and Wray, 2019). This strand of literature attributes the persistent negative results to a reduction in the health stock of children. While this might explain parts of the effects, I demonstrate that school inputs play a crucial role. The existing literature typically focuses on one specific major disaster event such as hurricanes (Sacerdote, 2012; Özek, 2023) or earthquakes (Di Pietro, 2018), which are often extreme outliers as the majority of disasters are not as severe. My paper is more closely related to Opper et al., 2023 who focus on the static effects of differently sized disasters on human capital and migration.[1] By leveraging recent advancements in the difference-in-difference methodology, I derive robust estimates regarding how these effects evolve over time.

The remainder of the paper is organized as follows: Section 3.2 illustrates the possible mechanisms and presents the existing empirical evidence. Section 3.3 describes the different data

[1] Opper et al., 2023, who developed their paper in parallel with this one, estimate the impact of disasters on net migration, average test scores, high school graduation rates, and post-secondary enrollment rates using FEMA disaster declarations and the SEDA database. Their study differs from my paper in three main ways: First, they focus on the effects of varying disaster sizes based on property damage, while this study also considers disaster frequency and type. Second, they present only static effects on first-difference outcomes, whereas this study provides robust event study results. Third, this paper offers additional insights into the impacts on student composition, adult mental health, and variations by gender, socio-economic background, and per-pupil spending.

sources in more detail. Section 3.4 contains the two main empirical strategies, including the event study design and the static two-way fixed effects model. The results are presented in Section 3.5, and Section 3.6 concludes.

3.2 Conceptual Framework and Existing Evidence

Natural disasters can affect educational outcomes through several mechanisms.[2] First, natural disasters affect the shadow price of quality education, reflecting its accessibility (Nguyen and Minh Pham, 2018). This includes infrastructure damage, temporary or permanent school closures, as well as effects on teaching staff.

Natural disasters and extreme weather events account for the most frequent cause of prolonged unplanned school closures in the US prior to the Covid-19 pandemic, as depicted in Figure 3.1 panel a). Panel b) splits those natural disaster school closures by disaster type. Hurricanes, ice and snowfall cause most of the school closures. There is a wide range of literature showing that instruction time correlates with student performance (Pischke, 2007; Lavy, 2015; Aucejo and Romano, 2016; Jaume and Willén, 2019; Wedel, 2021).[3] In Maryland, students experiencing reduced instruction time due to unscheduled closures during snowfall performed worse on high-stakes exams (Marcotte, 2007). Drawing from Massachusetts data, Goodman, 2014 provides evidence that coordination problems in the classroom, as outlined by Lazear, 2001, play an important role in this context. While heavy snowfall leads to coordinated school closures and no effect on achievement, moderate snowfall induces student absence and reduces math achievement by 0.05 standard deviations. The Covid-19 pandemic has also provided further insights into the detrimental impact of school closures. Both cognitive and socio-emotional development were significantly impeded, with students from disadvantaged homes experiencing more severe setbacks (Engzell et al., 2021; Werner and Woessmann, 2023). However, natural disasters do not necessarily increase the cost of quality education. Sacerdote, 2012 shows that students forced to switch school after the hurricanes Katrina and Rita experienced a sharp decline in test scores in the first year after the hurricanes. Yet, the long-run effects are mixed, with Sacerdote (2012) identifying an improvement in test scores among low-performing students placed in higher-quality educational settings.

Secondly, natural disasters can increase the costs of good health or impede the access to a healthy environment (Nguyen and Minh Pham, 2018). The evidence on wild fires is closely related to the broader literature on pollution exposure (Wu, 2022). Currie et al., 2009 show that a rise in pollution increases absence, possibly due to health issues, decreases cognitive skills (Lavy, 2015), but also hinders long-run knowledge acquisition due to impaired brain development (Block and Calderón-Garciduenas, 2009). Currie and Rossin-Slater, 2013 find

[2] Nguyen and Minh Pham, 2018 present a simple model to illustrate the different mechanisms through which natural disasters affect children's development, which guides the structure of this section.

[3] See Blanden et al., 2023 for a detailed review of the evidence.

negative effects of hurricane exposure during pregnancy on children. They demonstrate that the evidence regarding additional impacts of hurricanes during pregnancy on outcomes such as birth weight – subsequently influencing adult height, IQ, earnings, and education (Black et al., 2007) – is more mixed and depending on the specification. Klauber et al., 2024 observe no effect of cleaner air around birth on birth weights but show that children require less medication for at least five years. Fuller, 2014 shows that hurricane exposure during pregnancy in North Carolina translates into lower standardized test scores in math and reading by the third grade, while children exposed to floodings or tornadoes also exhibit somewhat diminished math performance. Karbownik and Wray, 2019 study the long-run effects of hurricane exposure in utero and as an infant by using World War I draft records linked to census data and find that white males had 5% lower income. The literature on in-utero exposure links these effects to stress during pregnancy. To test this channel, I incorporate data on adult mental health.

Thirdly, natural disasters can affect children's education through household income, wages, and increased costs of other commodities (Masozera et al., 2007; Arouri et al., 2015; Boustan et al., 2020; Pleninger, 2022). Deuchert and Felfe, 2015 demonstrate that damages to real estate redirect investment toward house reconstruction, potentially diverting resources away from children's health and education. I contribute to this channel by estimating the effects of natural disasters on the share of students receiving free or reduced lunch, which serves as an indicator of increased financial distress in affected households.

Certain characteristics influence the vulnerability to natural disasters (Cutter et al., 2008), including socio-economic status, race, ethnicity, and gender, as wealth and social safety nets can facilitate recovery. Boys often react more strongly to disruptive (family) events, showing fluctuations in test scores and increased disciplinary issues (Bertrand and Pan, 2013; Autor et al., 2016, 2019). My research enhances the literature by analyzing achievement patterns separately for boys and girls and assessing how natural disasters impact the gender achievement gap and the socio-economic gap.

Vulnerability is closely linked to disaster resilience, which Cutter et al., 2008 classify into social, economic, institutional, infrastructure, and community capital. Insurances play a crucial role in mitigation: Pleninger, 2022 demonstrates that unemployment insurance effectively reduces the impact of natural disasters on income. Moreover, insurance and revenue diversification enhance the resilience of municipal bonds against price drops following natural disasters (Auh et al., 2022). However, municipalities with above-average racial minority compositions experience greater expenditure losses and a heightened debt default risk compared to average municipalities in the decade following a hurricane (Jerch et al., 2023). In turn, this constrains public expenditures: Deryugina, 2017 illustrates that hurricanes lead to increased unemployment and disability insurance claims but a decline in educational assistance transfers. I contribute to this literature by studying whether counties with higher pre-existing per-pupil expenditures display enhanced resilience to natural disasters.

The literature underscores that the mechanisms might vary depending on the type of natural disaster. In this paper, I study the effect of different types of natural disasters. While numerous studies concentrate exclusively on individual types of disasters, such as hurricanes (Sacerdote, 2012) or earthquakes (Di Pietro, 2018), which are often extreme outliers and relatively infrequent occurrences, the majority of disasters are not as severe. I analyze heterogeneity across various dimensions of disaster size, including factors such as fatalities, costs, and duration.

3.3 Data

3.3.1 Natural Disasters

The OpenFEMA Dataset by the Federal Emergency Management Agency (FEMA) of the Department of Homeland Security contains all major disaster declarations since 1964. A disaster declaration made by the President of the United States and only in strongly affected areas that struggle to deal with the consequences, which rules out any inconsequential natural disasters. Every disaster declaration includes the date the disaster was declared, the area, the type of incident, and which assistance program was declared. One disaster can cause multiple disaster events across different counties. The years are adjusted to align with school years, so the year variable matches the education data.

For large disasters, there is information on the number of deaths, the number of injured, the number of people that got homeless, reconstruction costs, insured damage and total damage in the international emergency events database (EM-DAT) that I merge via county and start date. Generally, EM-DAT considers only disasters that caused more than 10 deaths, that affected more than 100 people, or that called for international assistance or an emergency declaration. This information refers to overall fatalities and damage for a disaster and is not specified on a county level.

Panel a) Figure 3.2 shows the number of disaster declarations in FEMA by year. Each disaster event is only counted once, even when it affected multiple counties. In 2011, the year with the most disaster declarations, there were 177 events, while in 2015, there were the fewest declarations (79).[4] Since then, the number has been increasing every year. Panel b) shows disasters by type: 502 fires, 414 storms, 4 earthquakes, and 2 volcanic activities. Panel c) displays the average fatalities for those disasters for which we can merge EM-DAT data. Hurricanes have the highest average fatalities with 45 deaths, followed by landslides with 43 fatalities even though those were only 4 events. Fires are the most frequent type, but they have no fatalities. Panel d) depicts the average disaster damage in thousands of dollars as recorded in EM-DAT. In EM-DAT, adjusted damage refers to the financial losses caused by a disaster, which are normalized to account for factors such as inflation using CPI. Hurricanes emerge as

[4] Smith, 2022 shows that the frequency and costs of severe natural disasters have been on the rise over the past four decades since the 1980s, with variations observed between individual years.

the most economically devastating disasters, with storms following closely behind, causing approximately half the financial losses incurred by hurricanes. In contrast, landslides and fires are characterized by minimal direct financial impact.

3.3.2 Outcome Data

Student achievement: The Stanford Education Data Archive (SEDA) offers a unique source for school district- and county-level student achievement measured by standardized test scores for the 2008/09 to 2017/18 school years (Fahle et al., 2021). Moreover, it contains achievement gaps by gender and socio-economic status, demographic, and socio-economic data. Achievement in SEDA is based on the EDFacts data system that contains test data for all students in grades three to eight in maths and reading each year. EDFacts does not contain individual student-level data but the number of students in each school, subgroup, subject, grade, and year scoring at each performance level. Unfortunately, every state can design their own test and benchmark for performance levels, such that the data are not comparable across states. SEDA transforms these state specific benchmarks and places them onto a common scale across states and years using the National Assessment of Educational Progress (NAEP) and calculates achievement for different subgroups and geographical units. They exclude cases with low participation and insufficient data. Finally, the score is standardized by subtracting the average of the four national cohorts that were in fourth grade in 2009, 2011, 2013, and 2015 and dividing by the national grade-subject-specific standard deviation of this reference cohort. Consequently, a county mean of 0.5 indicates that the average student scored approximately 0.5 standard deviations higher than the average national reference cohort in that same grade. One standard deviation on this scale is approximately three grade levels. Estimates on this scale are comparable across the US and over time by relying on the stability of the NAEP scale over time.[5]

Table 3.1 provides summary statistics extracted from the SEDA dataset for several key variables. The final analysis will focus on the youngest children in grade three. However, I will also show results for grades four and five.[6] The math grades for grades three, four, and five have mean scores ranging from -0.016 to -0.064. The socio-economic gap in math for grade three has a mean of -0.229. Conversely, the male-female gap appears relatively low, with a mean of -0.023, suggesting minimal disparity between genders in math and grade three. On average, there are 13% Black students and 70% White students. 60% of the students get a free or reduced lunch. School size is captured through log enrollment, which has a mean value of 5.969. Furthermore, SEDA provides financial aspects of education, revealing mean values of $11,524 for total per-pupil expenditure and $11,572 for total per-pupil revenue. I will show results by high and low per-pupil expenditure, where high per-pupil expenditure exceeds $11,000, which is close to the median value. Finally, Table 3.1 displays the population estimate for 2008 that I use to weight the regressions.

[5] See Reardon et al., 2019 for a more detailed description on the SEDA data construction.

[6] Grades six to eight have substantially more missing values in SEDA.

SEDA is a repeated aggregate cross-section. However, Reardon et al., 2019 show that it highly correlates with longitudinal data on school and district level. The correlation is higher at the district level ($r = 0.87$) than at the school level ($r = 0.80$), which they attribute to higher mobility between schools than between districts. On the county level, the mobility-induced measurement error in learning rates will be even lower.

In- and out-migration: A natural disaster might increase mobility between counties. To study this impact, I add county level in- and out-migration from the Statistics of Income Division migration data. The Statistics of Income Division uses the number of personal exemptions claimed to approximate the number of individuals. Total values for migration to and from other US counties and abroad are available for 2010 to 2018. I divide the total inflow and total outflow by the population estimate from 2008. Table 3.1 contains the summary statistics. One limitation of the data is that those who are not required to file United States Federal income tax returns are not included. Thus, elderly and people of low socio-economic status are underrepresented. Additionally, the data contains estimates for the aggregate adjusted gross income of the in- and outflows.

Adult mental health: The county-level mental health data for 2010 to 2018 come from the County Health Rankings & Roadmaps (CHR&R) (Remington et al., 2015), a program of the University of Wisconsin Population Health Institute. The main outcome variable is the average number of poor mental health days among adults. Table 3.1 contains the summary statistics.

3.4 Identification Strategy

3.4.1 Difference-in-Differences Event Study Design

To explore the evolution of the impact of a natural disaster over time and evaluate the validity of the parallel trend assumption, I adopt a difference-in-differences event study design. Given the staggered occurrence of natural disasters across different time periods, various counties may exhibit learning curves or encounter changes in external conditions, resulting in heterogeneous treatment effects. The presence of such heterogeneous treatment effects complicates the identification of a clean control group and conventional DiD models are prone to generating biased estimates (De Chaisemartin and d'Haultfoeuille, 2020; Callaway and Sant'Anna, 2021; Goodman-Bacon, 2021; Gardner, 2022; Borusyak et al., 2024). Sun and Abraham, 2021 illustrate this contamination concerning the coefficients related to lead and lag indicators within a dynamic specification of the two-way fixed effects model. They propose a re-weighting procedure to address this bias. Their target parameter is the cohort average treatment effect on the treated (CATT) for a treatment cohort e and a relative time period l:

$$CATT_{e,l} = E[Y_{i,e+l} - Y^{\infty}_{i,e+l} | E_i = e] \tag{3.1}$$

In my setting, a treatment cohort e are counties that are treated in the same year. Panel a) of Figure 3.3 shows the first treatment period of each county. l are the periods to the initial natural disaster of county i within the observed time frame. $Y_{i,e+l}$ is the outcome in response to the treatment. $Y^{\infty}_{i,e+l}$ is the potential outcome of county i in a world where it is untreated. Each $CATT_{e,l}$ is then the average treatment effect l periods from initial treatment for all counties first treated at time e. The main identifying assumptions are parallel trends in the baseline outcomes and no treatment effect on pre-treatment periods, and potential treatment effect heterogeneity.

I estimate all $CATT_{e,l}$ using a linear two-way fixed effects event study model that interacts relative period indicators with cohort indicators, excluding the last treated as control C and the year before treatment $l = -1$:

$$Y_{i,t} = \alpha_i + \lambda_t + \sum_{e \notin C} \sum_{l \neq -1} \delta_{e,l}(1\{E_i = e\} \cdot D^l_{i,t}) + \epsilon_{i,t}, \tag{3.2}$$

The county fixed effects α_i remove any geographic differences between the counties that are time-invariant, such as risk of experiencing a natural disaster, elevation, proximity to the coast or differences in time-constant spending. The year fixed effects λ_t capture any time-variant factors across different years. $1\{E_i = e\}$ is an indicator for a county i belonging to treatment cohort e, and $D^l_{i,t}$ is an indicator for county i being l years away from treatment. Alternatively, one can also use never-treated counties as control, which I implement as a robustness check.

$\hat{\delta}_{e,l}$ is the difference-in-differences estimator for $CATT_{e,l}$ that needs re-weighting. Sun and Abraham, 2021 estimate the weights $Pr(E_i = e | E_i \in [-l, T-l])$ with total periods T by sample shares of each cohort in the relevant period $l \in g$ from its corresponding set g.

To derive the interaction-weighted estimator, Sun and Abraham, 2021 take the weighted average over all estimates for $CATT_{e,l}$ multiplied by the sample share of each cohort in the period $Pr(E = e)$:

$$\hat{v}_g = \frac{1}{|g|} \sum_{l \in g} \sum_{e} \hat{\delta}_{e,l} \widehat{Pr}(E_i = e | E_i \in [-l, T-l]) \tag{3.3}$$

Under parallel trends, limited anticipation, and potential treatment effect heterogeneity, $\hat{v}_g$ is consistent. Unless specified otherwise, standard errors are clustered at the county level.

3.4.2 Static Two-Way Fixed Effects

Some counties experience multiple disasters, and disasters vary in intensity. To illustrate these channels, I estimate a static two-way fixed effects model where several post treatment indicators are collected in a set. The baseline equation is the following:

$$Y_{i,t} = \alpha_i + \lambda_t + \beta D_{i,t}^{1-5} + \epsilon_{i,t}, \tag{3.4}$$

where α_i and λ_t are county and year fixed effects, respectively. First, I binarize the treatment following Callaway et al., 2024 such that $D_{i,t}^{1-5} = 1\{D_{i,t} > 0\}$ is a dummy of whether a county experienced at least one disaster in the past five years in a county and year, similar to Deryugina, 2017 and Jerch et al., 2023.[7] The control group are counties that did not experience a natural disaster in the past five years. As a robustness check, I include state times year fixed effects to compare counties within states and account for time-varying factors specific to each state. Standard errors are clustered at the county level and all estimations are weighted by the county population in 2008.

The first intensification channel is the severity of a natural disaster. I estimate the equation:

$$Y_{i,t} = \alpha_i + \lambda_t + \delta_1 Major_{i,t}^{1-5} + \delta_2 Minor_{i,t}^{1-5} + \epsilon_{i,t}, \tag{3.5}$$

$Major_{i,t}^{1-5}$ is an indicator equal to 1 if a county experienced at least one major disaster in the past 5 years. $Minor_{i,t}^{1-5}$ is an indicator equal to 1 if the county experienced no major disaster but at least one minor disaster in the past five years, similar to Jerch et al., 2023. All other specifications are identical to Equation 3.4. $\hat{\delta}_1$ and $\hat{\delta}_2$ provide estimates of the impact of any type of disaster that falls into either of these two categories. I use three alternative methods to measure the severity of a disaster. In the first specification, I follow Boustan et al., 2020 and define major natural disasters as such that caused more than 25 deaths, which corresponds roughly to the median value of fatalities. Boustan et al., 2020 argue that the actual number of fatalities might be determined by economic development, which is why they avoid using the actual number of fatalities and prefer to use this simple threshold. Alternatively, one can distinguish between major and minor disasters based on costs. I define major natural disasters as such that cause more than one billion dollar (adjusted) total damage. Lastly, I split disasters on whether the event lasted more than 50 business days or less.

The second intensification channel is the number of natural disasters. More frequent disasters could be more detrimental if counties have no time to recover (Pleninger, 2022). Counties might also enhance their level of protection through increased exposure to disaster events.

Let the number of natural disasters be the dose d that a county experienced in the past five years. To study if more disasters cause more harm or whether counties are adapting, I estimate the following model:

[7] The dummy includes the past five years to capture longer term effects of the treatment.

$$Y_{i,t} = \alpha_i + \lambda_t + \sum_{j=1}^{J} 1\{D_{i,t}^{1-5} = d_j\}\gamma_j + \epsilon_{i,t}, \tag{3.6}$$

where $\{D_{i,t}^{1-5} = d_j\}$ is a series of dummy variables equal to 1 if county i in year t experienced a dose of d_j natural disasters in the past five years, with untreated units as the omitted category. The spatial distribution of the number of natural disasters is shown in panel b) of Figure 3.3. This treatment specification follows the multi-valued discrete setting in Callaway et al., 2024 and the OLS coefficients $\hat{\gamma} = (\hat{\gamma}_1, ..., \hat{\gamma}_J)$ are estimators of the average level treatment effect over treatment dosages. However, comparison between different dosages requires stronger assumptions than standard parallel trends. Under strong parallel trends, the path of outcomes for lower-dose units must reflect the path of higher dose units had they received the lower-dose. In absence of this condition, the comparison across dose groups can still be interpreted as a causal response. However, it is contaminated by the selection bias (Callaway et al., 2024).

3.5 Results

3.5.1 Event Study Results

To assess how student achievement evolves over time, I estimate the event study design from Equation 3.2, following Sun and Abraham, 2021.

Figure 3.4 shows that there is a negative effect of natural disasters on student achievement from the year a county is hit by a natural disaster (year = 0) until up to five years later. Comparing different student cohorts over time, math achievement in grade three is about 0.025 of a standard deviation lower one year after the natural disaster. Since one standard deviation is approximately three grade levels, students fall around 0.075 grade levels behind due to the natural disaster. The effect is very persistent, such that students in grade three in counties that experienced a natural disaster five years earlier are still 0.03 standard deviations behind. The results look similar for grades four and five, although the effects are somewhat smaller for grade four and again smaller for grade five. For grade five, only the effect one year after the disaster is statistically significant.

The analysis primarily focuses on grade three, as grades six to eight are not as extensively covered in the SEDA data. Despite these limitations, the results for grades six to eight are presented in Figure A3.1, which also indicate negative effects in the higher grades. However, these effects are not always significantly different from zero. Grade eight shows a positive point estimate in the year of the disaster, but it is close to zero and not statistically significant, making it difficult to draw any firm conclusions.

The coefficients in the years before the natural disaster in Figure 3.4 are all not significantly different from zero, reassuring that the parallel trends assumption holds. One issue discussed in the recent difference-in-differences literature is that pre-trend tests can be under-powered, such that one cannot reject the absence of pre-trends, nor can one reject the potential existence of pre-trends that would cause significant bias (Bilinski and Hatfield, 2018; Freyaldenhoven et al., 2019; Kahn-Lang and Lang, 2020; Roth, 2022).[8] Roth, 2022 finds that linear violations of parallel trends, which pre-trend tests detect only 50 percent of the time, can cause biases equal to or greater than the estimated treatment effect. Table A3.2 suggests that the pre-trend test from the baseline event study would detect a small linear trend of magnitude 0.006 with 50 percent power and a linear trend of 0.010 (0.009 for grade four and five) with 80 percent power. However, the low likelihood ratio of the observed coefficients under the linear trend of 0.010 relative to parallel trends favors parallel trends. For the small linear trend of 0.006 under 50 percent power, the likelihood ratio is still low for grade five but closer to one for grade three and four. I illustrate this linear trend as the red solid line in Figure A3.2. It is likely that a linear violation did not cause the estimated effect if it cannot fully explain the pattern in the event study. For grade three and five, the linear trend falls outside the confidence intervals. The linear trend would cause a bias of at most 0.013 for grade three in the year after the natural disaster, which is considerably smaller than the estimate. For grade four, the results are less robust and the bias would be larger or similar to the estimated treatment effect for the later post-treatment years if such a linear trend existed. I will conduct further robustness checks following Borusyak et al., 2024 and assess the sensitivity of the results in a number of ways.

Another issue highlighted by Roth, 2022 is that samples failing to detect a linear trend in the population means can suffer from selection bias. This bias often increases the bias arising from violations of parallel trends. To illustrate this, I assume that the true population means follow the red, solid line in Figure A3.2. The dotted blue line shows the expected coefficients on average, conditional on not finding a significant pre-trend. However, in this analysis, the blue dotted line is very close to the red, solid line, which suggests that the selection bias does not exacerbate the bias in my baseline event study.

Additionally, I estimate event study results following the imputation method by Borusyak et al., 2024 and the two-stage difference-in-differences method by Gardner, 2022. The imputation method proposed by Borusyak et al., 2024 separates pre-trend testing from the estimation of treatment effects, removing the correlation between treatment effects and pre-trend estimators.[9] This approach avoids the bias introduced by pre-testing, as highlighted by Roth, 2022. Figure A3.3 shows that the pre-trend coefficients, with the exception in grade three and five, are statistically insignificant. The F-test cannot reject the hypothesis that coefficients are

[8] Pre-trend tests reverse the roles of a type I and a type II error. A 95 percent confidence interval sets the probability of finding a violation when parallel trends actually hold to 5 percent, but the probability of failing to detect a pre-trend can be much higher (Bilinski and Hatfield, 2018).

[9] Borusyak et al., 2024 run the pre-trend test with a dynamic TWFE specification on the set of untreated observations only.

jointly equal to zero with a p-value of 0.184 for grade three and 0.706 for grade four, lending further support to the parallel trends assumption. Only for grade five, the F-test weakly rejects the hypothesis that coefficients are jointly equal to zero with a p-value of 0.025. The imputation method and the two-stage difference-in-differences method in Figure A3.4 generally confirm the negative effects on student achievement from grades three to five. However, both methods produce larger standard errors. Interestingly, they produce a positive point estimate for grade three, four years after the disaster event. As with Sun and Abraham, 2021, the period-four estimate is not significantly different from zero.

In the standard setting of Sun and Abraham, 2021 in Figure 3.4, the last treated cohort serves as the control group. As a sensitivity analysis, one can augment the control group by using counties that were never treated as a control. Figure A3.5 illustrates that this specification yields virtually identical results. Sun and Abraham, 2021 provide a method particularly for the event study setting, making it an especially appropriate baseline model for this study. Additionally, I estimate event study results following the imputation method by Borusyak et al., 2024 and the two-stage differences-in-differences method by Gardner, 2022. Figure A3.3 shows that these methods generally confirm the negative effects on student achievement from grades three to five. However, both methods produce larger standard errors. Interestingly, they produce a positive point estimate for grade three, four years after the disaster event. As with Sun and Abraham, 2021, the period-four estimate is not significantly different from zero.

The main specification with Sun and Abraham, 2021 focuses on a single grade level, comparing different cohorts over time. Alternatively, we can pool all grade levels and add cohort fixed effects to account for variations across cohorts. Panel a) of Figure A3.6 shows the pooled math score of all grade levels. Panel b) adds cohort fixed effects to the pooled math scores. Both plots appear nearly identical. The results in Figure A3.6 confirm a significant negative effect in the year of the natural disaster and the year after. In the following years, the point estimates show negative values, gradually approaching zero each year and becoming statistically insignificant. By the fifth year after the natural disaster, the point estimate slightly exceeds zero, but without significant deviation from it.

In summary, student achievement experiences a significant negative decline when a disaster strikes a county, particularly evident in the year of the natural disaster and one year after. These adverse effects persist for up to five years following the natural disaster. Notably, the impact is most pronounced and enduring among grade three students compared to those in grades four and five. This implies that disasters detrimentally affect cohorts, even those not yet enrolled in school at the time of the event. For example, a student showing lower math performance in third grade five years after the disaster would have been of preschool age at the time of the disaster, still three years away from starting elementary school. The negative effects may stem from a potential decline in preschool quality and supply. Part of this impact could also be linked to a reduction in children's health stock, consistent with findings by Fuller, 2014, who reports negative impacts on grade three achievement for children whose parents were exposed to hurricanes during pregnancy. Other possible mechanisms include financial difficulties, a decline in infrastructure and school quality, especially if high-quality teachers relocate from the affected areas or if teachers and parents experience significant mental distress. Such distress can reduce parental support and affect the quality of teaching, ultimately impacting student achievement.

Achievement gaps: The persistent negative effects in grade three could affect vulnerable children more strongly than others and thus widen achievement gaps. The existing literature suggests that boys and girls respond differently to family and school environments. (Bertrand and Pan, 2013; Autor et al., 2016, 2019), with boys often demonstrating more adverse outcomes in test scores and disciplinary issues in response to disruptive (family) events. Descriptively, the male-female gap in mathematics within the SEDA data is negligible. Figure 3.5 shows that in grade three, boys (panel b) face more negative effects than girls (panel a). However, panel c) indicates that the effect on the male-female gap is not significant, except for one year after the natural disaster. There is also no significant effect on the gender gap in grade four, as shown in Figure A3.7. However, in grade 5, there is a significant negative effect on the male-female gap, suggesting that boys may experience more pronounced setbacks in learning outcomes as a result of such events.

Panel a) of Figure 3.6 displays the negative effect on economically disadvantaged children in grade three.[10] However, panel b) shows no clear evidence of a widening socio-economic achievement gap post-disaster in grade three, except for the period five years after the natural disaster. For grade four and five, Figure A3.8 suggests a widening of the socio-economic gap, indicating that students from economically disadvantaged backgrounds are more adversely affected by the disaster.

Per-pupil spending: In terms of effect heterogeneities, an alternative approach involves exploring different county attributes to investigate whether certain counties demonstrate greater resilience to natural disasters. I split the sample into ex-ante low and high per-pupil spending counties. Counties with per-pupil spending above USD 11,000 in 2009, which was approximately the average, are classified as high-spending counties. All other counties are classified as low-spending counties. Figure 3.7 shows that high per-pupil spending counties only experience significant negative effects in the year following the natural disaster, then appear to recover more quickly. In contrast, low per-pupil spending counties exhibit negative results until three years after the natural disaster. However, Table A3.1 indicates that counties with lower per-pupil spending also exhibit lower ex-ante student achievement and a higher proportion of students eligible for free or reduced-price lunch. Consequently, the observed effect cannot be solely attributed to differences in per-pupil spending. Instead, it underscores the broader disparities in disaster resilience that are closely tied to financial resources, as highlighted by Cutter et al., 2008. The link between financial investment and resilience to

[10] SEDA does not provide a separate achievement score for children who are not economically disadvantaged.

natural disasters appears to be a consistent theme across different domains. Auh et al., 2022 find that municipal bonds backed by diversified revenue sources are generally resilient to natural disasters, except for those issued by municipalities in weak financial condition. This exception is attributed to the challenges faced by financially burdened municipalities in diversifying away the shock caused by severe natural disasters.

3.5.2 Heterogeneity by Disaster Characteristics

The previous section demonstrates that the effects on student achievement are consistently negative over time. Notably, the impact across the post-disaster periods remains similar, justifying their aggregation into a static effect. The specifications from Chapter 3.4.2 allow for more flexibility to study different channels, such as disaster severity, frequency, and type. Beginning with the baseline estimate for the static TWFE model in Equation 3.4, the initial column of Table 3.2 displays the outcomes without weighting. On average, third-grade students in counties experiencing at least one natural disaster within the past five years demonstrate a performance decline of 0.021 standard deviations. In the subsequent column, the preferred specification integrates population weights to address variations in population sizes across US counties. According to this specification, student achievement declines by 0.028 standard deviations. Model 3 incorporates state times school year fixed effects to enable comparisons among counties within states. The state-school year fixed effects control for unobserved time-varying state effects such as state fiscal shocks. However, this specification may inadvertently absorb some of the treatment effect (Wolfers, 2006). As anticipated, the coefficient is half the size, yet the overall interpretation of the results remains consistent. The static estimates closely align with the average across the five post-treatment periods when employing the event study design in Figure 3.4.

Severity of disasters: The negative coefficient observed in Table 3.2 may be driven by severe natural disasters, while smaller-scale disasters may not have a significant impact. To investigate this possibility, I estimate Equation 3.5 and differentiate between minor and major natural disasters. In the first column of Table 3.3, I define major disasters as those resulting in more than 25 deaths, following the definition by Boustan et al., 2020. All regressions are population-weighted. Counties in the aftermath of a major natural disaster exhibit a performance decline of 0.038 standard deviations compared to counties unaffected by natural disasters. However, the coefficient on minor natural disasters remains close to the overall effect with a significant of coefficient of -0.027, indicating that also below-average natural disaster cause severe harm.

In column two, I dissect disasters by their costs, employing a threshold of 1 billion dollars in damages as the cutoff point. The point estimate for major disasters is -0.042, but it is not only the 1 billion dollar disasters that impede learning. Minor disasters in this specification decrease student achievement by 0.025 of a standard deviation. Lastly, I redefine major disasters based on the duration of the disaster declaration, with durations exceeding 50 days classified as major disasters. Counties experiencing a disaster that lasted more than 50 days have a -0.068 lower student achievement, but again, also shorter disasters cause significant harm in student achievement with a significant coefficient of -0.025. These results align with Opper et al., 2023, who also find the largest effects from very large disasters, defined as those causing over $500 per capita property damage. They also show that even disasters exceeding $100 per capita property damage lead to significant negative impacts on student achievement.[11]

Multiple disasters: In some counties, multiple natural disasters occurred within the observed time frame. The occurrence of several disasters could intensify their adverse effects, leaving counties with limited time to recover. However, another perspective suggests that counties may develop greater resilience and adaptability to natural disasters over time, potentially reducing the harm caused by frequent occurrences. I estimate Equation 3.6, which includes a dummy variable for each count of disasters. The counterfactual are counties that did not experience any disaster in the past five years.

Figure 3.8 demonstrates a clear trend: as the number of natural disasters increases, students in affected counties experience progressively worse academic performance. This suggests that frequent occurrences of natural disasters may leave counties with insufficient time for recovery, contributing to the persistent decline in academic performance.

Disaster type: Certain types of natural disasters may have a more pronounced impact on student achievement than others. As depicted in Figure 3.9, volcanic activity and landslides cause the greatest harm. However, those are relatively rare events and are thus unlikely to drive the results. Earthquakes, though rare, can cause substantial infrastructure damage and incur high costs. These events are associated with a 0.05 standard deviation decrease in student achievement. More frequent events, such as hurricanes and fires, reduce student achievement by 0.05 standard deviations, while the effects of storms and floods are close to zero and not significant.

3.5.3 Mechanisms

The results show that disasters detrimentally affect cohorts, even those not yet enrolled in school at the time of the event. Thus, missed days in school cannot be the only factor at play. There are many potential mechanisms that may play a role, including long-lasting infrastructure damage, relocation of higher-performing students or high-ability teachers, financial difficulties of families, and a decline in the health stock of children and caregivers. Some of

[11] Unlike Equation 3.5, Opper et al., 2023 only consider disasters that occurred within the same year and estimate their effect on the first difference of student achievement.

these mechanisms can be tested directly with the available data. In this section, I discuss them in more detail, focusing on migration patterns, changes in the student composition, and mental health of the adult population.

Migration: The loss of human capital following a natural disaster poses a significant detriment to a region or county, whether it be through the decline in cognitive abilities of the population or through a brain drain. One concern is that some families might have moved away after the disaster event. If families move within counties, this is still reflected in the student achievement county score. However, the migration of selected families from or into the county might partially account for the decline in student achievement. Overall, around 5% of the population of the county baseline population leaves the county each year and 5% of the county baseline population moves in from another county.

Figure 3.10 shows the event studies for in- and out-migration of the county. Interestingly, there is an increase in the inflow and the outflow. Both variables are highly correlated. However, this increase is very small, with the largest point estimate being 0.003 percentage points. Although the effects on migration are small, the second question is if these migration patterns are selective, because certain groups have better financial resources and job opportunities to choose their place of residence more freely. If migration outflow was positively selected while migration inflow was negatively selected, it could lead to an overestimation of the effect on student achievement. However, Figure A3.9 demonstrates that the aggregate adjusted gross income (AGI) of the out-migrating population remains unchanged after a disaster event. This suggests that out-migrating families are not more positively selected after a disaster. On the contrary, the increase in aggregate AGI indicates that the incoming households may be slightly more positively selected compared to those before a disaster.

Since high-SES students tend to perform better academically than their low-SES peers, this implies that if there is any impact, it may be an underestimation of the effect of natural disasters on student achievement.

Student composition: Next, I test if natural disasters shift the student composition. As depicted in Figure 3.11, overall enrollment, the share of Black students and the share of White students remain unchanged. This underscores that selective migration or a selective relocation of students is unlikely. There is a significant rise in the number of students receiving free or reduced lunches two to four years after the disaster event. Given the results on aggregate adjusted gross income for both in-migration and out-migration, this shift is unlikely to be driven by migration. Instead, it appears to be more closely related to a financial deterioration of the population following a natural disaster. Although not entirely clear, this suggests that financial distress within families could be a key channel through which natural disasters negatively impact student achievement in the long term.

Adult mental health: Long-run effects of early life exposure to natural disasters may be linked to a reduction in the mental and physical health stock (Karbownik and Wray, 2019). Unfortunately, I cannot directly test the effect on children's mental health due to data limitations. However, mental health of adults could serve as a critical channel. Adults experiencing distress, anxiety, or depression following a natural disaster may struggle to provide the necessary support and stability for children, at home, at school, and within the broader community, which could have long-term negative consequences for children's cognitive and socio-emotional development. Moreover, adults grappling with their own mental health may find it challenging to model positive coping strategies or maintain consistent communication with schools, hindering the implementation of effective interventions to support student well-being and academic success. However, Figure 3.12 shows no effect on the average number of mentally unhealthy days.

3.6 Discussion and Conclusion

This research unveils a persistent negative impact on student achievement following natural disasters, with students experiencing setbacks for up to five years post-event. These dynamics signal a depletion in the human capital reservoir of these regions, resulting in long-term economic damage (Gust et al., 2024). Leveraging data from the SEDA database spanning 2009 to 2018 and FEMA disaster declarations, I apply state-of-the-art difference-in-differences techniques by Sun and Abraham, 2021 to produce event study estimates that account for heterogeneous treatment effects in this staggered framework. Boys and low-SES children experience somewhat stronger effects. Alternative difference-in-differences methods by Borusyak et al., 2024 and Gardner, 2022, along with an alternative control group specification, confirm the detrimental effects on students. These findings are particularly concerning given the increasing frequency and severity of natural disasters. Except for the Covid-19 pandemic, natural disasters are the most frequent reason for prolonged unplanned school closures in the US.

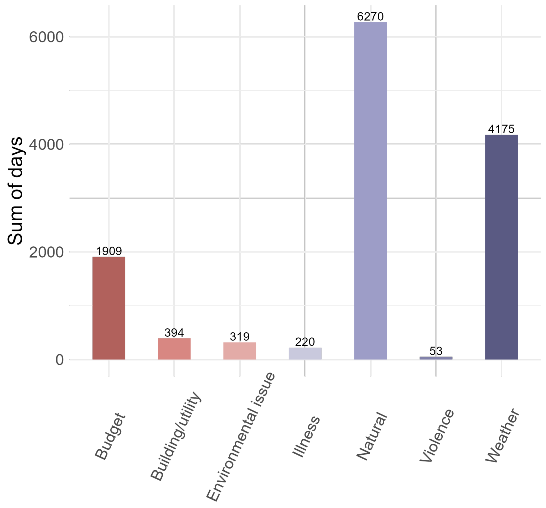
However, school closures represent just one facet through which natural disasters impact human capital. The observed negative effects on cohorts not yet enrolled during disasters suggest a depletion of health resources among younger children, damage of school buildings and facilities, lower financial resources of families and communities, and potential shifts in student and teacher composition of affected areas as contributing factors. While not all of these channels are directly testable, the increase in the share of students receiving free or reduced lunch suggests financial distress in families may be a key mechanism.

Significant divergences emerge when considering pre-disaster investment levels, with counties with higher per-pupil expenditure demonstrating swifter recovery compared to their lower-spending counterparts. However, countries with lower pre-disaster investment levels also have an ex-ante higher share of children receiving reduced or free lunch. Nevertheless, this finding holds important policy implications, indicating that augmenting per-pupil spending can enhance community resilience against human capital erosion caused by natural disasters. Future inquiries should delve into the specific types of investments pivotal in shielding against disaster-induced damage. Unfortunately, empirical evidence by Deryugina, 2017 suggests that governments typically curtail education spending in the aftermath of natural disasters, which could worsen the situation in future disasters.

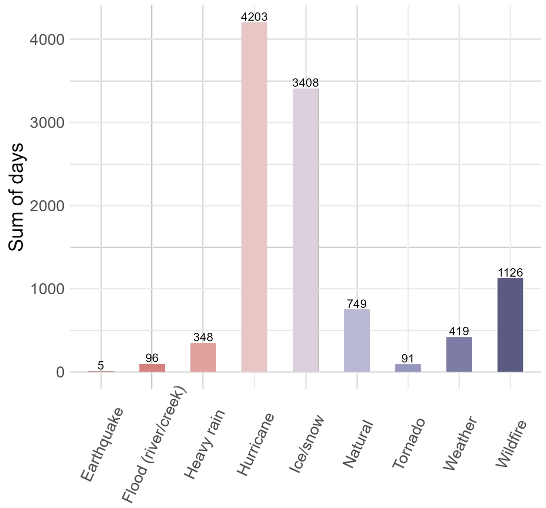
The results in this paper show that having multiple disasters in a row can cause more damage. Furthermore, the severity and nature of disasters play pivotal roles, with major events such as hurricanes exerting the most pronounced adverse impact on student performance. These findings underscore the imperative for proactive disaster preparedness and response measures, alongside targeted interventions.

Figures and Tables

Figure 3.1 : Days of school closures in the US between 2011 and 2019



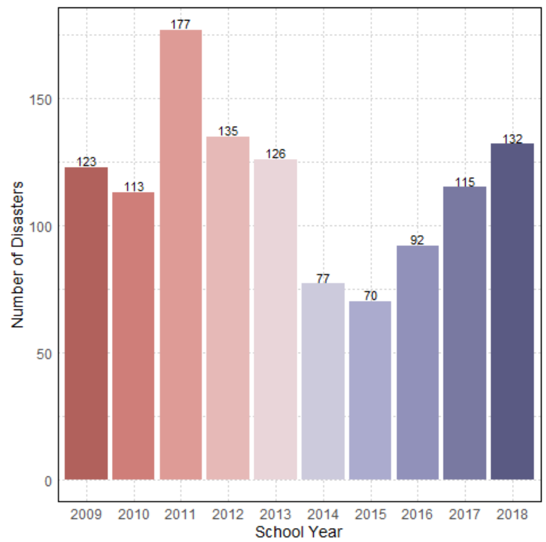
(a) Sum of school closure days overall



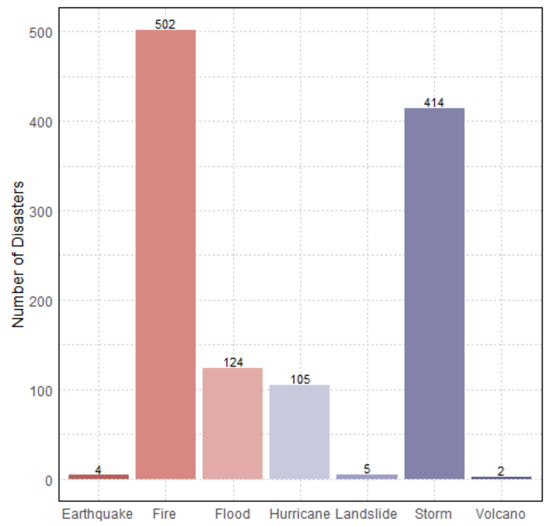
(b) Sum of school closure days due to natural disasters and extreme weather

Notes: The figure shows unplanned school closure days in the US by general causes on panel a) and by type of natural disaster in Panel b). Jahan et al., 2022 conducted daily systematic online searches to collect data on publicly announced unplanned school closures lasting at least one school days in the United States from August 1, 2011, through June 30, 2019.

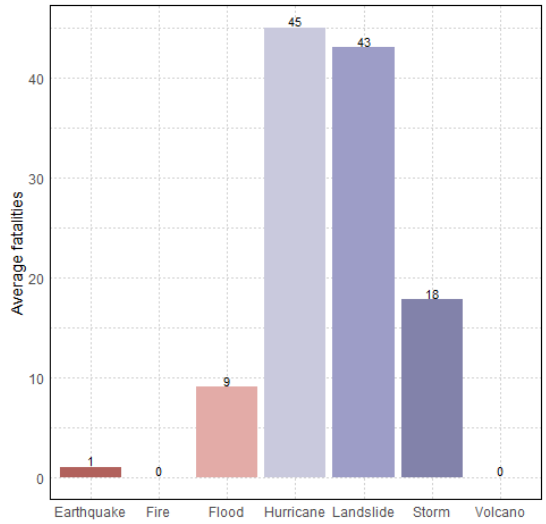
Figure 3.2 : Summary statistics: FEMA disaster declarations



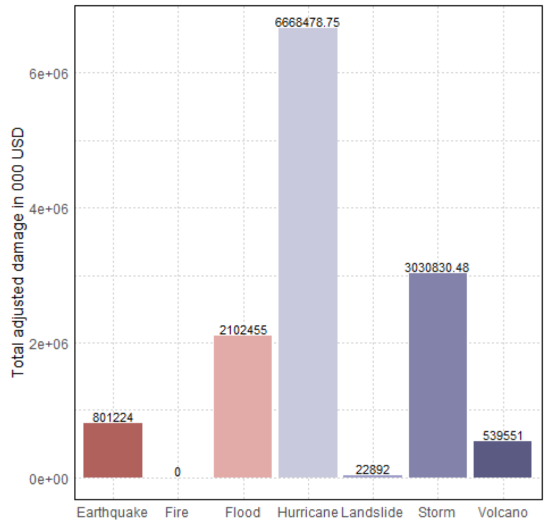
(a) Disasters by year



(b) Disasters by type



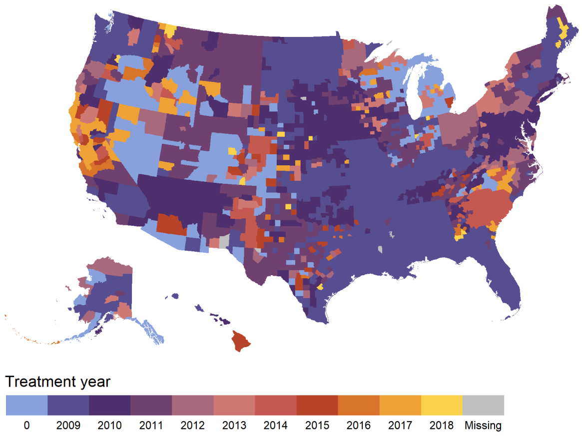
(c) Average fatalities



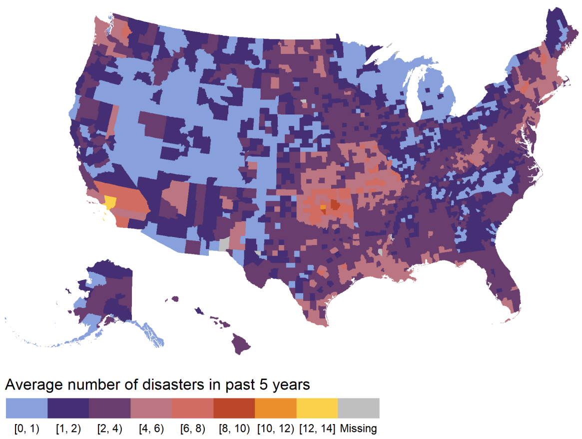
(d) Average costs

Notes: Panel a) shows total events declared as disaster in FEMA. Every disaster is counted only once even when declared in multiple locations. Panel b) shows number of disasters by disaster type in FEMA between 2009 and 2018. Panel c) shows average disaster fatalities if information on fatalities exists. Panel d) shows average value of the damage in thousands of US dollars at the moment of the event adjusted for inflation using CPI. Note that EM-DAT includes only disasters with at least 10 fatalities, 100 affected or if a country called for international assistance or an emergency declaration.

Figure 3.3 : Map: First year of disaster and number of disasters



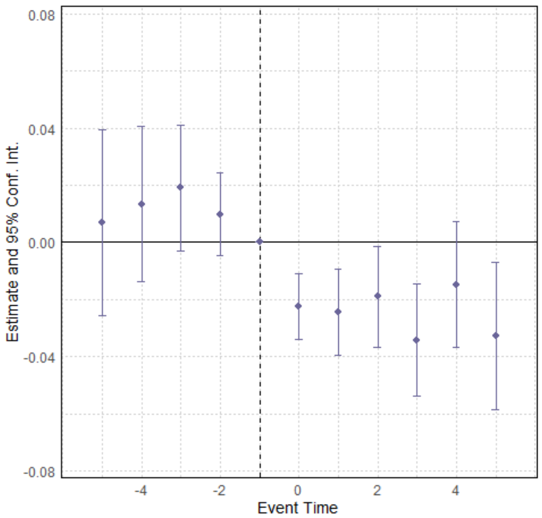
(a) First disaster event



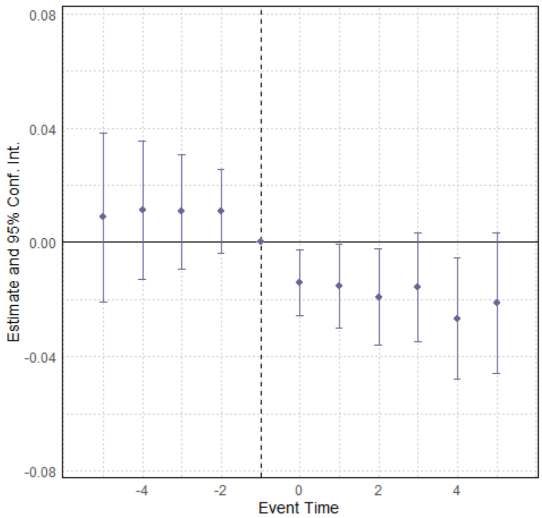
(b) Average number of disasters in past five years

Notes: The figure shows the first year of a FEMA disaster declaration between 2009 and 2018 for the Sun and Abraham, 2021 estimations and the average number of FEMA disaster declarations in the past five years.

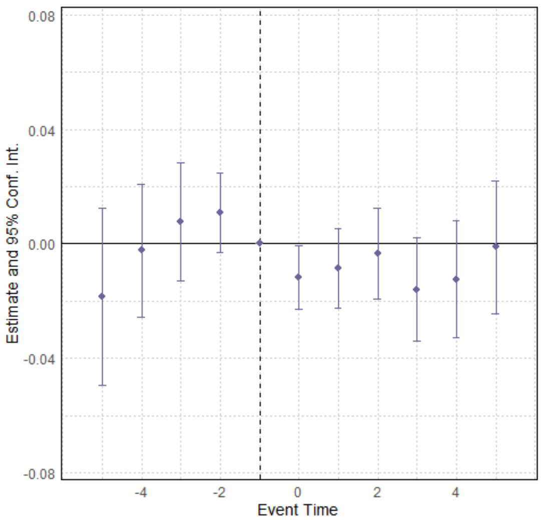
Figure 3.4 : Event study: Math achievement



(a) Math achievement in grade three



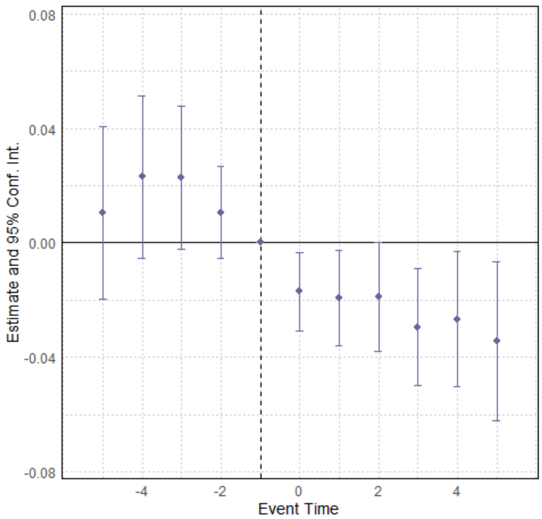
(b) Math achievement in grade four



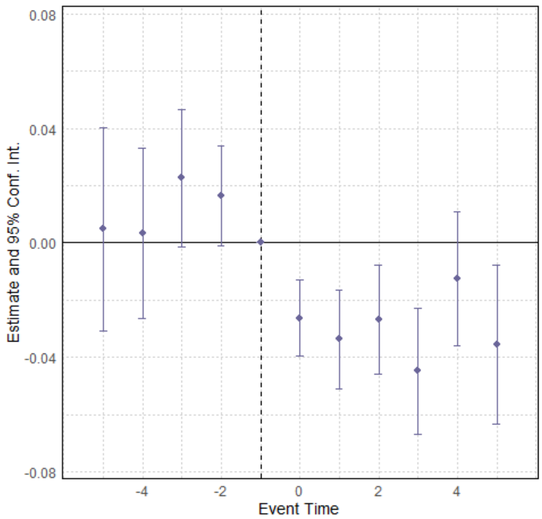
(c) Math achievement in grade five

Notes: The figure shows the main results from the event study analysis following Sun and Abraham, 2021 for grade three, four, and five. The x-axis represents years relative to the first natural disaster between 2009 and 2018. The y-axis represents the estimate with 95% confidence intervals. The figure shows a decrease in math achievement after the natural disaster.

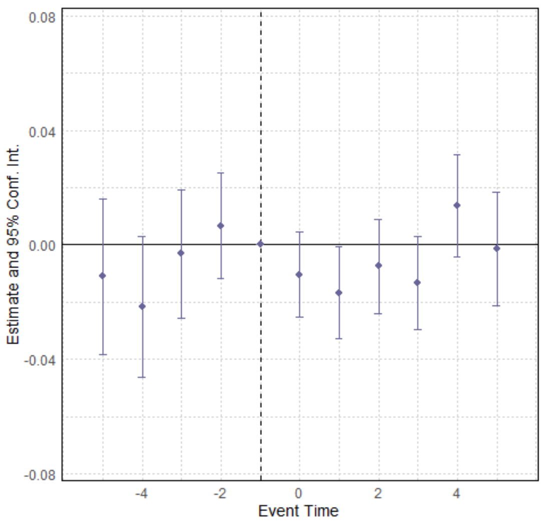
Figure 3.5 : Event study: Gender achievement gap



(a) Achievement of girls



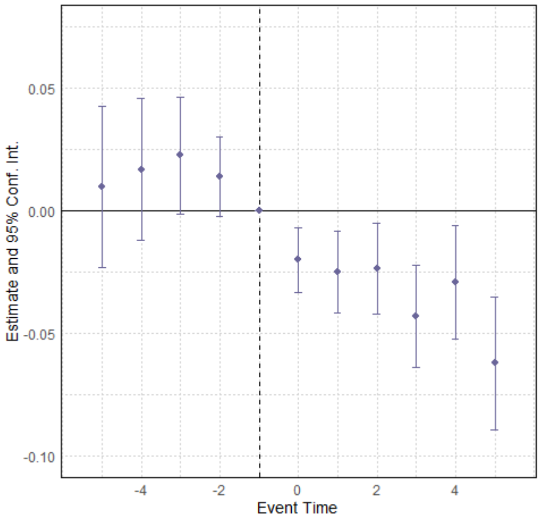
(b) Achievement of boys



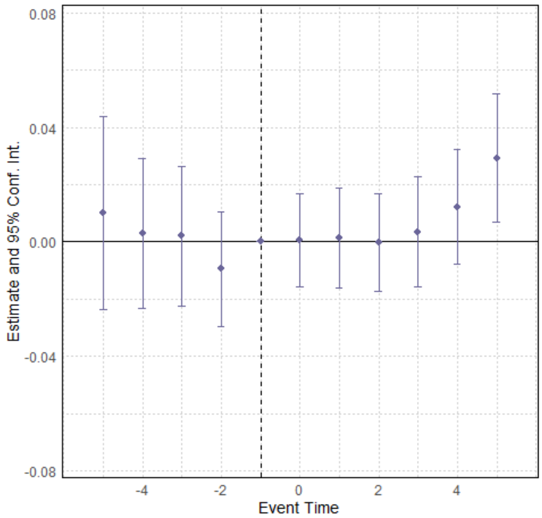
(c) Male-female gap

Notes: The figure shows the main results from the event study analysis following Sun and Abraham, 2021 for the male-female gap in grade three math achievement. The x-axis represents years relative to the first natural disaster between 2009 and 2018. The y-axis represents the estimate with 95% confidence intervals.

Figure 3.6 : Event study: Socio-economic gap



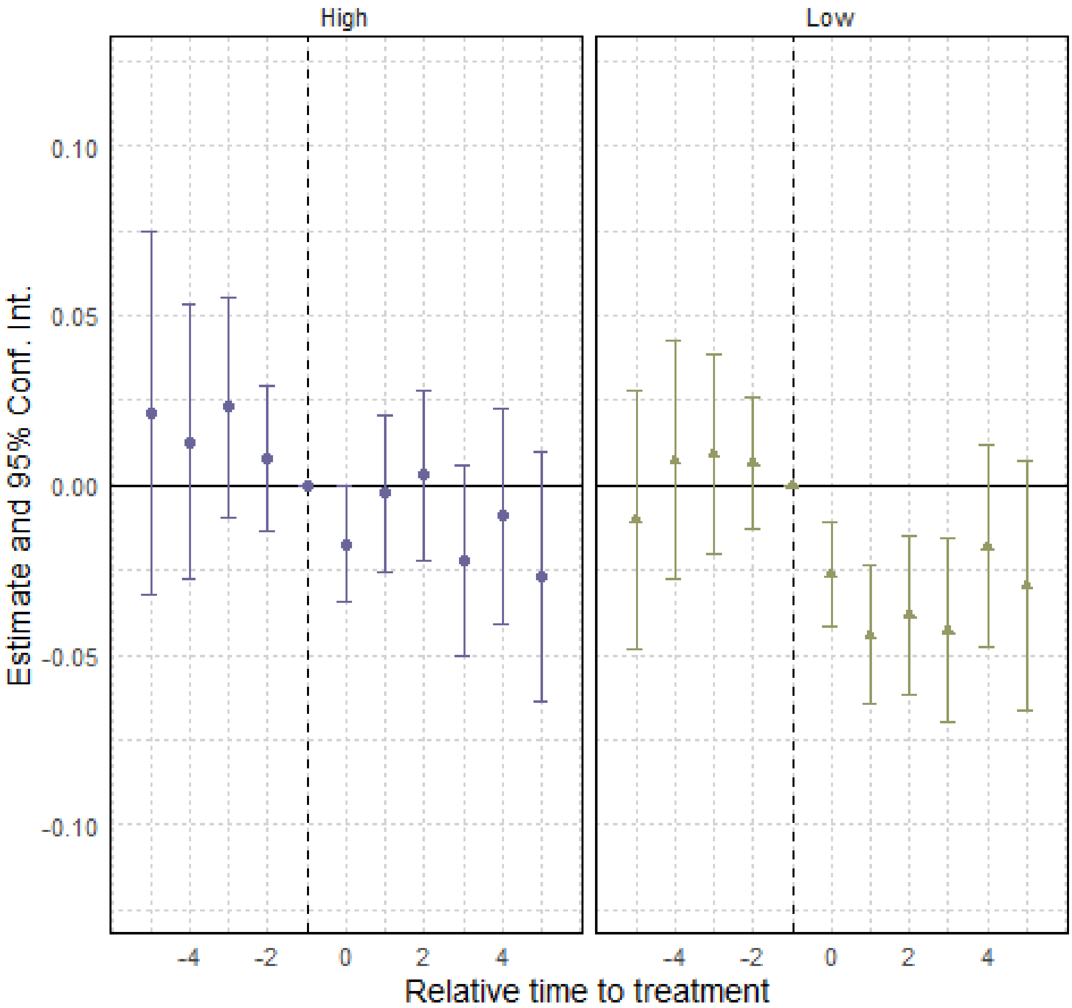
(a) Achievement of econ. disadvantaged



(b) Socio-economic gap

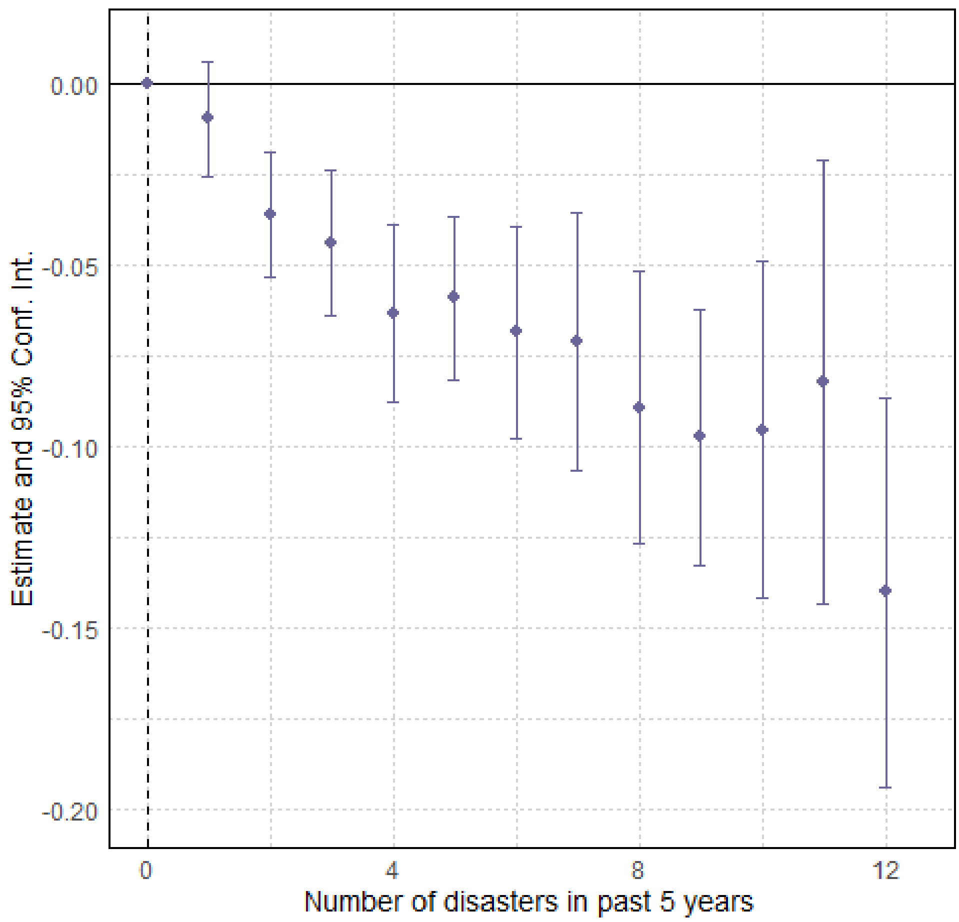
Notes: The figure shows the main results from the event study analysis following Sun and Abraham, 2021 for the socio-economic gap in grade three math achievement. The x-axis represents years relative to the first natural disaster between 2009 and 2018. The y-axis represents the estimate with 95% confidence intervals.

Figure 3.7 : Heterogeneity by per-pupil spending



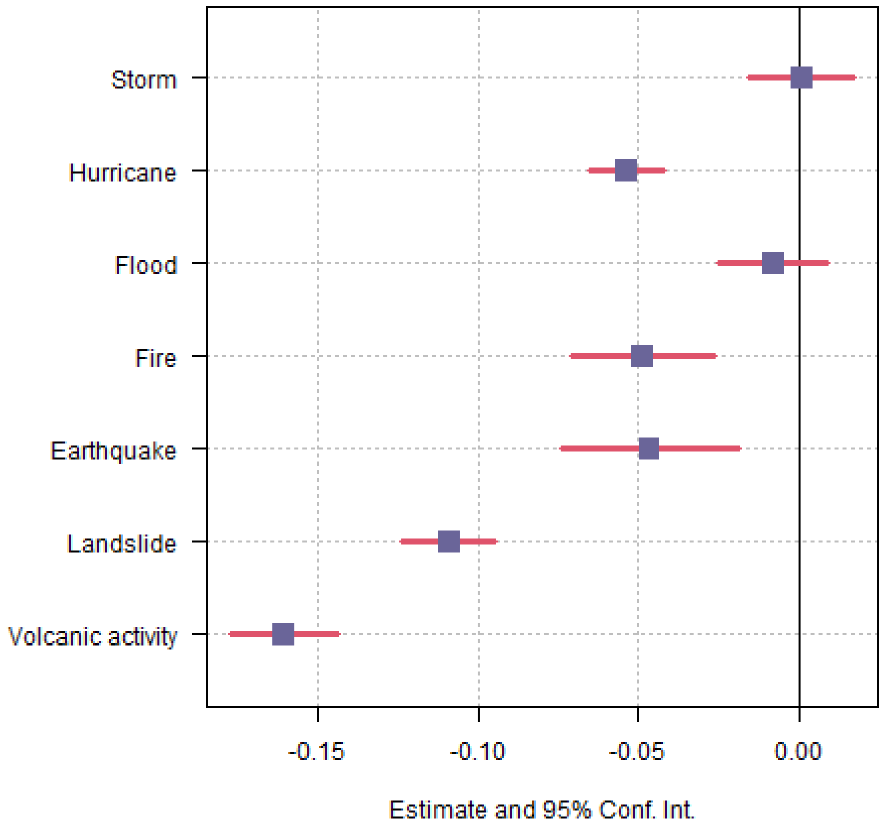
Notes: The figure shows the main results from the event study analysis following Sun and Abraham, 2021 for math achievement by high and low per-pupil spending. The x-axis represents years relative to the first natural disaster between 2009 and 2018. The y-axis represents the estimate with 95% confidence intervals.

Figure 3.8 : Heterogeneity by number of disasters



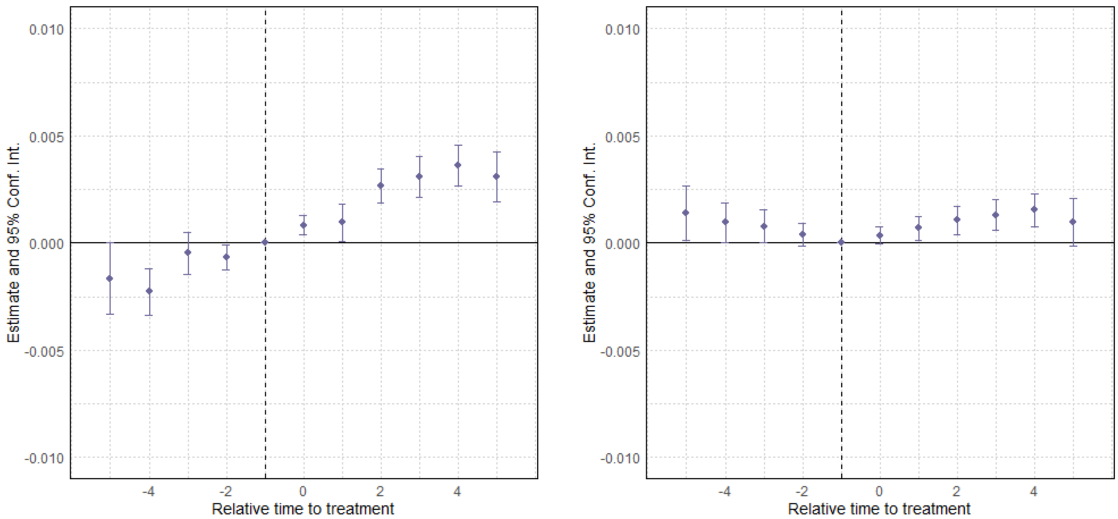
Notes: The figure shows the results from estimating equation 3.6 that contains county and school-year fixed effects, population weights and a dummy for each disaster count over the past five years.

Figure 3.9 : Heterogeneity by disaster type



Notes: The figure shows the results from estimating equation 3.6 that contains county and school-year fixed effects, population weights and a dummy for whether a disaster type occurred in the past five years.

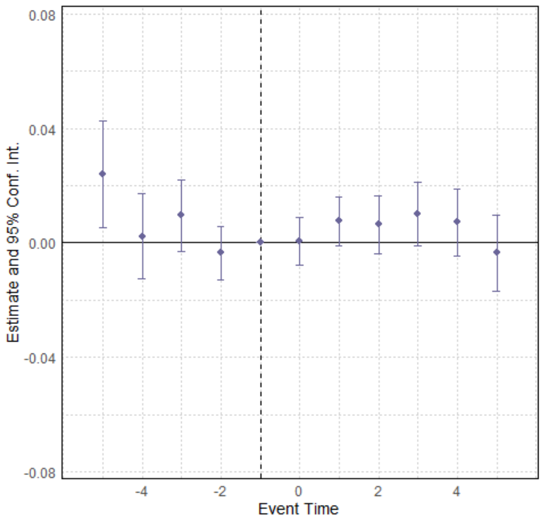
Figure 3.10 : Event study: Migration



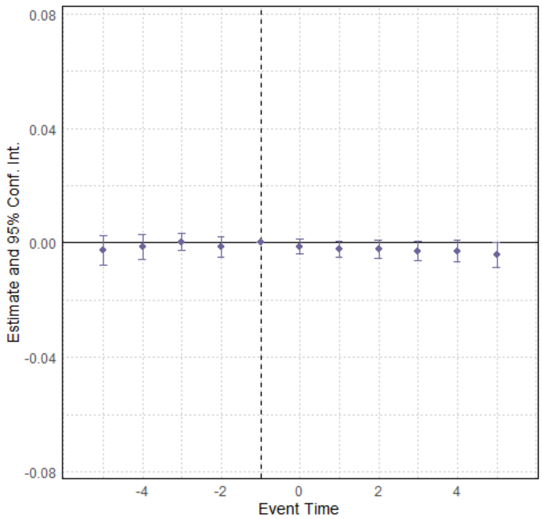
(a) Migration inflow (b) Migration outflow

Notes: The figure shows the results from the event study analysis following Sun and Abraham, 2021 for migration inflow and migration outflow. The x-axis represents years relative to the first natural disaster between 2010 to 2018. The y-axis represents the estimate with 95 confidence intervals.

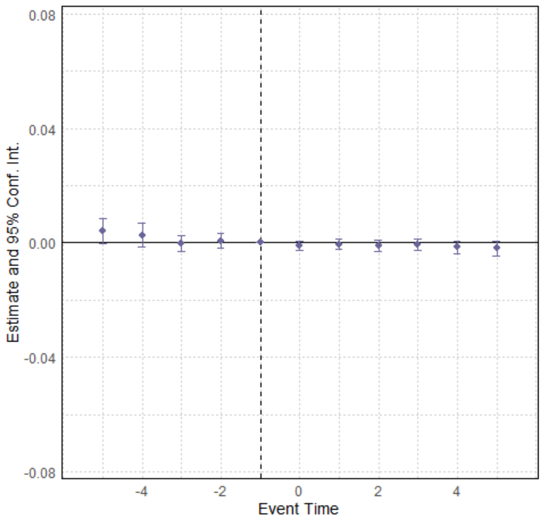
Figure 3.11 : Event study: Enrollment and student composition



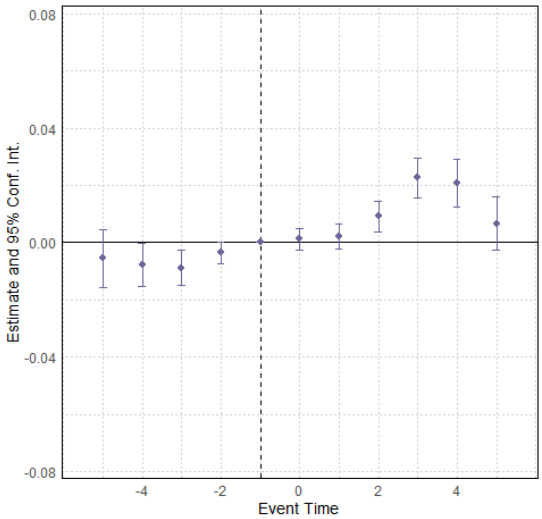
(a) Enrollment



(b) Black population



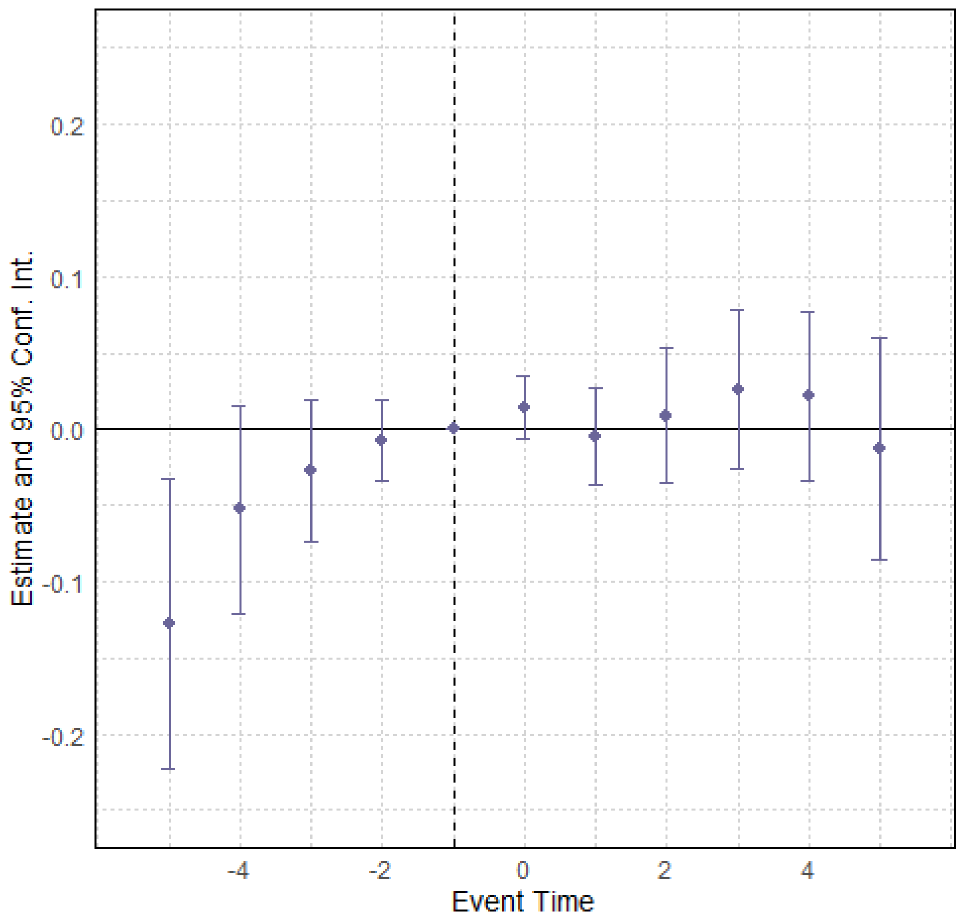
(c) White population



(d) Free or reduced lunch

Notes: The figure shows the results from the event study analysis following Sun and Abraham, 2021 for enrollment, share of White students, the share of Black students, and share of students receiving free or reduced lunch in grade three. The x-axis represents years relative to the first natural disaster between 2010 to 2018. The y-axis represents the estimate with 95% confidence intervals.

Figure 3.12 : Event study: Adult mental health



Notes: The figure shows the results from the event study analysis following Sun and Abraham, 2021 for mental health. The x-axis represents years relative to the first natural disaster between 2009 and 2018. The y-axis represents the estimate with 95% confidence intervals.

Table 3.1 : Summary statistics SEDA data

Statistic	N	Mean	Median	St. Dev.	Min	Max
Math grade 3	15,930	−0.016	−0.005	0.298	−1.513	1.120
Math grade 4	15,930	−0.047	−0.030	0.301	−1.621	1.052
Math grade 5	15,930	−0.064	−0.045	0.299	−1.589	1.075
Socio-economic gap	15,663	−0.229	−0.222	0.265	−1.539	0.778
Male-female gap	15,624	−0.023	−0.012	0.285	−1.496	1.053
Percent Black Students	15,930	0.134	0.032	0.206	0.000	1.000
Percent White Students	15,930	0.694	0.763	0.264	0.000	1.000
Percent free/reduced lunch	15,930	0.596	0.594	0.170	0.087	1.000
Log enrollment	15,930	5.969	5.814	1.230	2.890	11.158
Total per pupil expenditure	15,860	11,524.460	10,936.790	2,519.827	4,920.042	30,126.450
Total per pupil revenue	15,862	11,571.940	11,081.720	2,355.729	5,305.988	28,087.980
Population estimate 2008	15,930	82,383.660	27,729	233,250.100	1,469	5,256,705
Inflow	21,942	0.050	0.046	0.022	0.007	0.703
Outflow	21,942	0.049	0.046	0.019	0.012	0.419
Mental health	23,013	3.614	3.600	0.872	0.630	8.300

Note: Summary statistics for the county level data from the Stanford Educational Archive 4.1 for 2009 to 2018, county level in- and out-migration from the Statistics of Income Division migration data for 2010 to 2018, and average number of poor mental health days from CHR&R for 2010 to 2018.

Table 3.2 : Static effect on student achievement

	Math score grade 3		
	Model 1	Model 2	Model 3
	(1)	(2)	(3)
Disaster	-0.0206***	-0.0284***	-0.0141**
	(0.0054)	(0.0080)	(0.0067)
Population-weighted		✓	✓
Observations	15,930	15,930	15,930
R^2	0.76518	0.87385	0.90586
Within R^2	0.00153	0.00693	0.00151
County fixed effects	✓	✓	✓
School year fixed effects	✓	✓	
State-school year fixed effects			✓

Note: The table presents the results from estimating the TWFE model in equation 3.4 for math achievement in grade three. Model 1 contains only county and school-year fixed effects and no weights. Model 2 contains county and school-year fixed effects and is weigthed by the ex-ante county population size. Model 3 contains county and state times school year fixed effects and population weights. Clustered (county) standard-errors in parentheses. Signif. Codes: ***: 0.01, **: 0.05, *: 0.11

Table 3.3 : Heterogeneity by severity of disasters

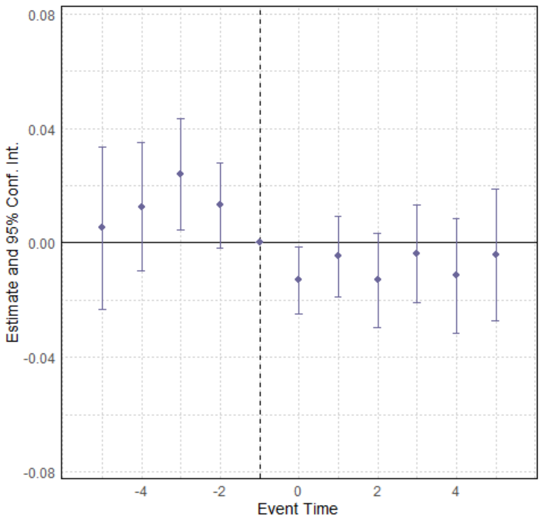
	Math score grade 3		
	By fatalities	By damage	By duration
	(1)	(2)	(3)
Minor disaster	-0.0274***	-0.0251***	-0.0247***
	(0.0079)	(0.0080)	(0.0077)
Major disaster	-0.0377***	-0.0423***	-0.0680***
	(0.0137)	(0.0115)	(0.0131)
Observations	15,930	15,930	15,930
R^2	0.87391	0.87422	0.87516
Within R^2	0.00738	0.00980	0.01723
County fixed effects	✓	✓	✓
School year fixed effects	✓	✓	✓

Note: The table presents the results from estimating the TWFE model in equation 3.5 with county and school year fixed effects for math achievement in grade three. Major disasters are those that caused more than 25 deaths (column 1), more than 1 billion dollar (adjusted) total damage (column 2), or that lasted more than 50 business days (column 3). All models contain population weights. Clustered (county) standard-errors in parentheses. Signif. Codes: ***: 0.01, **: 0.05, *: 0.11

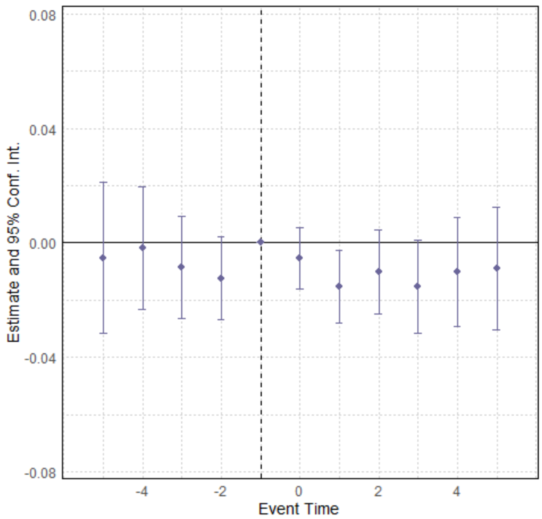
Appendix

A3.7 Appendix Figures and Tables

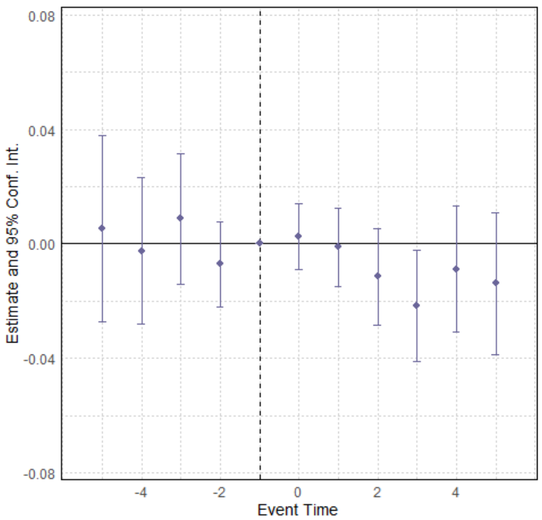
Figure A3.1 : Event study: Math achievement in higher grades



(a) Math achievement in grade six



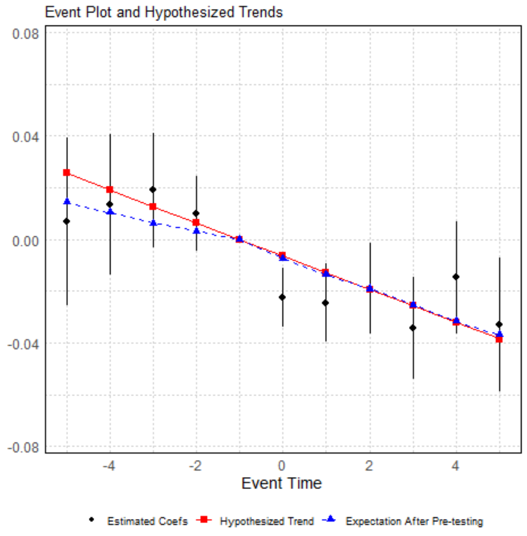
(b) Math achievement in grade seven



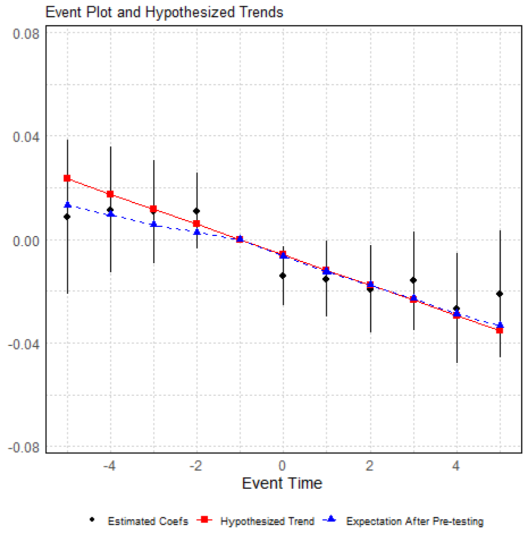
(c) Math achievement in grade eight

Notes: The figure shows the results from the event study analysis following Sun and Abraham, 2021 for grade six, seven, and eight. The x-axis represents years relative to the first natural disaster between 2009 and 2018. The y-axis represents the estimate with 95 confidence intervals. The figure shows a decrease in math achievement after the natural disaster.

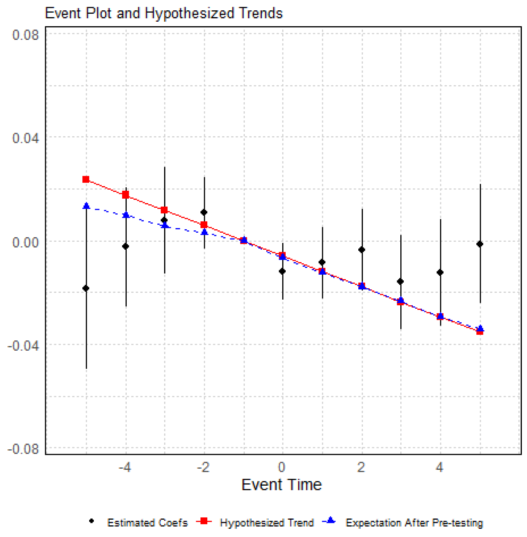
Figure A3.2 : Pre-trend diagnostics: Math achievement



(a) Pretest grade three



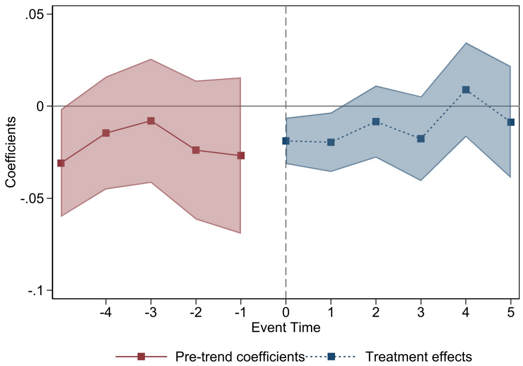
(b) Pretest grade four



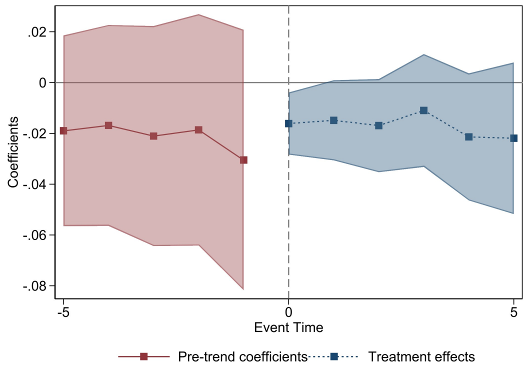
(c) Pretest grade five

Notes: The plots show the pre-trend diagnostics by Roth, 2022 for the Sun and Abraham, 2021 estimates. The red, solid line is the a linear violation of parallel trends that a pre-trend test would detect with 50 percent power. The dashed blue the expected coefficients conditional on not finding a significant pre-trend if the true population means were the hypothesized red, solid line.

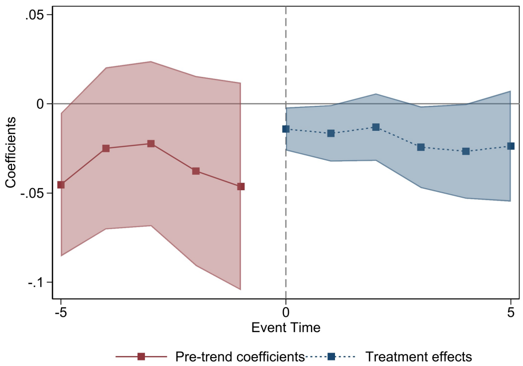
Figure A3.3 : Event Study: Math achievement with imputation method



(a) Grade three, imputation method



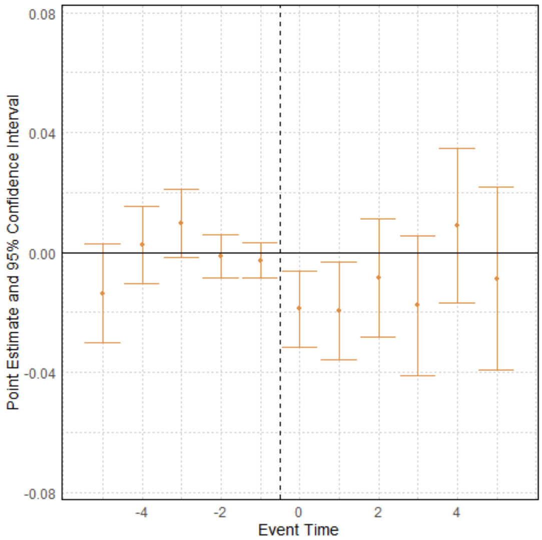
(b) Grade four, imputation method



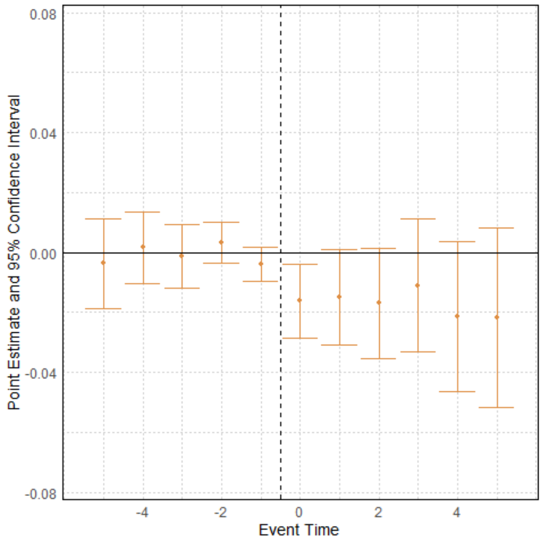
(c) Grade five, imputation method

Notes: Event study analysis following the imputation method by Borusyak et al., 2024 for grade three, four, and five. The x-axis represents years relative to the first natural disaster. The y-axis represents the estimate with 95% confidence intervals.

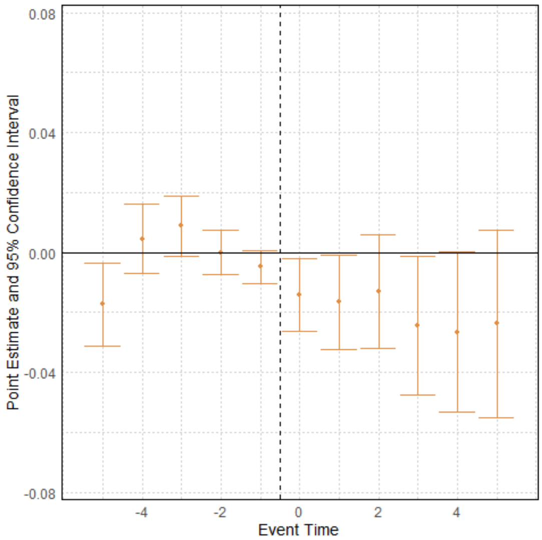
Figure A3.4 : Event Study: Math achievement with two-stage difference-in-differences



(a) Grade three, two-stage difference-in-differences



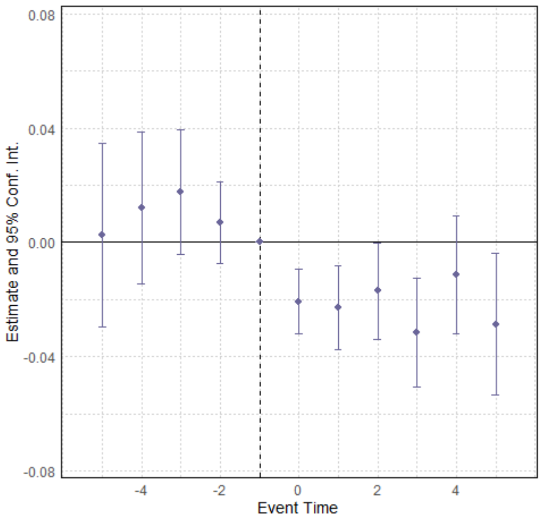
(b) Grade four, two-stage difference-in-differences



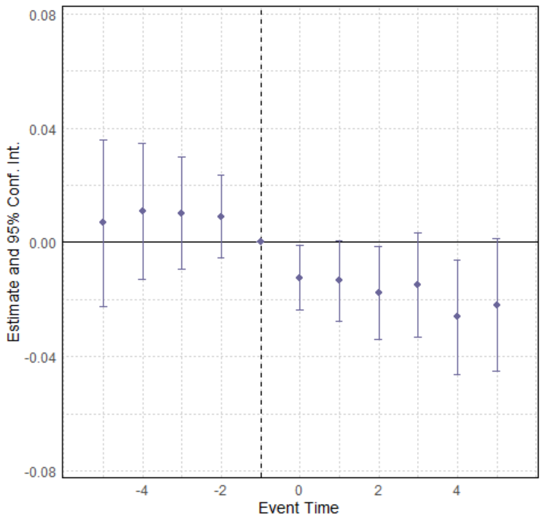
(c) Grade five, two-stage difference-in-differences

Notes: Event study analysis following the two-stage DiD method by Gardner, 2022 for grade three, four, and five. The x-axis represents years relative to the first natural disaster. The y-axis represents the estimate with 95% confidence intervals.

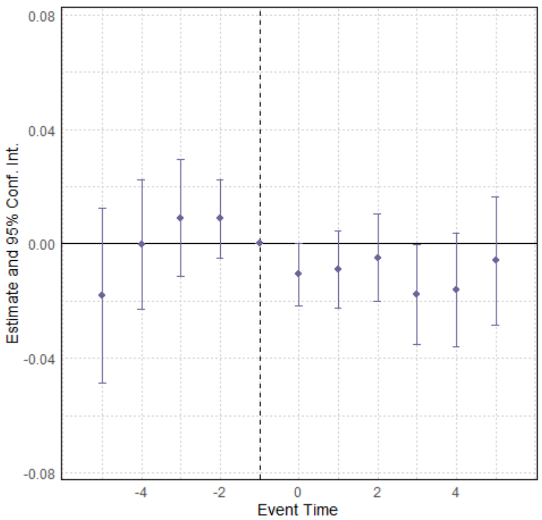
Figure A3.5 : Event study: Math achievement never-treated as control



(a) Math achievement in grade three



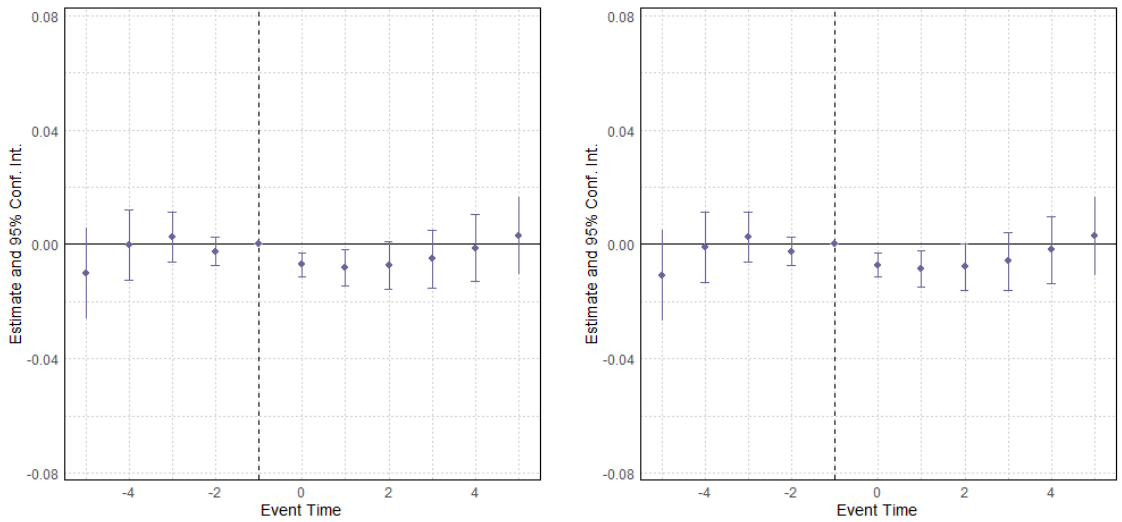
(b) Math achievement in grade four



(c) Math achievement in grade five

Notes: The figure shows the results from the event study analysis following Sun and Abraham, 2021 for grade three, four, and five with never-treated counties as control group. The x-axis represents years relative to the first natural disaster between 2009 and 2018. The y-axis represents the estimate with 95% confidence intervals. The figure shows a decrease in math achievement after the natural disaster.

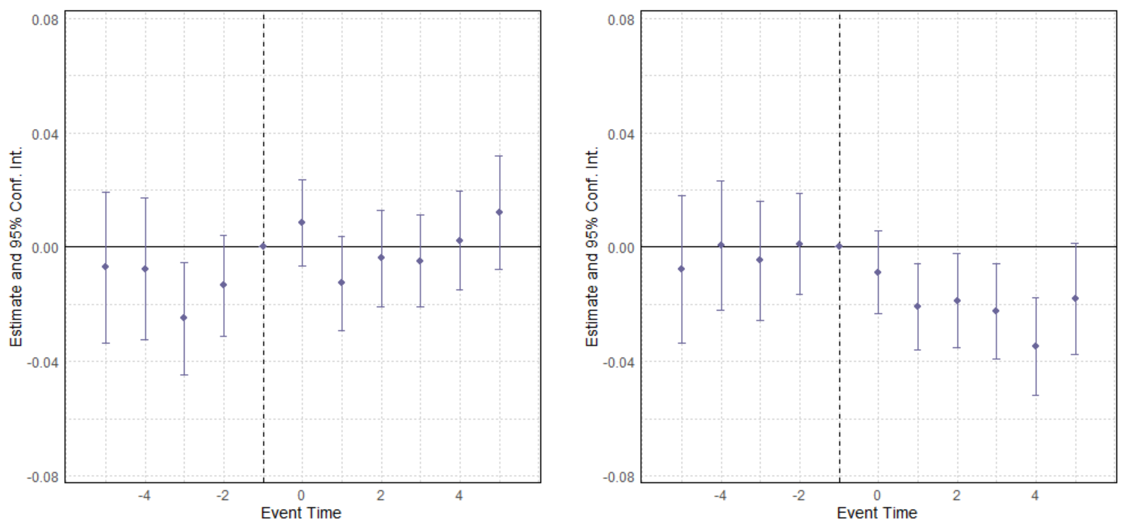
Figure A3.6 : Event study: Math achievement with pooled grades and cohort fixed effects



(a) Math achievement pooled (b) Math achievement with cohort fixed effects

Notes: The figure shows the results from the event study analysis following Sun and Abraham, 2021 for grade three to eight a) pooled and b) including cohort fixed effects. The x-axis represents years relative to the first natural disaster between 2009 and 2018. The y-axis represents the estimate with 95% confidence intervals. The figure shows a decrease in math achievement after the natural disaster.

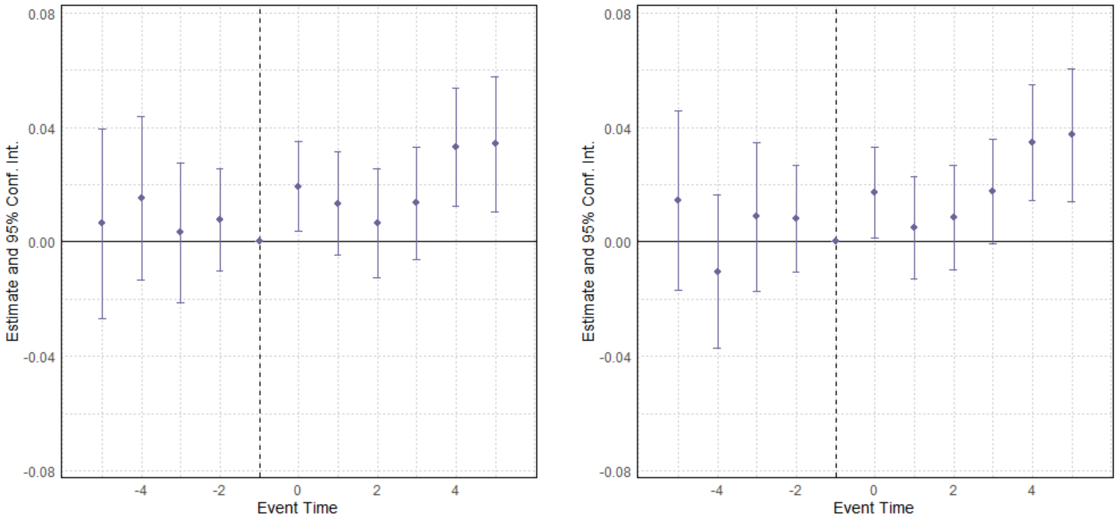
Figure A3.7 : Event study: Male-female achievement gap in grade four and five



(a) Male-female gap grade four (b) Male-female gap grade five

Notes: The figure shows the main results from the event study analysis following Sun and Abraham, 2021 for the male-female gap in grade four and five math achievement. The x-axis represents years relative to the first natural disaster between 2009 and 2018. The y-axis represents the estimate with 95% confidence intervals.

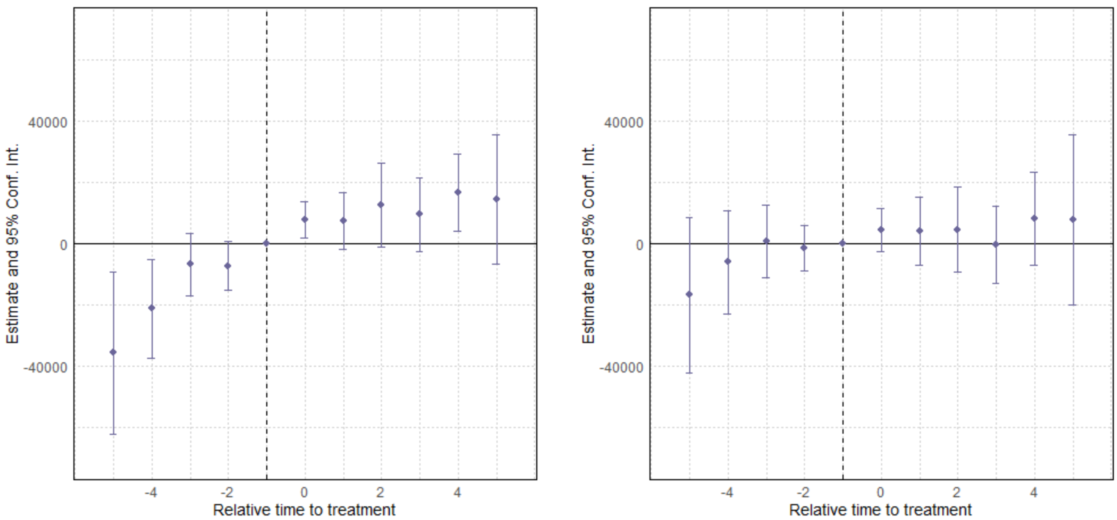
Figure A3.8 : Event study: Socio-economic achievement gap in grade four and five



(a) Socio-economic gap grade four (b) Socio-economic gap grade five

Notes: The figure shows the main results from the event study analysis following Sun and Abraham, 2021 for the socio-economic gap in grade four and five math achievement. The x-axis represents years relative to the first natural disaster between 2009 and 2018. The y-axis represents the estimate with 95% confidence intervals.

Figure A3.9 : Event study: Aggregate adjusted gross income



(a) Aggregate adjusted gross income inflow (b) Aggregate adjusted gross income outflow

Notes: The figure shows the main results from the event study analysis following Sun and Abraham, 2021 for the aggregate adjusted gross income. The x-axis represents years relative to the first natural disaster between 2009 and 2018. The y-axis represents the estimate with 95% confidence intervals.

Table A3.1 : Balancing table: High and low per-pupil spending

	High (N=659)		Low (N=928)			
	Mean	Std. Dev.	Mean	Std. Dev.	Diff. in Means	Std. Error
Math grade 3	−0.024	0.308	−0.084	0.272	−0.060	0.015
Percent Black students	0.133	0.211	0.144	0.209	0.010	0.011
Percent White students	0.681	0.273	0.733	0.251	0.052	0.013
Percent free/reduced lunch	0.511	0.177	0.564	0.151	0.053	0.008
Log enrollment	6.069	1.427	5.934	1.037	−0.135	0.065

Note: The table shows the mean and the standard deviation by high and low per-pupil spending, the difference in means, and the standard error.

Table A3.2 : Pre-trend test results

Metric	Grade 3	Grade 4	Grade 5
Power	50	50	50
Hypothesized trend	0.006	0.006	0.006
Bayes factor	0.585	0.586	0.235
Likelihood ratio	1.038	1.105	0.526
Power	80	80	80
Hypothesized trend	0.010	0.009	0.009
Bayes factor	0.234	0.234	0.235
Likelihood ratio	0.336	0.372	0.526

Notes: Summary statistics of power, slope and bias calculation (Roth, 2022). The table shows the probability that we would find a significant pre-trend (set to 50% or 80%), the slope of the differential trend that we would be able to detect with that power, the ratio of the probability of "passing" the pre-test under the hypothesized trend relative to under parallel trends (a smaller Bayes factor favors parallel trends over the hypothesized trend when the pre-trend is insignificant), the likelihood ratio of the observed coefficients under the hypothesized trend relative to under parallel trends.

4 Does Civic Education Foster Civic Engagement?[*]

In recent years, democracies worldwide face increasing challenges to their legitimacy and stability. Voter turnout has declined over the last decades (Kostelka and Blais, 2021). Political disillusionment among younger generations is prevalent (Kitanova, 2020) – with exceptions like Fridays for Future and the Last Generation movements. There has been a rise in support for populist parties in several European countries (Noury and Roland, 2020), and a resurgence of racism and antisemitism (Huneke, 2024). The COVID-19 pandemic further exacerbated political distrust, and the proliferation of fake news and hate speech, especially online, has targeted both political and civil society elites (Kalsnes, 2023). Amidst these multifaceted crises, often referred to as a "polycrisis" (Lawrence et al., 2024), there is a desperate search for solutions. Civic education in schools is frequently highlighted as a potential remedy (Manning and Edwards, 2014a; Hamm et al., 2023).[1]

Our study contributes to this discourse by examining the long-term impact of civic education in schools on students' civic engagement later in life. Specifically, we investigate the effects of introducing mandatory civic education in schools on civic engagement in adulthood, examining the impact of civic education at the extensive margin. Additionally, we explore whether an increase in the number of hours dedicated to civic education influences civic engagement, thus examining the impact of civic education at the intensive margin, a topic often addressed in political debate.

We leverage the distinctive German context characterized by significant post-war variations across states and school tracks due to reforms affecting the presence and intensity of civic education in the classroom. In the German federal education system, the education ministers of each state are responsible for implementing education reforms, including any changes to school hour schedules. These schedules specify the weekly hours allocated to various learning areas and subjects for each type of school.

We collect a unique new dataset which encompasses the weekly hours allocated to civic education across all German states and the three secondary school tracks since 1976. We restrict the analysis to the 10 West German states. East and West Germany had fundamentally

---

* This chapter is co-authored with Sven Resnjanskij, Larissa Zierow, Marcel Helbig, and Norbert Sendzik. It is based on the paper 'Does Civic Education Foster Civic Engagement?'. The project was supported by the Leibniz Competition (SAW 2019).

1 The role of schools in fostering civic engagement among future generations has long been emphasized by political leaders. For instance, Willy Brandt famously remarked, "Die Schule der Nation ist die Schule" (The school of the nation is the school), and Abraham Lincoln noted, "Teach the children so it won't be necessary to teach the adults."

different political, economic, and educational systems due to their division until 1990, which could otherwise confound our results.

We link the curricular records to individual level outcomes from the German Socio-Economic Panel (SOEP) based on state, school track, and time of secondary schooling in childhood. Our main variable of interest is an index of civic engagement that Hener et al. (2016) constructed specifically for the SOEP. Our final data set enables us to analyze the impact of civic education across multiple decades of school cohorts and investigate long-term outcomes.

Descriptively, we show that civic engagement increases with age, conditional on cohort and without controlling for civic education. Additionally, it positively correlates with higher education, being male, and having a high socio-economic status. These patterns are in line with existing literature (Putnam, 2000; Dee, 2004).

Our causal approach is based on the reform-induced variation in civic education hours, contingent upon state, school-entry year, and attended school track. Employing the new continuous difference-in-differences methodology developed by De Chaisemartin and d'Haultfoeuille (2024), we identify effects from differences in adult outcomes between cohorts within the same school track in states that experienced alterations in civic education hours, relative to cohorts in state and school track combinations that were unaffected by changes in civic education hours.

Our analysis shows that introducing civic education as a subject (at the extensive margin) has a positive impact on civic engagement. However, findings regarding the intensive margin, namely the impact of increased average instructional hours, are less definitive, showing negligible effects on average. Our heterogeneity analysis underscores that individuals from lower socio-economic backgrounds tend to derive greater benefits from civic education at the extensive margin than individuals from higher socio-economic backgrounds.

Our study contributes to the following strands of literature. In the field of political economy and education science, empirical investigations have yielded mixed findings on the link between educational interventions and subsequent political attitudes, with some studies revealing limited effects on attitudes despite increases in knowledge (Green et al., 2011). The existing studies on this question either do not allow making causal claims (e.g., Callahan et al., 2008; Campbell and Niemi, 2016; Hoskins et al., 2016; Neundorf et al., 2016; Borge, 2017; Reichert and Print, 2018; Stadelmann-Steffen and Sulzer, 2018; Nelsen, 2021; Sampermans et al., 2021; Alscher et al., 2022; Otache et al., 2023; Rinnooy Kan et al., 2023), estimate the effect of general rather than specialized civic education (e.g., Dee, 2004; Milligan et al., 2004; Siedler, 2010; Sondheimer and Green, 2010; Persson et al., 2016; Meyer, 2017; Lindgren et al., 2019; Ahlskog, 2021; Apfeld et al., 2022; Harka and Rocco, 2022), or focus on the general classroom environment or democratic educational aspects of other subjects (e.g., Persson et al., 2020; Hoskins et al., 2021; Deimel et al., 2024). The studies that causally identify effects of educational intervention on civic engagement do not estimate the effect of civic education as its own subject in school, but focus on interventions that are small enough to be analyzed in randomized controlled trials like community programs, mock elections, teacher training, or information provision (Syvertsen et al., 2009; Donbavand and Hoskins, 2021; Kalla and Porter, 2022). One exception is Jung and Gopalan (2023) who use nationally representative data to show that adopting civics tests for high school graduation does not affect youth voter turnout. While their study focuses on the effects of civics tests, our study specifically explores the influence of civic education as a standalone subject, providing insights into its unique effects on civic engagement later in life. Unlike previous studies, we study large-scale civic education programs and provide causal insights into their impact.

A broad literature in the economics of education studies the effect of different school reforms. While this research has traditionally focused on academic outcomes, such as academic achievement and labor-market success (e.g., Hanushek, 1986; Woessmann, 2016b, our study extends the scope to encompass non-academic outcomes. This expansion of the analytical lens is motivated by recent contributions highlighting the importance of non-academic outcomes in educational policy, such as personality traits (Almlund et al., 2011) or soft skills (Koch et al., 2015). Arold et al., 2022 investigate the impact of school reforms on religious attitudes and Arold, 2024 analyzes how changes in school curricula influence beliefs in evolution.

The remainder of the paper proceeds as follows. In the second chapter, we present theoretical considerations regarding how civic education taught in schools could impact civic engagement later in life. The third chapter provides a historical and institutional background. In the fourth section, we present the data. In the fifth chapter, we outline our empirical model, followed by the presentation of our results in the sixth section. In the seventh section, we conclude our study.

A3.1 Conceptual Framework

Civic engagement is a repertoire of different activities aiming to defend interests, express opinions, influence decisions of authorities (Theocharis and Van Deth, 2018), and contribute to society (Putnam, 2000). Political attitudes and actions are critical components of civic engagement. Political attitudes generally include political interest, trust, and the feeling of political efficacy, as well as support for norms and values that foster democratic behavior (Quintelier and Van Deth, 2014; Weiss, 2020). These attitudes are often measured by indicators such as satisfaction with democracy (Dompnier and Berton, 2012; Claassen et al., 2023). Political actions, viewed through the lens of democratic theory in established democracies, are often referred to as political participation. It includes activities undertaken by citizens to influence political decisions (Verba et al., 1995; Quintelier and Van Deth, 2014). These general definitions underscore the normative core of political socialization, emphasizing the importance of both attitudes and actions in shaping civic engagement.

Adolescence, often referred to as the "impressionable years", is a critical phase for political socialization (Neundorf et al., 2016). This phase coincides with significant cognitive development, particularly as the frontal lobe of the brain undergoes intense transformation and thus presents an opportunity for societal interventions, as adolescents begin to exert some agency in their decision-making processes (Hoxby, 2021). Imprinting theory highlights that experiences during sensitive developmental stages can have lasting effects and that learning experiences activated by the environment during critical developmental periods lead to long-lasting outcomes (Marquis and Tilcsik, 2013).

Political attitudes and actions can be shaped by various factors during adolescence. Understanding the role of school-based civic education in political socialization requires distinguishing among these factors. Deimel (2023) uses Bronfenbrenner's (1981) socio-ecological model to differentiate between relevant environments and mechanisms: the family environment, the school environment, interactions between family and school environments, and broader societal influences (for a more in-depth discussion, see Sendzik et al., 2024). We will use this theoretical framework to explore the potential mechanisms at play when analyzing the impact of civic education taught in schools on civic engagement in later adult life.

1. *Family Environment*: The family is the primary instance for direct and indirect political socialization. Frequent discussions about political topics within the family and higher socio-economic status (SES) are associated with greater willingness for political participation (Quintelier, 2015; Hoskins et al., 2016; Janmaat and Hoskins, 2022). Janmaat and Hoskins (2022) indicate that political attitudes and actions are transmitted from parents to children, with this transmission occurring during early adolescence and concluding by age 16. Families have a persistent influence on their children's political engagement through the observation of political behaviors, discussion about political issues, and exchange of political information (Jennings et al., 2009). Because political knowledge and civic engagement is related to SES, the exposure to political influence of a child will depend on SES, including factors such as parents' education, occupation, and income (Deimel, 2023). Moreover, children can influence family inputs, as shown by Dahlgaard (2018), who, using a regression discontinuity design, finds that parents are more likely to vote when their child becomes enfranchised, indicating that trickle-up socialization can occur where children's civic engagement impacts their parents'.

2. *School Environment*: Schools influence students through curricular content, principles of political education, and participation experiences. The manner of political education, including open classroom environments and participation in school decision-making bodies, impacts the acquisition of political knowledge, attitudes, and behaviors. Civic education can compensate for differences in family inputs or accelerate these differences depending on the educational system and the quality of civic education provided (Deimel, 2023). Learning about the political system, different parties, and the role of citizens can decrease the costs of political participation. This could be especially important for children receiving lower exposure to political knowledge transfer at home

(Neundorf et al., 2016), which means that schools have a *compensating effect* for family inputs. However, in a tracked educational system, such as Germany, children will likely interact with peers of similar SES and, thus, with a similar socialization at home (Janmaat et al., 2014). It could also be that students in higher school tracks receive more and higher-quality civic education. In this case, civic education at school has an *accelerating effect* on already existing differences by family background (Deimel, 2023). However, civic education can also produce socially unequal effects and even negative impacts for students from more privileged backgrounds. Deimel et al. (2021) show that in the Netherlands, Denmark, and North Rhine-Westphalia, high levels of formal civic education are correlated with lower willingness to vote among 14-year-olds from high-SES families, while being correlated with higher willingness to vote among those from low-SES backgrounds. The reasons for this divergent effect remain unclear, but the authors suggest that students from high-SES families may already possess substantial political knowledge, prompting teachers to focus on more critical content.

3. *Broader societal and institutional factors*: Broader societal and institutional factors form the framework within which families and schools operate. Democratic systems and societal values exert significant influence on political attitudes: several studies find that time spent under a democratic system tends to increase support for democratic principles (Fuchs-Schündeln and Schündeln, 2015; Acemoglu et al., 2021; Kotschy and Sunde, 2022). Moreover, societal and institutional factors shape the opportunities for political socialization provided by schools. Deimel (2023) highlights the significance of these macro-level influences, emphasizing that variations in school resources, curricular content, and teacher qualifications due to differences in teacher training for political education can significantly affect political socialization outcomes. Studies like Quintelier et al. (2011) suggest that countries where political education is a distinct subject tend to have higher levels of participation in legal protests. However, systematic investigations into the effects of education policy-prescribed hours for political education have been limited and our study aims to fill this gap.

In this study, we focus on the impact of compulsory civic education in schools on civic engagement. Specifically, we assess the effect of the number of hours in the school curriculum dedicated to civic education while holding constant the dimensions of family background and societal context. Although our primary interest lies in the school channel, particularly the effect of the number of hours civic education is taught, we also address the other dimensions of the above presented theoretical framework.

In a heterogeneity analysis, we focus on the role of parental background and discuss the societal context of educational reforms, assessing their impact on different cohorts. To provide a comprehensive understanding of the broader societal and institutional context, the next chapter will delve into the historical and institutional background.

A3.2 Institutional Background: Civic Education in Germany

Based on an extensive literature review on the history of civic education in (West) Germany, this section explores the evolution and goals of civic education over time. We examine its historical development, organizational structure, educational content, and the reforms of the number of instructional hours; for a more in-depth discussion, see Sendzik et al. (2024).

History of Civic Education in Germany After World War II, the Western Allies, particularly the United States, prioritized civic education in schools to promote democratic development in their occupation zones (Gagel, 2005; Herrlitz et al., 2005; Detjen, 2013; Kuhn, 2013).

The re-education approach of the United States, reacting to the atrocities committed under the Nazi regime, aimed for significant educational reforms, including the introduction of political education modeled after Social Studies in the U.S. However, the actual implementation faced resistance, particularly in Bavaria, where local authorities preferred a Christian renewal as a safeguard against fascism. Despite these challenges, the U.S. efforts led to the introduction of political education as a subject in Berlin, Schleswig-Holstein, and Hesse by 1946, and in Württemberg-Hohenzollern by 1949. This period also saw the establishment of social science departments in universities, enhancing the professional training of teachers in political education (e.g. Kuhn, 2013).

France and Britain pursued their own educational reforms regarding political education in their occupation zones. Historical evaluations suggest these efforts were less forceful than those of the U.S. The British left it to the governments of their zone (present-day Schleswig-Holstein, Hamburg, Lower Saxony, North Rhine-Westphalia) to decide whether to introduce political education as a subject. The French adopted a laissez-faire approach to political education (present-day Rhineland-Palatinate, parts of Baden-Württemberg, Saarland). In the Soviet occupation zone, political education initially did not hold the same significance in reform considerations until 1949, although foundational elements for the later state-centered subject "Staatsbürgerkunde" were laid early on.

The late 1960s and 1970s witnessed a significant shift as political education gained prominence due to student and youth movements. The criticism of existing civic education as ineffective in preventing unrest led to calls for more critical and realistic political education. Chancellor Kurt Georg Kiesinger acknowledged that political education should foster a critical attitude in 1968. Consequently, the government promoted reforms to make political education more relevant and responsive to contemporary issues, emphasizing critical thinking over rote learning of institutional knowledge (Gagel, 2005).

Subsequently, political education became a mandatory, standalone subject with extended instructional hours. Simultaneously, political education gained traction in teacher training, exemplified by the increasing establishment of professorships for political education. In the populous, social-democratic-governed states of Hesse, North Rhine-Westphalia, and Lower Saxony, the content of civic education was reoriented to focus less on institutional knowledge, such as the structure of the state or principles of electoral law, and more on developing a critical attitude toward state actions and the capitalist economic system. This reorientation triggered criticism from the conservative CDU party, resulting in discussions about a joint standard for political education, namely the Beutelsbach Consensus.

The Beutelsbach Consensus of 1976 marked a pivotal moment, establishing foundational principles for political education that remain influential today. It outlined three key principles: *prohibiting indoctrination, presenting controversial topics in science and politics as controversial in class*, and *empowering students to analyze and influence political situations according to their interests* (Reinhardt, 2017). For the first time, different parties agreed on a minimum framework, shifting the focus from competing political systems to educational practice. It maintained the spirit of political pluralism, which holds specific importance considering the exploitation of schools for propaganda purposes during the Nazi era.

Organization and Educational Content of Civic Education in Germany Germany's education system is decentralized, with each of the 16 states responsible for education within its jurisdiction. While specific details may vary across states, the overall structure is uniform. After four years of primary school, children are divided into three lower secondary school tracks at age ten: the basic track (five years), the middle track (six years), and the high track (eight or nine years). The high track, also known as the academic track, prepares students for university entrance, providing them with the university entrance degree (Abitur). The basic track prepares students for vocational school and apprenticeship training. The middle track offers routes for further vocational training and the option to attend a university of applied sciences. All tracks teach the same core subjects.

Civic education is introduced as a standalone subject in schools, starting from the secondary level. To prevent political influence from the ruling party within the classroom, civic education in post-World War II Germany is designed as a pluralistic, federally structured, and subsidiarity-based cooperative system. This is reflected in the varying names and combinations of the subject across different German states. In 2003, recognizing the impracticality and undesirability of strong standardization, the Society for Didactics of Politics and Political Youth and Adult Education agreed on output-oriented educational standards. These standards focus on three key areas of competence (Sander, 2005):

1. Analytical Competence: The ability to judge and analyze political events, problems, and debates.

2. Action Competence: The ability to formulate opinions, beliefs, and interests, and the skills to negotiate and compromise.

3. Methodological Competence: The ability to independently navigate current political, economic, and legal issues.

The educational content of civic education in Germany is thus framed by a liberal understanding of the state, which aims to empower citizens politically and ensure the "primacy of the individual over the state" (Gagel, 2005; Kuhn, 2013). This framework contrasts with historical examples such as monarchic authoritarian states in the 17th to 19th centuries, which sought to cultivate obedient subjects (Detjen, 2013). The former German Democratic Republic aimed to educate its citizens into loyal, socialist personalities, notably through state-mandated civic education in schools (Kerbel, 2016).

In terms of teaching methodology, the Federal Agency for Civic Education provides teaching materials to support educators in delivering civic education effectively. Political pluralism is a central tenet, reflected in the decentralized and cooperative nature of the system.

For our analysis, we focus on individuals who entered secondary school after 1975, due to the lack of information and comparability before the Beutelsbach Consensus 1976. We also exclude the East German states for the sake of comparability.

Reforms of the Number of Hours Civic Education is Taught Calls for enhancing both the quantity and quality of civic education in schools are recurrent across Germany. Over the last fifty years, these calls have prompted reforms in the allocation of hours dedicated to civic education. Recently, there has been a growing emphasis on the role of civic education in addressing societal challenges (Manning and Edwards, 2014a; Hamm et al., 2023). The German government emphasized the need for a stronger integration of civic education across all levels and stages of schooling, in its response to the 16th Children and Youth Report in 2020 (BMFSFJ, 2020).

In the German federal education system, education ministers of states are responsible for implementing any reforms in civic education including any changes to "Stundentafeln", or school hour schedules. School hour schedules are enacted through ordinances as part of training and examination regulations, typically with the approval of the school committees of the state parliaments. These regulations are based on the respective state's school law (Gökbudak and Hedtke, 2018). Thus, school hour schedules articulate the educational policy intentions of the ruling parties in the state parliament and government, quantified in the form of weekly hours allocated to learning areas and school subjects. They set binding guidelines for schools, subsequently affecting the amount of civic education received by cohorts of students. These reforms thus reflect the dynamic nature of educational policies and their impact on student learning experiences.

A3.3 Data

In this chapter, we present the data for analyzing the impact of civic education reforms in Germany. Our data encompass a detailed record of civic education hours across various school tracks and states, individual-level data including outcome variables and controls, as well as state-level controls, such as other educational reforms and party affiliation of education ministers.

A3.3.1 A Novel Measure of Civic Education

For this study, we digitized historical legal records from several archives across Germany and compiled a novel database on the hours of civic education taught in all three major school tracks in each grade across all West German states.[2] This database is based on coding all available legal records that document changes in the curricula. We coded the weekly compulsory hours of civic education in lower secondary school grades, typically from grade five to grade ten. That way, we can reconstruct how many hours per week of civic education students should have received while attending lower secondary school for each birth cohort. The result is a unique longitudinal dataset detailing both the intensive (how many hours) and extensive margins (any hours at all) of civic education taught in German school tracks.

Our dataset leverages approximately 400 schedules and legal regulations, providing comprehensive coverage of mandatory civic education in secondary schools from 1976. "Stundentafeln", or school hour schedules, play a crucial role in our study – as an example see Figure A3.1. These schedules, which are implemented via legal orders, quantify the education policy of state governments in terms of weekly hours dedicated to various subjects. They reflect the political will regarding the emphasis placed on different educational topics and the allocation of instructional time (Tillmann, 2020). Therefore, school hour schedules are a suitable measure for assessing the quantity of civic education provided, although they do not account for the quality of instruction.[3]

The schedules use different names for the civic education, including Politikunterricht (politics), Gesellschaftskunde/-lehre (social studies), Sozialkunde (social studies), and Weltkunde (world studies). Civic education is often combined with other subjects. For instance, as shown in Figure A3.1, politics is integrated with history and geography. To measure only the politics component, we allocate one-third of the total weekly hours to politics, assuming that the time is equally divided among the three subjects. This method also accounts for potential quality variations when politics is combined with other subjects. Civic education quality might be higher if taught as a standalone subject because it allows for more focused and in-depth exploration.

[2] For more details on data collection and preparation, see Sendzik et al. (2024).

[3] Only under very rare circumstances, schools are allowed to deviate from the prescribed schedules due to factors such as lack of teachers or temporary school-specific adjustments.

To visualize the substantial variation in civic education across states and school tracks over time, we present detailed patterns of civic education hours for each state for the basic track in Figure A3.2, the middle track in Figure A3.3, and the highest track in Figure A3.4. Each line represents the changes in weekly civic education hours for a specific state. The x-axis indicates the year a cohort began grade five, illustrating the progression over time. While some states had no civic education for extended periods, most states implemented civic education with varying intensities in their curricula since the 1970s.

Additionally, Figure A3.5 shows the average civic education hours across different school tracks and states. It shows that students in Hesse and North Rhine-Westphalia receive the most civic education hours, while those in Schleswig-Holstein and Bavaria receive the fewest. There is no indication that any specific track consistently has more civic education hours overall. In most states, the basic school track is assigned more hours of civic education than the highest school track, while in Baden-Württemberg and North Rhine-Westphalia, the middle school track has the most hours.

In conclusion, our civic education dataset provides a comprehensive and detailed historical record of civic education hours in West Germany, enabling longitudinal analyses of education policies and their impact on political socialization during secondary schooling in all major school tracks.

A3.3.2 Individual Data

We match the civic education records with individual-level data from the German Socio-Economic Panel (SOEP) from the survey years 1984 to 2020. The SOEP allows us to link our individual-level outcomes of interest measured at several points in time to the state, school track, and time of their secondary schooling in childhood. With the panel structure, we can analyze the effect of civic education on outcome variables over the life course of individuals.

To match the outcome data with the curricular measure of civic education, we use the best information available on state of schooling, school track, and entry year. Preferably, we use available information on the secondary school entry year form the SOEP. In the worst case, we use the birthdate information and state-time specific school entry cutoff dates that we collected.[4]

[4] For 16 percent of the observations, the SOEP provides information on the year the individuals entered secondary school, another 77 percent have information on the school entry year. We approximate the entry year to secondary school for the remaining 7 percent with birthdate information and state-time specific school entry cutoff dates. We derive direct information on the school track for 15 percent. For another 15 percent, we use information on their initial school leaving certificate, and for the remaining 70 percent, we use information on their highest school certificate, i.e., secondary general school degree after 9 years (Hauptschulabschluss), middle school degree after 10 years (Realschulabschluss), or university entrance degree after 12/13 years (Abitur). 78 percent of the observations have information on the state of schooling. For another 8 percent, we approximate state of schooling with state of birth. For 14 percent of the observations, we approximate state of schooling with the state of current residence.

Our final dataset combines 79,604 observations of individuals who entered secondary school between 1976 and 2010 (see Table A3.1 for an overview). In this period, students in grade five to ten had 0 to 1.67 hours of civic education per week. The average number of weekly hours is 0.78. 33.7 percent of the students in our data attended the highest track (Gymnasium), 35.0 percent attended the middle track (Realschule), and 31.3 percent attended the basic track (Hauptschule). 52.6 percent are women, and 25.3 percent have a direct or indirect migration background. A direct migration background indicates that the individual was not born in Germany. An indirect migration background means that the individual was born in Germany but has parents with direct migration background. We classify individuals as having a high socio-economic status if at least one parent obtained a university entrance degree (Abitur), which corresponds to completing the highest secondary school track in the German education system. In the final sample, 19.5 percent have a high SES.

Civic Engagement

As our main outcome variable, we use an index consisting of several measures to compare political actions and attitudes over time, considering the changing conceptions of political engagement in the West. Conceptual notions of political attitudes and actions in the West have evolved, often as a result of societal transformations and sometimes marked by conflicts (see Theocharis and Van Deth, 2018; Weiss, 2020 for a detailed discussion). This is crucial when searching for appropriate dependent variables to measure the impact of civic education across several cohorts. For example, since the late 1960s, participation research has expanded its understanding of what constitutes political activities, partly in response to the 1968 movement (Weiss, 2020). In addition to traditional activities such as voting, party membership, and contacting officials, research on political actions since the 1970s has included participation in demonstrations and petitions. From the 1990s onwards, volunteering and social engagement were added, and recently, non-institutional forms like consumer boycotts or social media protests have been considered. We use an index of civic engagement to consistently measure and reflect individual involvement in society across different generations. This method accommodates the evolving nature of civic engagement.

Our main variable of interest is an index of civic engagement that Hener et al. (2016) constructed specifically for the SOEP. The index aggregates four variables – *political interest, democratic party identification, organizational involvement,* and *volunteering* – by assigning equal weight to their respective z-scores. However, each of these components contains interesting information on its own, which is why we also report results for each outcome separately. After normalization, each of the variables has a mean of zero and a standard deviation of 1, such that coefficients can be interpreted in percent of a standard deviation. All outcomes point in the same direction and higher scores mean more civic engagement.

Although political interest is one determinant for political action (Verba et al., 1995), we combine this with further variables that pin down actual engagement. As such, political interest represents a first-stage outcome. Hener et al. (2016) add democratic party identification to the index because it has a strong link to political behavior. The concept goes back to Campbell et al. (1960), who describe it as a social or political identity that is distinct from voting preferences. It can encourage individuals to vote, to become a party member, or to support the election campaigns. There is debate about whether party identification contributes to more or less informed decision making (see Dalton, 2016 for a detailed discussion). We code party identification with right-wing extremist parties such as "NPD" (since 2023 "Die Heimat"), "Die Rechte", and "Die Republikaner" as zero because these parties are officially not considered democratic parties. In contrast, the "AfD" is, so far, only a "suspected case" of right-wing extremism. Organizational involvement is a stronger concept than party identification and captures active involvement in a citizens' group, political party, or local government. Volunteering in clubs and social services allows for a more general contribution to public affairs.

Below, we report the SOEP questions from which we draw the information for our civic engagement index. The brackets contain any further transformation of the original coding.

> *Political interest*. Generally speaking, how much are you interested in politics? [Very interested, moderately interested = 1; not much interested, or completely disinterested = 0]
>
> *Democratic party identification*. Many people in Germany lean towards one party in the long term, even if they occasionally vote for another party. Do you lean towards a particular party? [y/n] + Which party do you lean toward? [0 for non-democratic right-wing party]
>
> *Organizational involvement*. Which of the following activities do you take part in during your free time? [Involvement in a citizens' group, political party, local government]
>
> *Volunteering*. Which of the following activities do you take part in during your free time? [Volunteer work in clubs or social services]

Political interest and democratic party identification are available throughout all survey years of the SOEP. The questions for organizational involvement and volunteering are roughly contained in every second year (23 of 37 survey years). In our baseline specification, civic education is missing if any of the component variables are missing. We also conduct robustness checks where we ignore missing values in the components.

State-Level Controls

Additionally, we draw information on the most important school reforms in Germany, such as years of compulsory schooling and shortened upper secondary school (German G8 reforms) from Helbig and Nikolai (2015). The duration of compulsory schooling differs by state, ranging from 9 to 10 years, as illustrated in Table A3.1. Table A3.1 shows that only 5 percent of our sample were affected by the G8 reform.

Historically, center-left parties are more supportive of civic education (Sendzik et al., 2024). To account for this trend, we gather data on the party affiliation of the education minister in each state (Irmert et al., 2023). In the final sample, 41.9 percent of the education ministers were from center-left parties.

In sum, the combination of these different data sources provides a unique perspective on the evolution and impact of civic education in Germany. In the following, we first conduct a descriptive analysis of civic engagement across state and time before we present our empirical strategy to estimate the causal effect of civic education on civic engagement.

A3.4 Correlates of Civic Engagement

In this section, we explore various correlates of civic education outcomes to illustrate the analytical potential of our dataset. There is an ongoing debate about whether today's youth are becoming increasingly apolitical (Kitanova, 2020). However, it is often unclear whether this reflects the typical lower civic engagement seen in younger age groups, which tends to increase with age, or if the current young generation is distinctly less civically engaged. By leveraging the long time span of our data, we can observe individuals from different cohorts at various ages. For example, those born in 1970 are tracked at ages 20, 30, and 40. We can compare 20-year-olds born in the 1970s with 20-year-olds born in the 1990s. This allows us to differentiate between age specific patterns (changes in civic engagement due to age) and cohort specific pattern (differences based on the time period of school enrollment). Moreover, the consistent measurement of key variables across numerous cohorts allows us to examine the relationship with demographic factors such as socio-economic status and gender. This comprehensive analysis not only tests established theories but also exemplifies opportunities our dataset offers for advancing future research on civic education and political behavior.

In a first step, we describe some general patterns in our main outcome variable and the different components without controlling for civic education. Figure **??**) shows that, in line with existing literature, civic engagement increases with age (Putnam, 2000) and education (Dee, 2004). Specifically, adults that went to a higher secondary school track report higher civic engagement. In the graph, we control for survey years and cluster standard errors on the individual level. Men report more civic engagement than women (Figure **??**). While the educational and age patterns can be observed in most OECD countries, the gender gap in civic engagement is particularly large in Germany (OECD, 2020a). Panel c) shows that there is also a gap in civic engagement by SES.

It is worth looking at the different components of civic education separately. Political interest (Figure A4.1) shows a similar pattern to the overall index but, here, the gender gap increases with age. For democratic party identification, the gender gap is somewhat lower, while the SES gap remains unchanged (Figure A4.2). Organizational involvement does not show the same strong increase over age, and there is a smaller gap between basic and middle track. Also, the gender and SES gap are substantially smaller with no gap on the tails (Figure A4.3). For people in the highest school track, volunteering increases from their late teens to their early twenties and then decreases until their early thirties (Panel **??**). Note that the highest school track qualifies students for university while people in the middle and basic track enter the labor market at a younger age. It seems plausible that students at the university are more involved in volunteering activities than people in the labor market. The pattern of the highest track is very similar to the high SES graph in Panel c) as there is a large overlap between high SES and the highest track. In the late thirties, volunteering increases again. This increase seems to be driven by women who are more involved in volunteering activities than men in their forties (Panel **??**).

In Figure A3.7, we present the pattern of civic education with cohorts entering grade five (secondary school) at the x-axis, controlling for age groups. Overall, the pattern is very flat, with little variation across cohorts. Again, the graph shows the expected gaps by school track, gender, and SES. Figure A4.5 to Figure A4.8 display the pattern separately for political interest, democratic party identification, organizational involvement. Political interest in Figure A4.5 shows no clear pattern across cohorts. There is a decrease in democratic party identification for younger cohorts in Figure A4.6, indicating that strong bonds to one specific party have become less important. While organizational involvement is relatively flat across cohorts 1975 to 1995, there is an increase in organizational involvement for cohorts entering secondary school after 2000 (Figure A4.7). For volunteering in Figure A4.8, there is an overall increase across cohorts. Generally, it is difficult to disentangle cohort, age, and year effects from each other (Kotschy and Sunde, 2022). However, these patterns demonstrate that different cohorts focus on different forms of civic engagement, and illustrate why it is consistent in our context – studying school reforms across different cohorts and age groups – to use the civic engagement index conceptualized by Hener et al., 2016.

A3.5 Empirical Model, Identification, and Estimation Strategy

To estimate the effect of civic education on civic engagement later in life, we exploit the variation in civic education across German secondary school tracks, states, and cohorts. Employing the new difference-in-differences methodology developed by De Chaisemartin and d'Haultfoeuille, 2024, we identify effects on adult outcomes by comparing state-school-track groups that experienced changes in civic education hours between cohorts with those that did not, starting from the same baseline.

Two qualities of our treatment prevent us from estimating it with a conventional two-way fixed effects regression. First, the number of hours in civic education is (quasi) continuous. Second, the number of hours in civic education increases and decreases over time. Recent advances in the staggered difference-in-differences literature propose solutions to heterogeneous treatment effects but usually assume that the treatment is binary and absorbing (Callaway and Sant'Anna, 2021; Goodman-Bacon, 2021; Sun and Abraham, 2021; Borusyak et al., 2024). Some heterogeneity-robust estimators allow for a non-absorbing treatment but assume a binary treatment (Wooldridge, 2021). Continuous treatments require parallel trends for all combinations of different doses of the treatment d and d' (Callaway et al., 2024). If the strong parallel trend assumption does not hold, the comparison across dose groups is still a causal response but the selection bias could contaminate the estimate.

De Chaisemartin and d'Haultfoeuille, 2024 propose a heterogeneity-robust difference-in-differences estimator for when treatment is non-binary and non-absorbing.

Estimand of Interest: Let $Y_{g,t}(d_1, ..., d_t)$ denote the potential outcome for a state-school-type group g and cohort t. Every group g is exposed to a treatment $D_{g,t}$ that is equal to some dose of civic education d_t for l periods. Group g's cohort-t outcome, $Y_{g,t}$, may be affected by D_g, t, group g's cohort-t treatment, but also by g's lagged treatments. F_g denotes the first cohort for which group g's treatment changes. Then, the expected difference between g's actual outcome for $F_g - 1 + l$ and the counterfactual outcome that g would have experienced had it remained at its initial value $D_{g,1}$ until $F_g - 1 + l$ is

$$\delta_{g,l} = E[Y_{g,F_g-1+l} - Y_{g,F_g-1+l}(D_{g,1}, ..., D_{g,1})] \tag{4.1}$$

Estimation: To estimate $\delta_{g,l}$, De Chaisemartin and d'Haultfoeuille, 2024 propose an estimator $DID_{g,l}$ that compares the $F_g - 1$-to-$F_g - 1 + l$ outcome evolution between group g and groups g' with the same starting dose, $D_{g',1} = D_{g,1}$, whose treatment has not changed yet, $F_{g'} \geq F_g - 1 + l$, requiring that there is a group with the same initial treatment dose. N_t^g denotes the number of groups g' with the same period-one treatment as g that have not (yet) switched their treatment dose. For groups that never change their treatment dose, let $F_g = T + 1$, with T_g being the past period where a group g' exists. For every g, such that $F_g \leq T_g$ and $N^g_{F_g-1+l} > 0$, we can estimate $\delta_{g,l}$ using

$$DID_{g,l} = Y_{g,F_g-1+l} - Y_{g,F_g-1} - \frac{1}{N^g_{F_g-1+l}} \sum_{g':D_{g',1}=D_{g,1},F_{g'}>F_g-1+l} (Y_{g',F_g-1+l} - Y_{g',F_g-1}) \tag{4.2}$$

The estimator assumes that a group's outcome does not depend on its future treatment (*no anticipation*) and that two groups with the same period-one treatment follow *parallel trends*.

$\delta_{g,l}$ can be the effect of being exposed to a higher or lower treatment dose for l periods. If groups switch from a higher to lower treatment, De Chaisemartin and d'Haultfoeuille, 2024 multiply the $DID_{g,l}$ by -1 which yields an estimator of having been exposed to a higher treatment dose for l periods for the aggregated $DID_{g,l}$. By default, all standard errors are clustered at the level of the group variable.

To study treatment effect heterogeneity by treatment path, we estimate trajectory-specific DID_l. The DID_l drops trajectories without comparison group. We show results for continuous, binary, and binned treatments which retain more groups in the estimator. For the extensive margin, we create an indicator variable equal to 1 if a group received more than zero hours of civic education, that is $1\{d > 0\}$. Additionally, we construct an indicator variable, $1\{d > 0.7\}$, that is equal to 1 if a group received more than 0.7 hours of civic education per week, which corresponds roughly to the mean value. This treatment compares receiving an above-average amount of civic education to a below-average amount. We refer to this as the intensive margin. We construct a binned treatment where we round the continuous measure of civic education to values of $\{0, 0.5, 1, 1.5\}$, which enables us to look more closely at different treatment paths. Table A3.2 presents continuous mean of civic education hours that different subgroups were exposed to, and observations for the binary and binned treatment. The continuous mean shows that the exposure to civic education is very similar across school tracks and cohorts. Also, along individual characteristics such as gender, parental education, and migration background, there is no sign of an unequal access to the amount of civic education. Overall, only 3 percent of our groups receive zero treatment. Specifically, only Schleswig-Holstein, Lower Saxony, Bavaria, and Saarland did not offer civic education at some point in time. Zero treatment is fairly balanced across school tracks and cohorts. Only the 2010 (grouped) cohort did not receive zero treatment. However, due to the characteristics of our data, which is cut off at 2010, there are very few observations for this cohort. Treatment and control group size for the intensive treatment dummy are more balanced by construction.

A3.6 Results

In this section, we present the results from our empirical model, aimed at quantifying the impact of civic education on civic engagement later in life. First, we examine the extensive margin, which compares individuals who received any civic education to those who did not. Next, we present the intensive margin, investigating how varying amounts of civic education hours influence civic engagement outcomes. Finally, we explore heterogeneity in these effects, analyzing how the impact of civic education differs across gender and socio-economic status.

A3.6.1 Main Results

Extensive Margin To investigate how civic education affects civic engagement, we start by estimating Equation 4.2 with the binary treatment equal to 1 if a group received some civic education versus no civic education. This corresponds to the extensive margin. In this setting, DID_l compares groups that switch from untreated to treated to groups that have not yet been treated or vice versa.

Column 1 of Table A3.3 reports the results for civic engagement without controls. The point estimate suggests that individuals that had civic education in secondary school show civic engagement that is 24.4 percent of a standard deviations higher. The specification is robust to including various sets of controls. Including state controls that vary over time (column 2) leaves the point estimate unchanged. Including individual controls (column 3) or both state and individual controls (column 4) reduces the point estimate only marginally to 24.0 percent. All point estimates are significant.[5]

Civic engagement is an index combining political interest, democratic party identification, organizational involvement, and volunteering. Table A3.4 presents the results for each component separately. Civic education increases political interest by 45.5 percent of a standard deviation, organizational involvement by 29.9 percent of a standard deviation, and volunteering by 33.4 percent of a standard deviation. There is no significant effect on democratic party identification.

In the next step, we drop each variable one by one from the index. The results are presented in Table A3.5. Dropping political interest decreases the civic engagement index to 19.7 percent of a standard deviation. Dropping organizational involvement or volunteering decreases the coefficient to 22.1 percent and 20.8 percent respectively. Dropping democratic party identification increases the coefficient to 36.6 percent.

Intensive Margin Instead of looking at the effect of moving from 0 to any civic education, one might be interested in the effect of moving from a below-average to an above-average value of civic education which corresponds more to an intensive margin of civic education. We construct a dummy variable that is equal to 1 if individuals received more than 0.7 hours of civic education on average per week in secondary school, i.e., $1\{d > 0.7\}$. Figure A3.8 shows an insignificant, negative point estimate, robust to including various sets of controls.[6]

In the next step, we create a treatment variable with four bins, where we round the continuous measure of civic education to no civic education, 0.5 hours per week on average, 1 hour per week on average, or 1.5 hours per week on average. Figure A3.8 shows that the aggregated effect of the binned treatment is negative and not significant. Lastly, Figure A3.8 presents the aggregated effect for the continuous treatment. The effect is close to zero and not significant.

[5] Note that this effect is identified from the basic and middle school track in Lower-Saxony and all school tracks Saarland switching into treatment in 1977 with Schleswig-Holstein as the control group and Bavaria's highest school track switching out of treatment in 1998 with a broad control group.

[6] The corresponding Tables A3.6, A3.7, and A3.8 contain more details.

While De Chaisemartin and d'Haultfoeuille (2024) propose a framework for continuous treatment, they emphasize that estimating treatment effects separately for different paths can be more conclusive than examining the aggregate effect, provided there is sufficient statistical power. Figure A3.9 shows the treatment effect for each treatment path of our binned treatment separately.[7] For groups moving from 0 to 0.5 hours of civic education, we observe a significant increase in civic engagement by 30.7 percent of a standard deviation. Moving from 0.5 hours to 1 hour has a negative point estimate corresponding to 2.9 percent of a standard deviation with controls. The effect is not statistically significant when including state controls. Groups moving from 1 to 1.5 hours have a negative point estimate that is not significant for all specifications. Similar to the results from our binary treatment, we find only a positive significant result at the extensive margin. The effect of an increase in civic education from a lower to a higher dose is ambiguous.

Table A3.10 shows that, when estimating the effect at the intensive margin, political interest, organizational involvement, and volunteering each have positive but statistically insignificant point estimates, whereas democratic party identification has a negative effect. Democratic party identification values stable democratic party preference as civic engagement. However, informed swing voters may choose different parties based on the issues or candidates in each election and not lean towards any party long-term. These voters might be penalized in this measure, leading to the negative estimate. In Table A3.11, we drop each component variable from the civic engagement index one by one. When we drop political interest, organizational involvement, or volunteering, the negative but insignificant point estimate remains. However, when we drop democratic party identification, the index moves slightly above zero, though it remains very close to zero and insignificant. Therefore, the interpretation of zero effects on civic engagement at the intensive margin is unchanged even when excluding democratic party identification.

Discussion: In summary, our analysis shows a significant positive effect of civic education on civic engagement, at the extensive margin. This means that individuals having any amount of weekly hours of civic education in school instead of zero hours have higher values of civic engagement later in life. However, we do not find significant positive effects of civic education at the intensive margin. This suggests that individuals having more weekly hours of civic education in school than the (non-zero) hours of the comparison group do not have higher values of civic engagement later in life.

There are several explanations, why moving from a positive dose of civic education to a higher dose does not increase civic engagement. First, civic education in the production of civic engagement might exhibit diminishing returns. While essential knowledge gained through civic education is crucial for boosting civic engagement, additional details, such as in-depth descriptions of voting systems, may offer limited benefits or even become counterproductive.

[7] Also note the corresponding Table A3.9.

This implies that there are limits to how increasing the quantity of civic education can enhance civic engagement. Second, those with more civic education may become more aware of the failures of politicians and, out of frustration or resignation, choose not to participate in civic engagement activities. For example, Harka and Rocco, 2022 provide evidence that more educated individuals in Italy are more likely to abstain from voting as a form of protest against politics. Third, civic education might influence the behavior of parents (Dahlgaard, 2018), who play a crucial role as key contributors to the development of civic engagement in their children. If parents see that their child is receiving civic education at school and discussing politics with teachers and peers, they may feel less compelled to engage in civic education at home. This reduction in parental input could hinder the child's civic engagement development, potentially offsetting or even reversing the positive effects of school-based civic education. Fourth, increasing the weekly hours of civic education in schools might have come at the cost of other subjects (to the best of our knowledge, especially language, history, and geography classes). This reduced focus on language education could undermine reading competencies, since strong reading skills are crucial for critically analyzing political actions, debates, and news. History classes offer valuable opportunities for civic learning (Bowen and Kisida, 2020). This trade-off could diminish civic engagement and offset the benefits of enhanced civic education. Fifth, crowding out could also happen with respect to open classroom environment policies and democratic educational aspects in other subjects that are sometimes effective (Hoskins et al., 2021; Deimel et al., 2024). As a consequence, this would counteract the potentially positive impact of increased hours of civic education.

In sum, analyzing both the extensive and intensive margins of civic education reveals the importance of considering the nature of school reforms, i.e., differentiating between moving from no treatment to some treatment vs. moving from some dose to a higher dose: this is most crucial in contexts in which we do not assume education production functions to feature constant returns to scale. Moving forward, we explore further heterogeneity in these effects across school track, gender, and SES.

A3.6.2 Heterogeneity

School Track To explore potential differences more precisely, we split the sample by school track, replacing the state-school track fixed effects with state fixed effects. Students in various tracks may have distinct educational experiences and socio-economic backgrounds, which could influence how civic education impacts their civic engagement. Table A3.12 illustrates the estimated effects of civic education on civic engagement on the intensive margin across different school tracks: general, middle, and high. The effects are small and statistically insignificant across all tracks. Specifically, the general and high tracks show slightly positive but non-significant effects, while the middle track shows a slightly negative and non-significant effect. These findings suggest that the impact of civic education on civic engagement does not vary meaningfully by school track. However, we run into power issues with this specification, limiting the robustness of our estimates. Additionally, we cannot estimate this for the extensive margin, further constraining our ability to draw definitive conclusions from the data for specific school tracks.

Gender Column 1 and 2 of Table A3.13 show significant positive effects on civic engagement later in live for women and men at the extensive margin of introducing civic education at school. The point estimate is larger for women and corresponds to 38.4 percent of a standard deviation. Due to large confidence intervals, we cannot reject that the effect for men has the same size. At the intensive margin, shown in Column 1 and 2 of Table A3.14, both effects for women and men on civic engagement remain insignificant. Thus, there are no significant gender differences both for introducing and increasing civic education. This indicates that the descriptive differences in civic engagement between men and women found in our dataset, as well as by OECD (2020a), cannot be attributed to a differential causal impact of civic education on men and women.

SES Column 3 and 4 of Table A3.13 show that the positive significant effect at the extensive margin is driven by individuals from low-SES families, i.e., if no parent had a university entrance degree (Abitur). The effect for individuals from high-SES families, where at least one parent had an Abitur, is insignificant. This is in line with a compensating effect (Neundorf et al., 2016), where learning about politics decreases the cost of political participation, especially for children receiving, on average, lower exposure to political knowledge transfer at home.

At the intensive margin in Column 3 and 4 of Table A3.14, individuals from low-SES backgrounds show a positive but statistically insignificant point estimate. Conversely, those from high-SES households exhibit a negative impact when exposed to increased civic education. The potential crowding-out effects discussed above could thus be especially strong for individuals from high-SES backgrounds, which is probably most plausible when it comes to crowding-out of parental input. However, the sample of high-SES individuals is small, with 7591 total observations and 76 switchers. Therefore, any conclusions drawn from this data should be interpreted with caution. Our dataset does not provide definitive insights into this mechanism, leaving it as an open question for future investigation.

A3.7 Alternative Specifications

Handling of Missing Values In the main specification, we set the civic engagement index to missing if one of the component variables is missing. Table A4.1 shows the effect of civic education on civic engagement ignoring missing values in one of the component variables. This gives more weight to variables that are observed in more survey years (political interest and democratic party identification). At the extensive margin, the effect remains positive and significant at around 18.5 percent of a standard deviation. At the intensive margin in

Table A4.2, the effect remains insignificant and close to zero. Table A4.3 and A4.4 present this specification for each component variable separately. Overall, this specification does not change our interpretation of the results.

State Approximation For 14 percent of our observations we approximate state of schooling with the state of current residence. Generally, mobility is lower in Germany than in the U.S. People who start vocational training typically remain closer to their home town as people who start studying. Thus, the state of schooling and the state of current residence (if both are known) are often identical, but with variations across school tracks: In the basic school track, where students typically continue with vocational training, the correlation coefficient between current and schooling state is 90 percent. In middle track, the correlation coefficient is 86 percent. For the highest school track, the correlation coefficient is only 62 percent. In Table A4.5, we drop individuals from the highest school track for which we approximate state of schooling with state of current residence. The effect is smaller than in our main specification with 11 percent of a standard deviation. At the intensive margin in Table A4.6, the effect remains close to zero and insignificant.

Pure Civic Education The school hour schedules often combine civic education with other subjects, as shown in Figure A3.1, where politics is integrated with history and geography. For our baseline treatment, when civic education is one of three subjects, we allocate one-third of the total weekly hours to politics. In Figure A4.7, we focus exclusively on civic education when it is taught as a standalone subject, which we refer to as pure civic education. There is no significant effect of pure civic education on civic engagement. Thus, we find no indication that introducing civic education as a separate subject is more effective than combining it with other subjects in the schedule.

Drop One State at a Time In Table A4.8, we drop one of the 10 West German states from the sample one by one. Schleswig-Holstein (SH), Lower-Saxony (NI), Bavaria (BY), and Saarland (SL) are the few states where students did not receive any civic education at some point in time. The effects remain positive significant for dropping Lower-Saxony, Bavaria, and Saarland. In contrast, dropping Schleswig-Holstein from our sample reverses the sign of our coefficient, since it is the only state that can serve as the control group for the groups switching into treatment. It is therefore critical in identifying the extensive margin effect. When we drop Schleswig-Holstein, the remaining negative effect is identified solely from Bavaria's highest school track switching out of treatment in 1998.

At the intensive margin in Table A4.9, all estimates remain close to zero and insignificant for dropping one state at a time.

A3.8 Conclusion

Our study investigates the relationship between civic education in schools and civic engagement in adult life, drawing from a comprehensive dataset spanning over four decades. Through a detailed examination of weekly civic education hours across German states, we employ a generalized difference-in-differences framework to estimate the causal effect of civic education. Our findings highlight the positive impact of introducing civic education as a subject, revealing a statistically significant positive effect on civic engagement at the extensive margin. When considering the intensive margin of civic education, however, we find negligible average effects.

We contribute to the literature in several ways. Firstly, we provide robust causal evidence on the impact of civic education on individuals' civic engagement trajectories. Secondly, our dataset, capturing nuanced variations in civic education provision in schools, sheds light on the longitudinal dynamics of education policies. The nuanced variations enables us to distinguish between different treatment paths. Thirdly, our methodological approach, employing continuous difference-in-differences techniques by De Chaisemartin and d'Haultfoeuille (2024), underscores the importance of these very recent analytical strategies in understanding policy interventions. This method allows us to estimate causal effects of a non-absorbing and continuous treatment, conditions often encountered in education policies, where adjustments such as the increase and decrease of hours or the implementation and reversal of reforms are common.

However, our study faces certain limitations. While our dataset offers detailed insights into the quantitative aspect of civic education provision, it does not capture the qualitative dimensions of pedagogical content. Although we can approximate some aspects of quality by distinguishing whether civic education is taught alongside other subjects or as a separate course, we lack information on the classroom environment and teacher quality. Additionally, relying on survey-based data from the German Socio-Economic Panel imposes constraints on the breadth of outcomes. For example, administrative election participation data and data on students' knowledge of civic education topics could deliver further important insights on the effects and mechanisms.

Looking ahead, our findings hold important implications for policy formulation and future research. Analyzing the content of civic education and its impact on civic engagement would be one important future research topic. Furthermore, exploring alternative pathways to bolster civic engagement beyond increasing civic education hours warrants further investigation – as our findings suggest that increasing the intensive margin of hours in school has limited impact. Thus, our research sets the stage for future research aimed at finding strategies for fostering active citizenship and democratic participation among future generations.

Figures and Tables

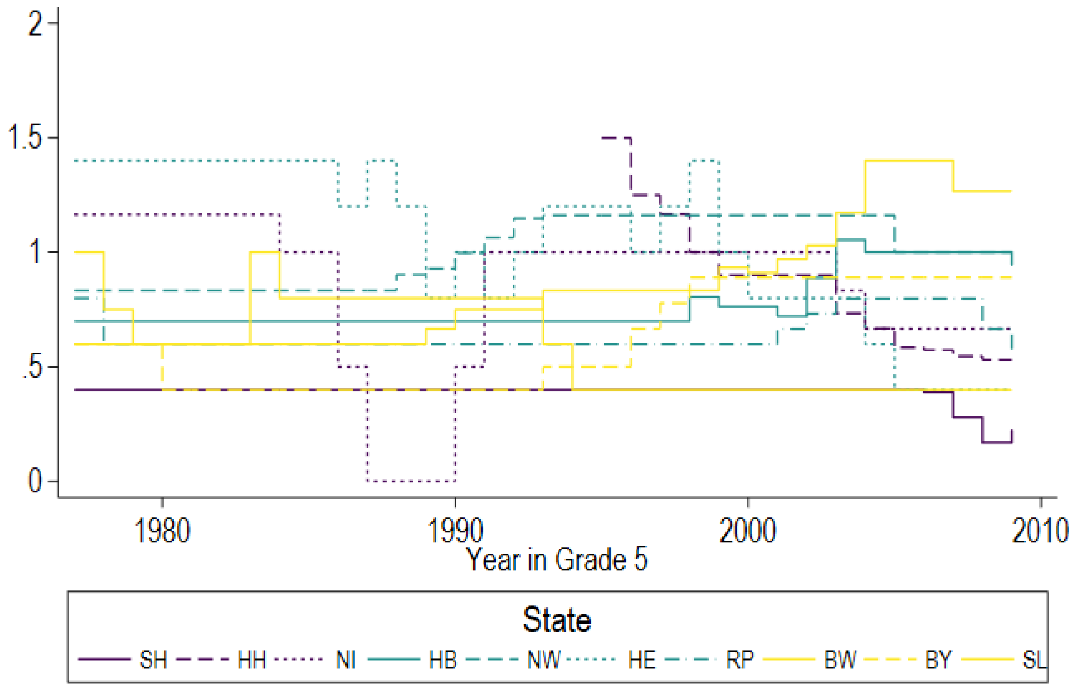
Figure A3.1 : Example of official school hour schedule indicating the number of weekly hours for each grade

Stundentafeln für die Sekundarstufe I - Hauptschule -

Klasse	5	6	7	8	9	10	Gesamtwochen- stunden
Wochenstundenrahmen*)	27-29	28-30	29-31	29-31	30-32	30-32	179
Lernbereich/Fach							
Deutsch	5-6	4-6	4-5	4-5	4-5	4-5	25-27
Gesellschaftslehre 1)2)	2-3	4-5	3-4	3-4	3-4	3-4	18-22
Geschichte, Erdkunde, Politik							
Mathematik	4-5	4-5	4-5	4-5	4-5	3-4	23-25
Naturwissenschaften 1)2)	4-5	3-4	2-4	3-4	3-4	2-4	17-21
Biologie, Physik, Chemie							
Englisch	5-6	5-6	4-5	3-4	3-4	3-4	23-25
Arbeitslehre 1)3)	-	-	2-4	2-4	3-4	3-4	11-13
Technik, Wirtschaft, Hauswirtschaft							
Kunst, Musik, Textil- gestaltung 1)4)	3-4	3-4	2-4	2-4	2-3	2-3	15-18
Religionslehre	2	2	2	2	2	2	12
Sport	2-4	2-4	2-4	2-4	2-4	2-4	17-19
Wahlpflichtunterricht 5)	-	-	2	2	2-4	2-4 6)	8-12
Förderunterricht	-	-	-	-	1-3	1-3 6)	2-6
zusätzlich: Muttersprachlicher Unterricht im Umfang von in der Regel 5 Wochenstunden							

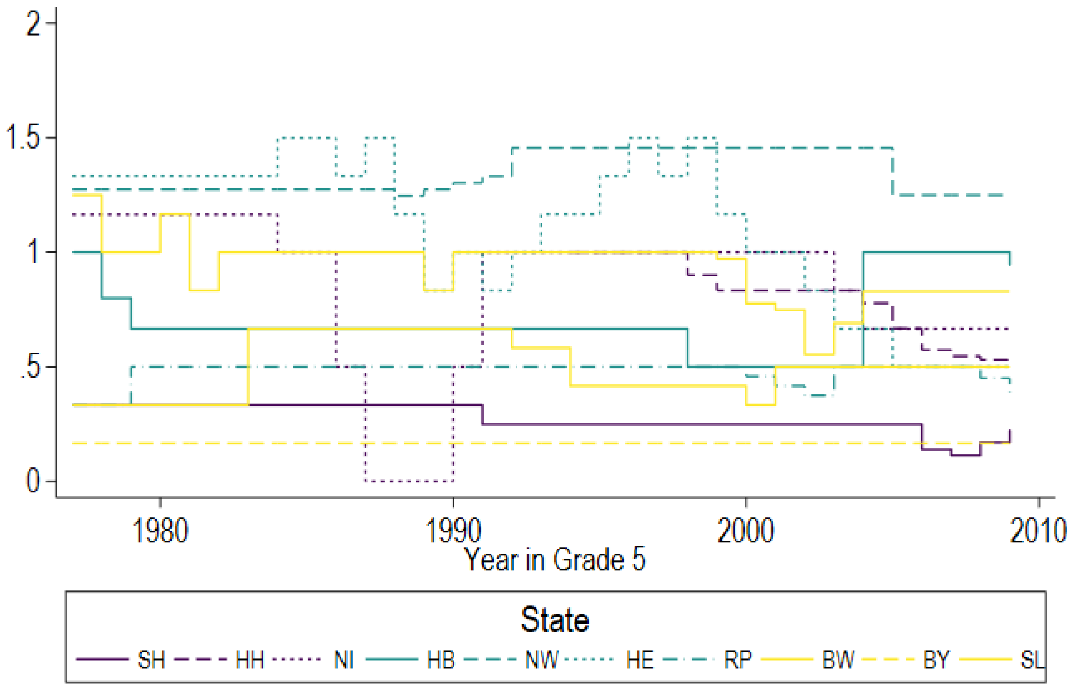
Notes: Source: North Rhine-Westphalia, 1998-10-21 Neuerlass, S. 638-9

Figure A3.2 : Civic education in the basic track



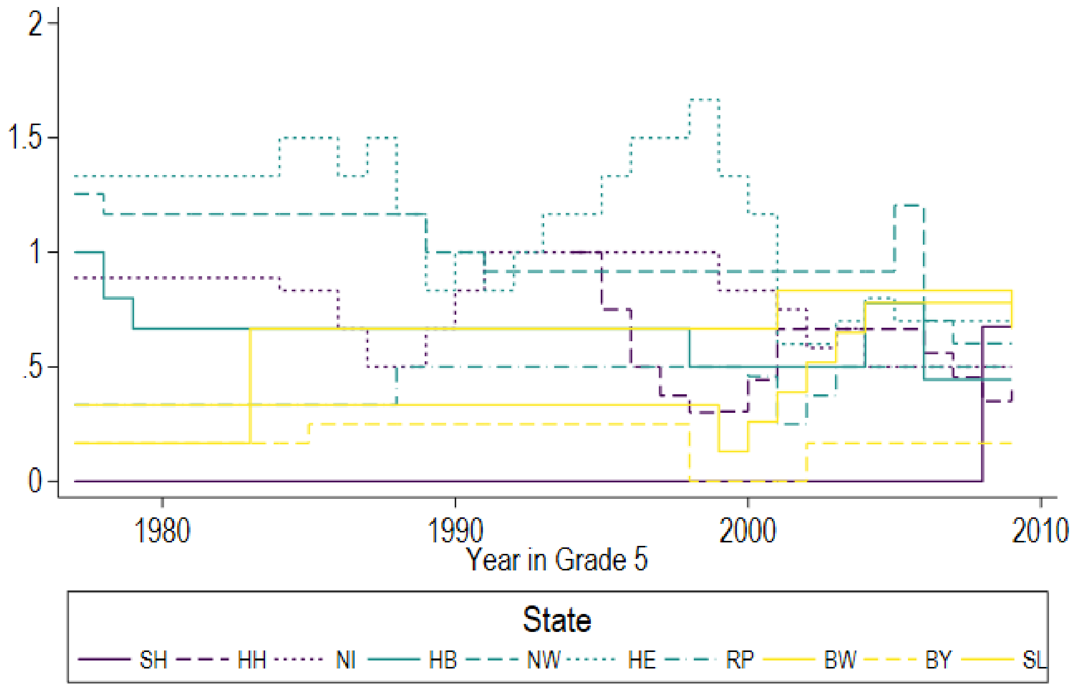
Notes: Average weekly hours of civic education in the basic track. Each line represents one state. Data source: Own data on civic education.

Figure A3.3 : Civic education in the middle track



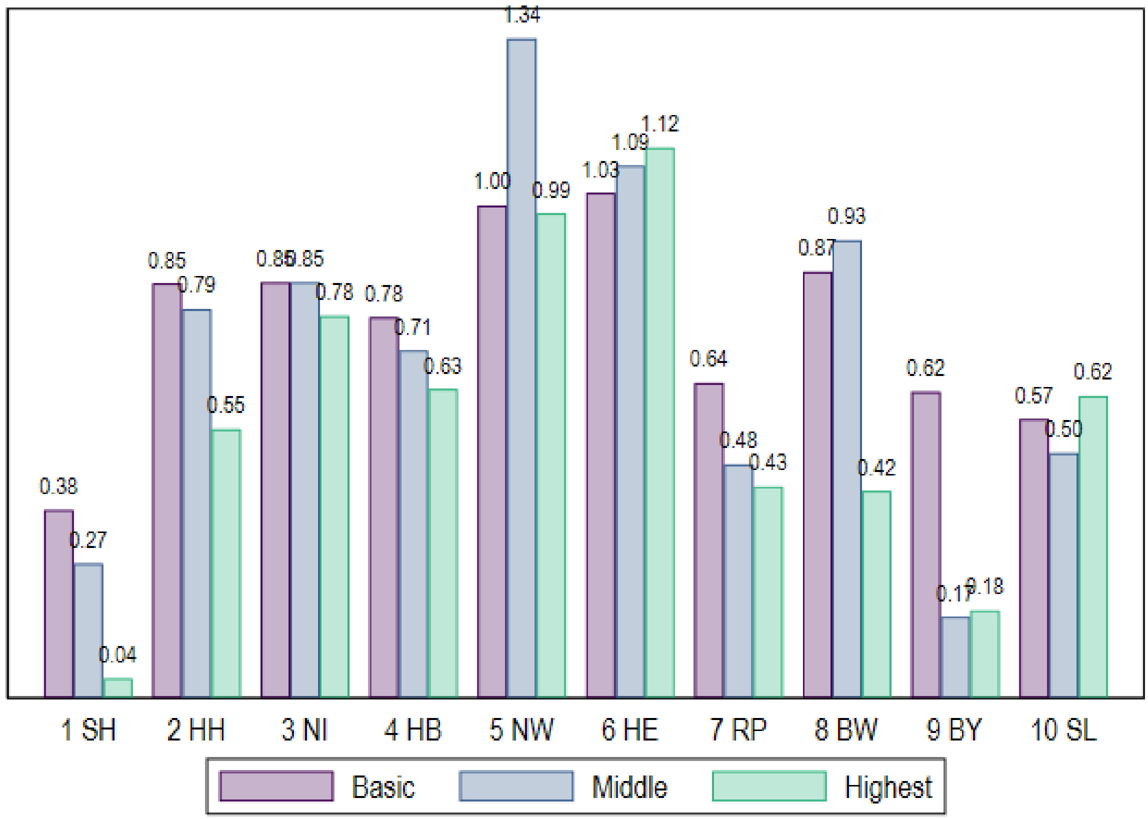
Notes: Average weekly hours of civic education in the middle track. Each line represents one state. Data source: Own data on civic education.

Figure A3.4 : Civic education in the highest track



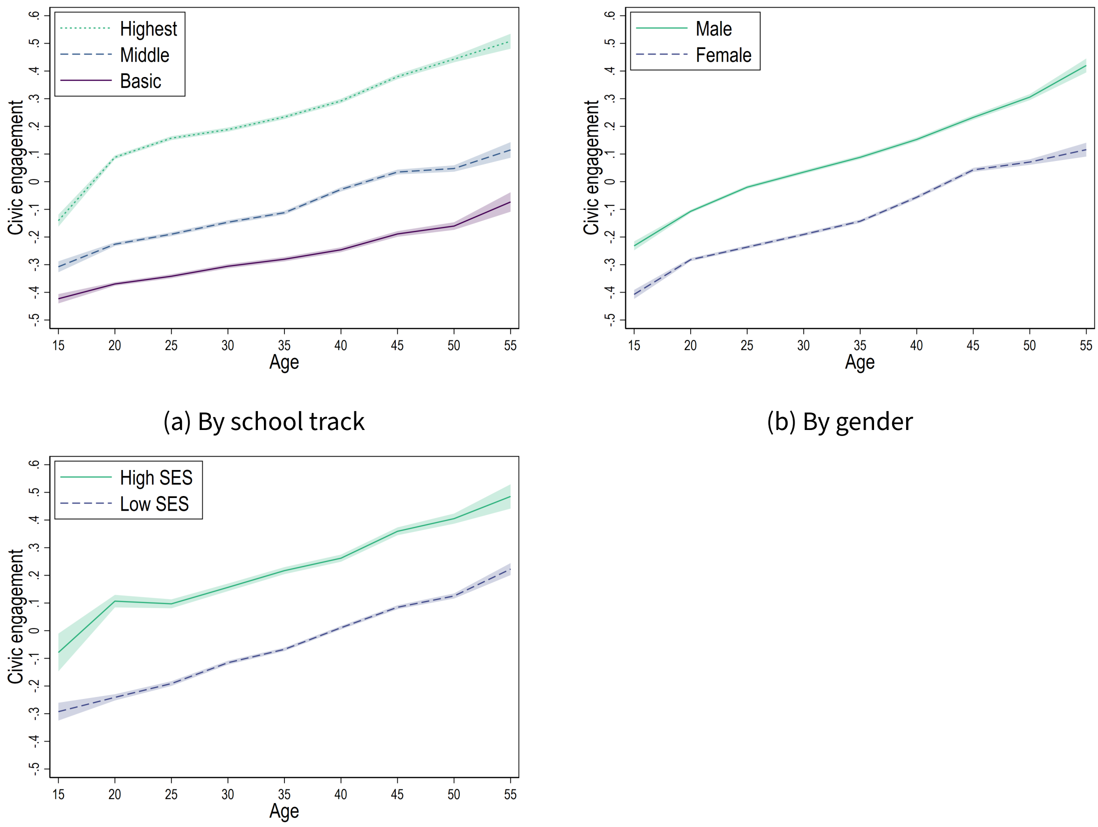
Notes: Average weekly hours of civic education in the highest track. Each line represents one state. Data source: Own data on civic education.

Figure A3.5 : Average number of hours of civic education in different school tracks and states



Notes: The figures show the average number of hours per week in civic education from grade 5 to 10 by school track and state. Data: own data on civic education for cohorts entering grade 5 between 1976 and 2010.

Figure A3.6 : General pattern of civic engagement by age



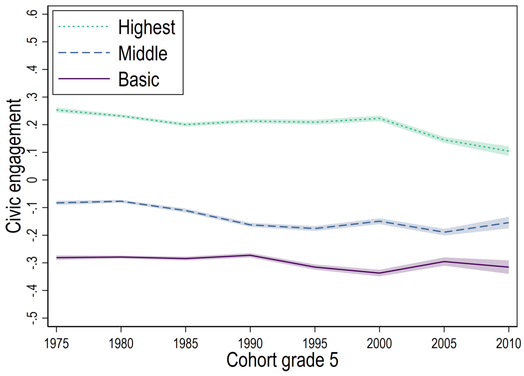
(a) By school track

(b) By gender

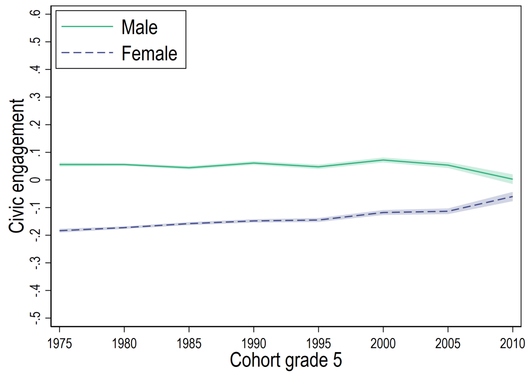
(c) By SES

Notes: Civic engagement by five-year age groups with a 95 percent confidence band. All plots conditional on cohort group, standard errors clustered on individual level. High SES is defined as having at least one parent with an Abitur. Data: SOEP.

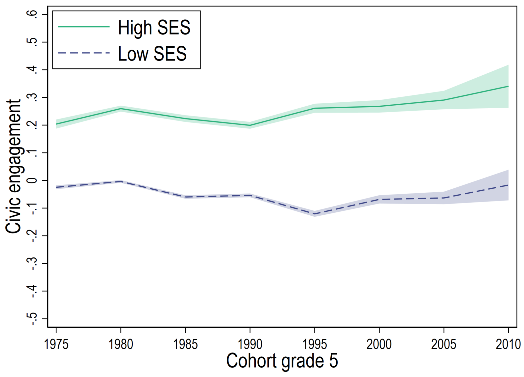
Figure A3.7 : General pattern of civic engagement by cohort



(a) By school track



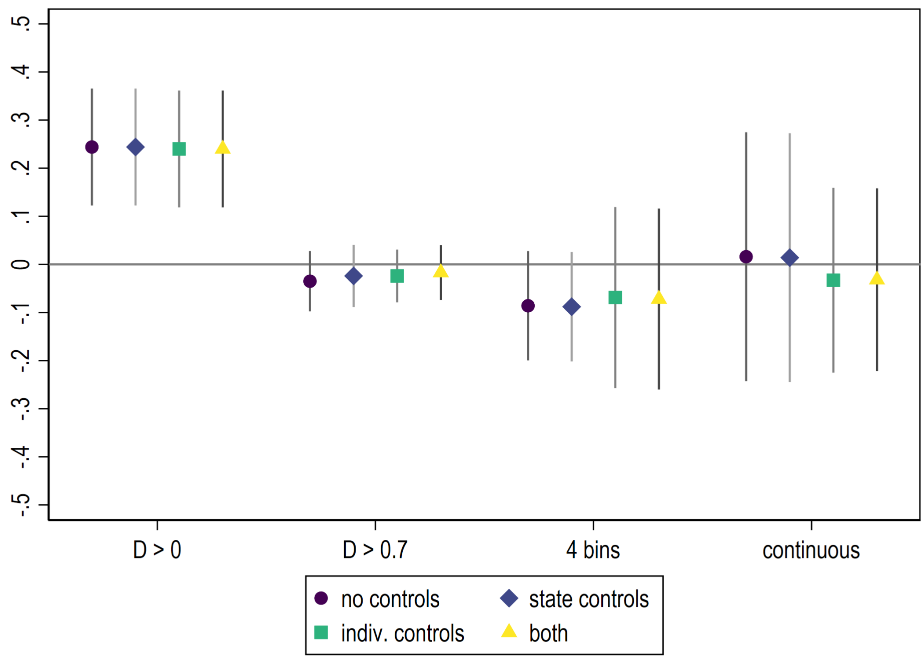
(b) By gender



(c) By SES

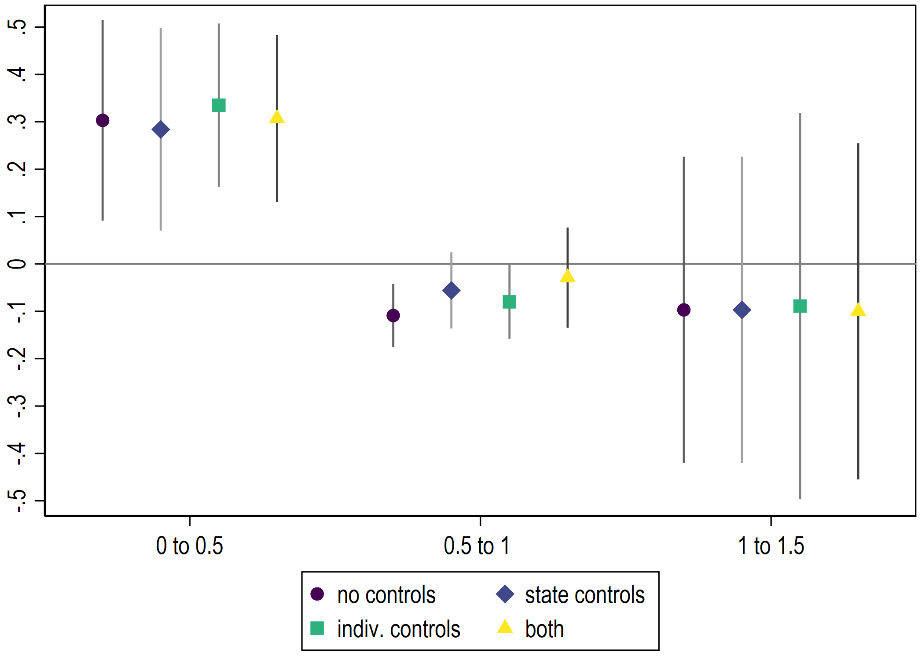
Notes: Civic engagement by five-year cohort groups with a 95 percent confidence band. All plots conditional on age group, standard errors clustered on individual level. High SES is defined as having at least one parent with an Abitur. Data: SOEP.

Figure A3.8 : The effect of civic education on civic engagement for different treatment definitions



Notes: This figure presents the results from Equation 4.2 with treatment $1\{d > 0\}$, $1\{d > 0.7\}$, the binned treatment with $\{0, 0.5, 1, 1.5\}$ hours of civic education, and the continuous treatment with 95 percent confidence intervals. State controls include the party in power at time of reform, compulsory school years, and G8. Individual controls include gender, individual born in Germany, and parental education. Standard errors are clustered at the state-school type level. Data: SOEP and own data on civic education.

Figure A3.9 : The effect of civic education on civic engagement by treatment path



Notes: This figure presents the results from Equation 4.2 with the binned treatment $\{0, 0.5, 1, 1.5\}$ hours of civic education for different treatment paths with 95 percent confidence intervals. State controls include the party in power at time of reform, compulsory school years, and G8. Individual controls include gender, individual born in Germany, and parental education. Standard errors are clustered at the state-school type level. Data: SOEP and own data on civic education.

Table A3.1 : Descriptive statistics

	Obs	Mean	Std.Dev.	Min	Max
Civics Mean gr. 5-10	79,604	0.779	0.413	0.000	1.667
Cohort gr. 5	79,604	1,987	9	1,976	2,010
Basic school track	79,604	0.313	0.464	0.000	1.000
Middle school track	79,604	0.350	0.477	0.000	1.000
Highest school track	79,604	0.337	0.473	0.000	1.000
Outcomes					
Civic engagement	79,604	-0.047	0.623	-0.652	1.768
Pol. interest	79,604	-0.124	0.955	-0.750	1.334
Party ident.	79,604	-0.134	0.977	-0.901	1.109
Org. involv.	79,604	-0.035	0.951	-0.330	3.033
Volunt.	79,604	0.103	1.044	-0.627	1.595
Controls					
Left government	79,604	0.419	0.493	0.000	1.000
Comp. years schooling	79,604	9.232	0.422	9.000	10.000
G8	79,604	0.049	0.216	0.000	1.000
Female	79,602	0.526	0.499	0.000	1.000
1st or 2nd gen mig.	79,604	0.253	0.435	0.000	1.000
High SES	38,983	0.195	0.396	0.000	1.000
Parental education					
No degree/general degree	38,983	0.557	0.497	0.000	1.000
Intermediate degree/ 10th grade	38,983	0.249	0.432	0.000	1.000
Upper secondary degree	38,983	0.195	0.396	0.000	1.000

Notes: The table reports descriptive statistics (mean, standard deviation, minimum, maximum, and the number of observations) for treatment, outcome and control variables. Data: SOEP and own data on civic education.

Table A3.2 : Treatment characteristics

	Continuous	D > 0		D > 0.7		Binned treatment			
		0	1	0	1	0	0.5	1	1.5
Total	0.779	2531	77073	34719	44885	9987	24928	31312	13377
School track									
Basic	0.760	425	24518	11702	13241	432	11298	11934	1279
Middle	0.863	862	26971	9943	17890	5876	4141	8110	9706
Highest	0.709	1244	25584	13074	13754	3679	9489	11268	2392
Cohort									
1970	0.818	895	21053	8839	13109	3260	5583	7840	5265
1980	0.762	935	30025	14760	16200	3754	11006	11228	4972
1990	0.793	408	17309	6694	11023	1535	5159	8818	2205
2000	0.714	293	8313	4193	4413	1393	2992	3340	881
2010	0.692	0	373	233	140	45	188	86	54
State									
SH	0.239	943	2019	2962	0	977	1985	0	0
HH	0.685	0	312	188	124	0	192	109	11
NI	0.824	1263	9114	3137	7240	1263	1874	7240	0
HB	0.752	0	787	591	196	0	595	124	68
NW	1.113	0	22316	326	21990	0	420	13917	7979
HE	1.230	0	7228	539	6689	0	539	1855	4834
RP	0.520	0	5192	4800	392	0	4820	372	0
BW	0.711	0	14354	7041	7313	97	7018	6754	485
BY	0.322	291	15100	14524	867	7576	6948	867	0
SL	0.519	34	651	611	74	74	537	74	0
Age									
10	0.772	233	8754	3942	5045	979	3012	3772	1224
20	0.771	930	29464	13329	17065	3525	9949	12437	4483
30	0.785	780	22072	9864	12988	2824	7040	8967	4021
40	0.788	477	14182	6404	8255	2192	4212	5304	2951
50	0.792	111	2601	1180	1532	467	715	832	698
Gender									
male	0.779	1290	36451	16549	21192	4303	12323	14952	6163
female	0.779	1241	40620	18168	23693	5684	12603	16360	7214
Parental education									
high SES	0.732	287	7304	3650	3941	1179	2476	2842	1094
low SES	0.792	1150	30242	13435	17957	4083	9374	11975	5960
Migration background									
no mig. background	0.770	2133	57362	26653	32842	8484	18303	22255	10453
1st or 2nd gen mig.	0.806	398	19711	8066	12043	1503	6625	9057	2924
Government									
Right-Center Party	0.564	2137	44132	31314	14955	9593	21893	14044	739
Left-Center Party	1.077	394	32941	3405	29930	394	3035	17268	12638

Notes: The table shows the average number of civic education from grade 5 to 10 in secondary school by subgroups and the number of observations by treatment groups for treatment $1\{d > 0\}, 1\{d > 0.7\}$, and the binned treatment with $\{0, 0.5, 1, 1.5\}$ for different subgroups. High SES is defined as having at least one parent with an Abitur. Total number of observations: 79,604, Data set: SOEP and own data on civic education.

Table A3.3 : Effect of civic education on civic engagement (extensive margin)

	(1)	(2)	(3)	(4)
ATE	.244	.244	.240	.240
	(.062)	(.062)	(.062)	(.062)
Observations	79604	79604	79604	79604
Switchers	678	678	678	678
Stayers	1302	1302	1302	1302
StateByType FE	Yes	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes
State Controls	No	Yes	No	Yes
Indiv. Controls	No	No	Yes	Yes

Notes: This table presents the results from Equation 4.2 following De Chaisemartin and d'Haultfoeuille (2024) with treatment $1\{d > 0\}$. Switchers are the treated groups that switch in or out of treatment and have a valid control group. Stayers are the valid control group with the same initial treatment dose whose treatment has not (yet) changed. State controls include the party in power at time of reform, compulsory school years, and G8. Individual controls include gender, individual born in Germany, and parental education. Standard errors are clustered at the state-school type level. Data: SOEP and own data on civic education.

Table A3.4 : Effect of civic education on political interest, democratic party identification, organizational involvement, and volunteering (extensive margin)

	(1)	(2)	(3)	(4)
VARIABLE	Pol. Interest	Dem. Party Ident.	Org. Involv.	Volunt.
ATE	.455	-.129	.299	.334
	(.030)	(.113)	(.078)	(.073)
Observations	79604	79604	79604	79604
Switchers	678	678	678	678
Stayers	1302	1302	1302	1302
StateByType FE	Yes	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes
State Controls	Yes	Yes	Yes	Yes
Indiv. Controls	Yes	Yes	Yes	Yes

Notes: This table presents the results from Equation 4.2 with treatment $1\{d > 0\}$. Switchers are the treated groups that switch in or out of treatment and have a valid control group. Stayers are the valid control group with the same initial treatment dose whose treatment has not (yet) changed. State controls include the party in power at time of reform, compulsory school years, and G8. Individual controls include gender, individual born in Germany, and parental education. Standard errors are clustered at the state-school type level. Data: SOEP and own data on civic education.

Table A3.5 : Effect of civic education on civic engagement dropping one variable from index (extensive margin)

	(1)	(2)	(3)	(4)
	Drop pol. interest	Drop party ident.	Drop org. involv.	Drop volunt.
ATE	.197	.366	.221	.208
	(.075)	(.045)	(.058)	(.071)
Observations	80121	82036	79745	79696
Switchers	696	714	679	680
Stayers	1304	1307	1305	1302
StateByType FE	Yes	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes
State Controls	Yes	Yes	Yes	Yes
Indiv. Controls	Yes	Yes	Yes	Yes

Notes: This table presents the results from Equation 4.2 following De Chaisemartin and d'Haultfoeuille (2024) with treatment $1\{d > 0\}$. The dependent variable is the civic engagement index, excluding one component at a time: (1) political interest, (2) democratic party identification, (3) organizational involvement, and (4) volunteering. Switchers are the treated groups that switch in or out of treatment and have a valid control group. Stayers are the valid control group with the same initial treatment dose whose treatment has not (yet) changed. State controls include the party in power at time of reform, compulsory school years, and G8. Individual controls include gender, individual born in Germany, and parental education. Standard errors are clustered at the state-school type level. Data: SOEP and own data on civic education.

Table A3.6 : Effect of civic education on civic engagement (intensive margin)

	(1)	(2)	(3)	(4)
ATE	-.035	-.024	-.024	-.017
	(.032)	(.033)	(.028)	(.029)
Observations	79604	79604	79604	79604
Switchers	1891	1891	1891	1891
Stayers	13593	13593	13593	13593
StateByType FE	Yes	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes
State Controls	No	Yes	No	Yes
Indiv. Controls	No	No	Yes	Yes

Notes: This table presents the results from Equation 4.2 following De Chaisemartin and d'Haultfoeuille (2024) with treatment $1\{d > 0.7\}$. Switchers are the treated groups that switch in or out of treatment and have a valid control group. Stayers are the valid control group with the same initial treatment dose whose treatment has not (yet) changed. State controls include the party in power at time of reform, compulsory school years, and G8. Individual controls include gender, individual born in Germany, and parental education. Standard errors are clustered at the state-school type level. Data: SOEP and own data on civic education.

Table A3.7 : Effect of civic education on civic engagement (binned treatment)

	(1)	(2)	(3)	(4)
ATE	-.086	-.088	-.069	-.072
	(.058)	(.058)	(.096)	(.096)
Observations	79604	79604	79604	79604
Switchers	2930	2930	2930	2930
Stayers	6015	6015	6015	6015
StateByType FE	Yes	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes
State Controls	No	Yes	No	Yes
Indiv. Controls	No	No	Yes	Yes

Notes: This table presents the results from Equation 4.2 following De Chaisemartin and d'Haultfoeuille (2024) with treatment bins $\{0, 0.5, 1, 1.5\}$. Switchers are the treated groups that switch in or out of treatment and have a valid control group. Stayers are the valid control group with the same initial treatment dose whose treatment has not (yet) changed. State controls include the party in power at time of reform, compulsory school years, and G8. Individual controls include gender, individual born in Germany, and parental education. Standard errors are clustered at the state-school type level. Data: SOEP and own data on civic education.

Table A3.8 : Effect of civic education on civic engagement (continuous)

	(1)	(2)	(3)	(4)
ATE	.016	.014	-.033	-.032
	(.132)	(.132)	(.098)	(.097)
Observations	79604	79604	79604	79604
Switchers	3531	3531	3531	3531
Stayers	16839	16839	16839	16839
StateByType FE	Yes	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes
State Controls	No	Yes	No	Yes
Indiv. Controls	No	No	Yes	Yes

Notes: This table presents the results from Equation 4.2 following De Chaisemartin and d'Haultfoeuille (2024) with continuous treatment. Switchers are the treated groups that switch in or out of treatment and have a valid control group. Stayers are the valid control group with the same initial treatment dose whose treatment has not (yet) changed. State controls include the party in power at time of reform, compulsory school years, and G8. Individual controls include gender, individual born in Germany, and parental education. Standard errors are clustered at the state-school type level. Data: SOEP and own data on civic education.

Table A3.9 : Effect of civic education on civic engagement by treatment path

	(1)	(2)	(3)
Treatment Path	0 to 0.5	0.5 to 1	1 to 1.5
ATE	.307	-.029	-.100
	(.090)	(.054)	(.181)
Observations	79604	79604	79604
Switchers	310	1564	977
Stayers	1382	7116	1230
StateByType FE	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes
State Controls	Yes	Yes	Yes
Indiv. Controls	Yes	Yes	Yes

Notes: This table presents the results from Equation 4.2 following De Chaisemartin and d'Haultfoeuille (2024) for each treatment path for the binned treatment. Switchers are the treated groups that switch in or out of treatment and have a valid control group. Stayers are the valid control group with the same initial treatment dose whose treatment has not (yet) changed. State controls include the party in power at time of reform, compulsory school years, and G8. Individual controls include gender, individual born in Germany, and parental education. Standard errors are clustered at the state-school type level. Data: SOEP and own data on civic education.

Table A3.10 : Effect of civic education on political interest, democratic party identification, organizational involvement, and volunteering (intensive margin)

	(1)	(2)	(3)	(4)
VARIABLE	Pol. Interest	Dem. Party Ident.	Org. Involv.	Volunt.
ATE	.034	-.155	.038	.017
	(.044)	(.061)	(.044)	(.043)
Observations	79604	79604	79604	79604
Switchers	1891	1891	1891	1891
Stayers	13593	13593	13593	13593
StateByType FE	Yes	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes
State Controls	Yes	Yes	Yes	Yes
Indiv. Controls	Yes	Yes	Yes	Yes

Notes: This table presents the results from Equation 4.2 with treatment D > 0.7. Switchers are the treated groups that switch in or out of treatment and have a valid control group. Stayers are the valid control group with the same initial treatment dose whose treatment has not (yet) changed. State controls include the party in power at time of reform, compulsory school years, and G8. Individual controls include gender, individual born in Germany, and parental education. Standard errors are clustered at the state-school type level. Data: SOEP and own data on civic education.

Table A3.11 : Effect of civic education on civic engagement dropping one variable from index (intensive margin)

	(1)	(2)	(3)	(4)
	Drop pol. interest	Drop party ident.	Drop org. involv.	Drop volunt.
ATE	-.028	.030	-.035	-.028
	(.030)	(.033)	(.029)	(.036)
Observations	80121	82036	79745	79696
Switchers	1921	1957	1894	1894
Stayers	13779	14015	13626	13606
StateByType FE	Yes	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes
State Controls	Yes	Yes	Yes	Yes
Indiv. Controls	Yes	Yes	Yes	Yes

Notes: This table presents the results from Equation 4.2 following De Chaisemartin and d'Haultfoeuille (2024) with treatment $1\{d > 0.7\}$. The dependent variable is the civic engagement index, excluding one component at a time: (1) political interest, (2) democratic party identification, (3) organizational involvement, and (4) volunteering. Switchers are the treated groups that switch in or out of treatment and have a valid control group. Stayers are the valid control group with the same initial treatment dose whose treatment has not (yet) changed. State controls include the party in power at time of reform, compulsory school years, and G8. Individual controls include gender, individual born in Germany, and parental education. Standard errors are clustered at the state-school type level. Data: SOEP and own data on civic education.

Table A3.12 : Effect of civic education on civic engagement by school track

	(1)	(2)	(3)
ATE	.031	-.093	.036
	(.052)	(.049)	(.036)
Observations	24943	27833	26828
Switchers	1099	385	315
Stayers	3184	1924	1748
State FE	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes
State Controls	Yes	Yes	Yes
Indiv. Controls	Yes	Yes	Yes
Sample	Basic	Middle	Highest

Notes: This table presents the results from Equation 4.2 following De Chaisemartin and d'Haultfoeuille (2024) with with treatment $1\{d > 0.7\}$ for each school track separately. Switchers are the treated groups that switch in or out of treatment and have a valid control group. Stayers are the valid control group with the same initial treatment dose whose treatment has not (yet) changed. State controls include the party in power at time of reform, compulsory school years, and G8. Individual controls include gender, individual born in Germany, and parental education. Standard errors are clustered at the state level. Data: SOEP and own data on civic education.

Table A3.13 : Effect of civic education on civic engagement by gender and SES (extensive margin)

	(1)	(2)	(3)	(4)
ATE	.384	.227	-.047	.039
	(.053)	(.100)	(.068)	(.010)
Observations	41861	37741	7591	31392
Switchers	398	280	21	306
Stayers	668	620	325	260
StateByType FE	Yes	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes
State Controls	Yes	Yes	Yes	Yes
Indiv. Controls	Yes	Yes	Yes	Yes
Sample	Female	Male	High SES	Low SES

Notes: This table presents the results from Equation 4.2 following De Chaisemartin and d'Haultfoeuille (2024) with with treatment $1\{d > 0\}$ by gender and socio-economic background. High SES: At least one parent holds an Abitur. Switchers are the treated groups that switch in or out of treatment and have a valid control group. Stayers are the valid control group with the same initial treatment dose whose treatment has not (yet) changed. State controls include the party in power at time of reform, compulsory school years, and G8. Individual controls include gender, individual born in Germany, and parental education. Standard errors are clustered at the state-school type level. Data: SOEP and own data on civic education.

Table A3.14 : Effect of civic education on civic engagement by gender and SES (intensive margin)

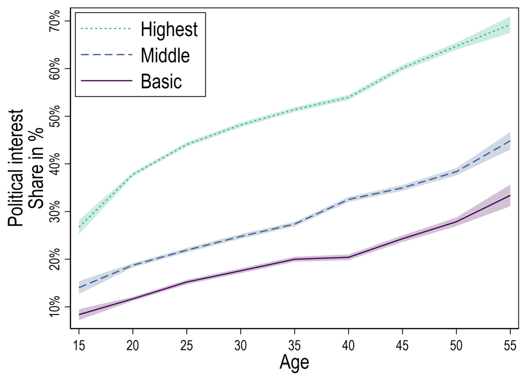
	(1)	(2)	(3)	(4)
ATE	-.045	.020	-.241	.055
	(.049)	(.042)	(.050)	(.038)
Observations	41861	37741	7591	31392
Switchers	975	916	76	665
Stayers	6797	4549	524	5526
StateByType FE	Yes	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes
State Controls	Yes	Yes	Yes	Yes
Indiv. Controls	Yes	Yes	Yes	Yes
Sample	Female	Male	High SES	Low SES

Notes: This table presents the results from Equation 4.2 following De Chaisemartin and d'Haultfoeuille (2024) with with treatment $1\{d > 0.7\}$ by gender and socio-economic background. High SES: At least one parent holds an Abitur. Switchers are the treated groups that switch in or out of treatment and have a valid control group. Stayers are the valid control group with the same initial treatment dose whose treatment has not (yet) changed. State controls include the party in power at time of reform, compulsory school years, and G8. Individual controls include gender, individual born in Germany, and parental education. Standard errors are clustered at the state-school type level. Data: SOEP and own data on civic education.

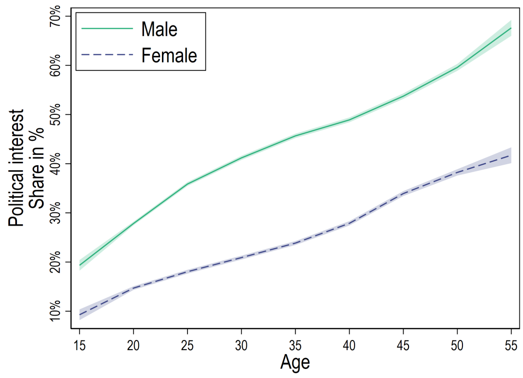
Appendix

A4.9 Appendix Figures and Tables

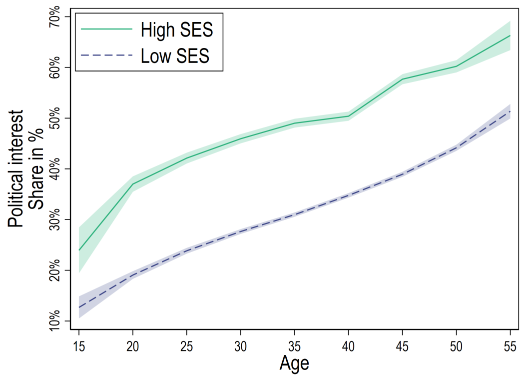
Figure A4.1 : General pattern of political interest by age.



(a) School track



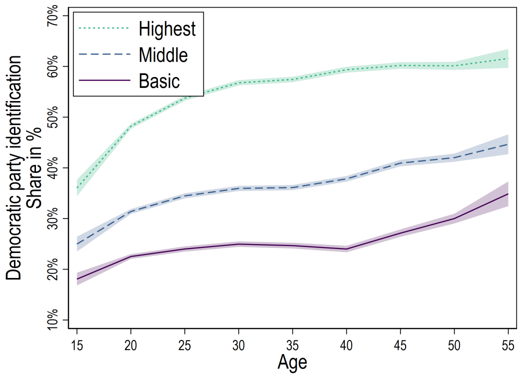
(b) Gender



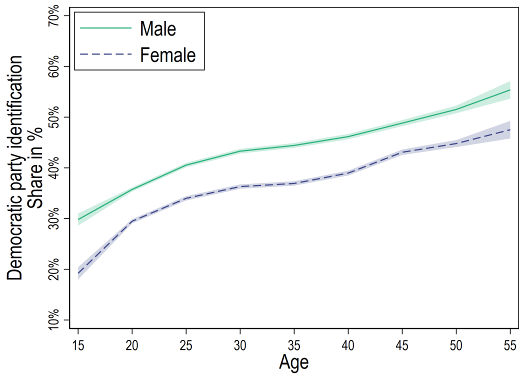
(c) SES

Notes: Political interest by five-year age groups with a 95 percent confidence band. All plots conditional on age group, standard errors clustered on individual level. High SES is defined as having at least one parent with an Abitur. Data: SOEP

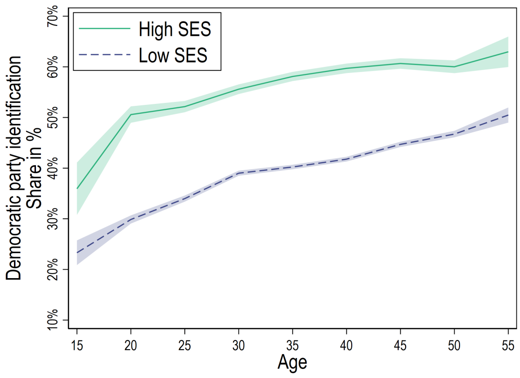
Figure A4.2 : General pattern of party identification by age.



(a) School track



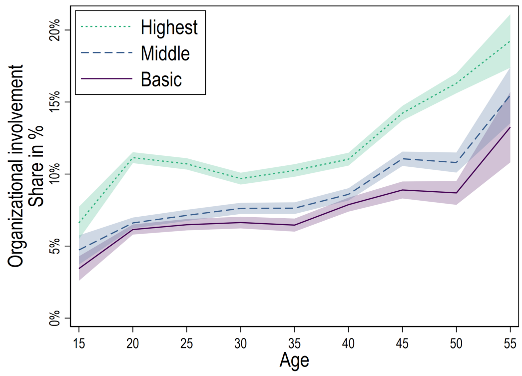
(b) Gender



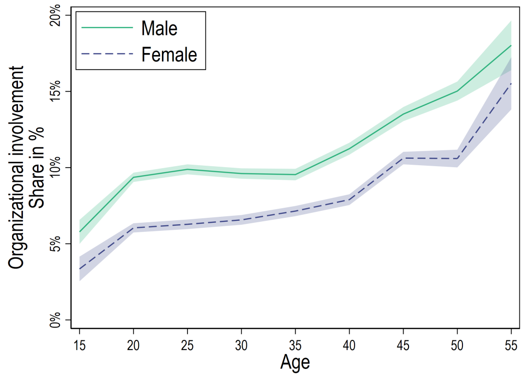
(c) SES

Notes: Party identification by five-year age groups with a 95 percent confidence band. All plots conditional on age group, standard errors clustered on individual level. High SES is defined as having at least one parent with an Abitur. Data: SOEP

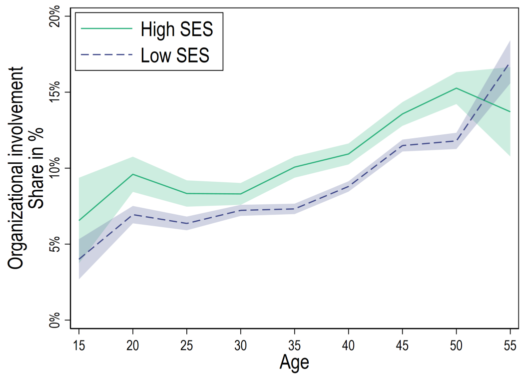
Figure A4.3 : General pattern of organizational involvement by age.



(a) School track



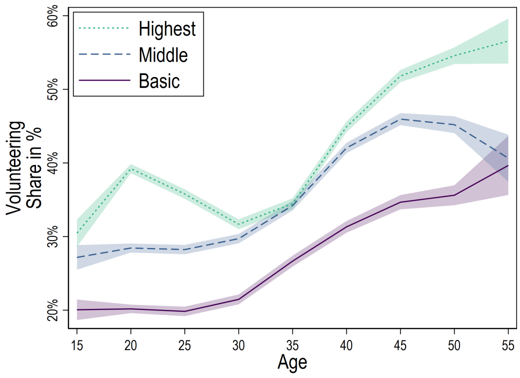
(b) Gender



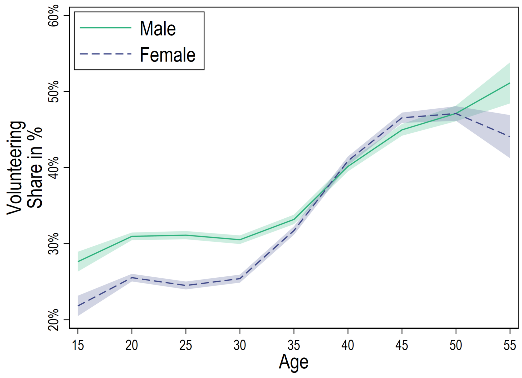
(c) SES

Notes: Organizational involvement by five-year age groups with a 95 percent confidence band. All plots conditional on age group, standard errors clustered on individual level. High SES is defined as having at least one parent with an Abitur. Data: SOEP

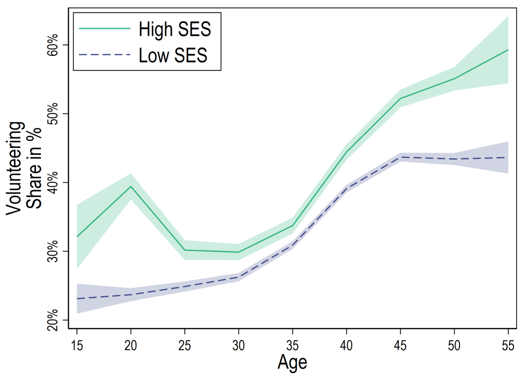
Figure A4.4 : General pattern of volunteering by age.



(a) School track



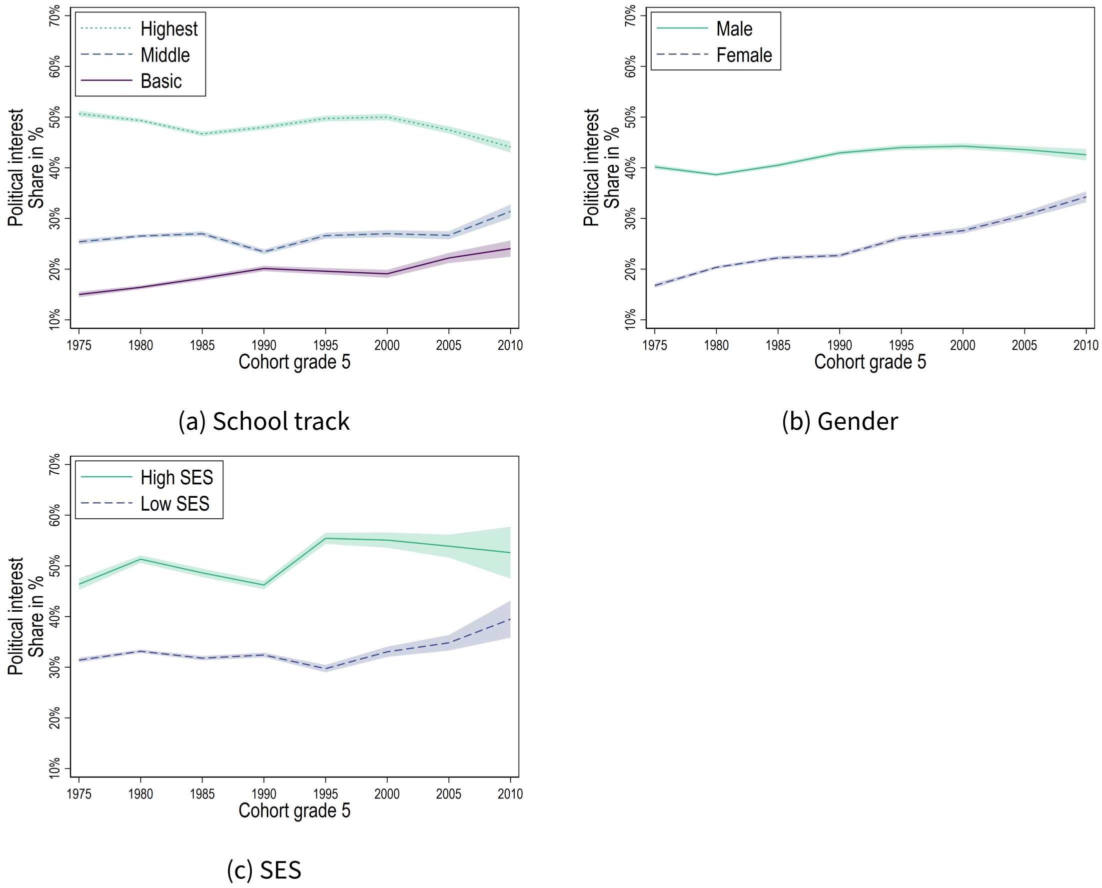
(b) Gender



(c) SES

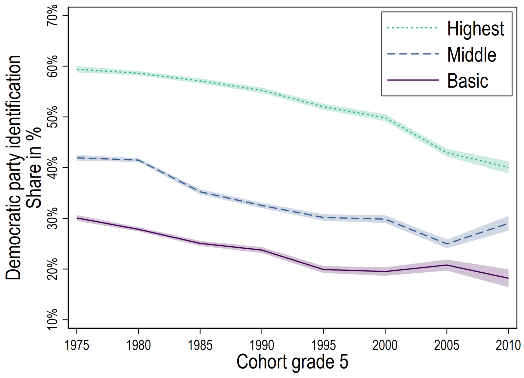
Notes: Volunteering by five-year age groups with a 95 percent confidence band. All plots conditional on age group, standard errors clustered on individual level. High SES is defined as having at least one parent with an Abitur. Data: SOEP

Figure A4.5 : General pattern of political interest by cohort.

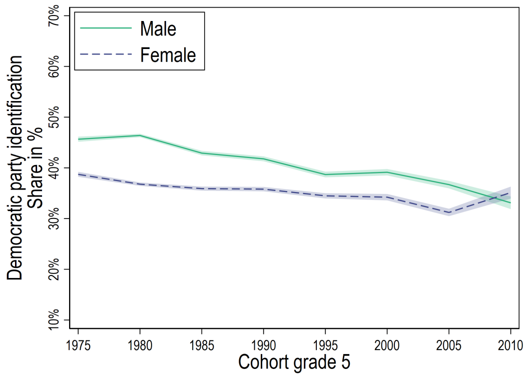


Notes: Political interest by five-year cohort groups with a 95 percent confidence band. All plots conditional on age group, standard errors clustered on individual level. High SES is defined as having at least one parent with an Abitur. Data: SOEP

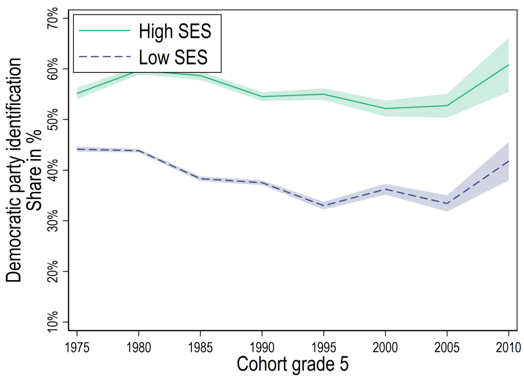
Figure A4.6 : General pattern of party identification by cohort.



(a) School track



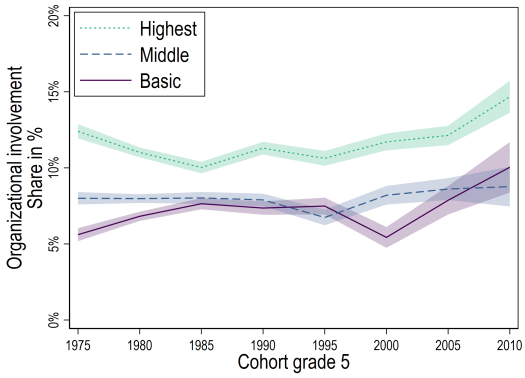
(b) Gender



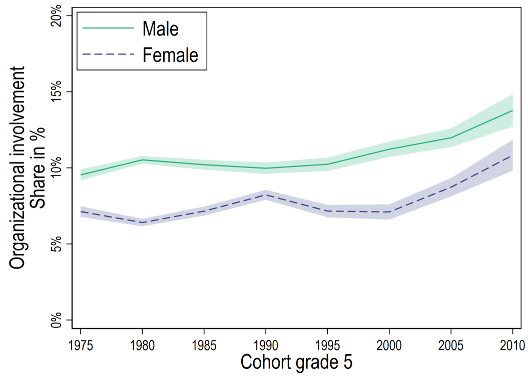
(c) SES

Notes: Party identification by five-year cohort groups with a 95 percent confidence band. All plots conditional on age group, standard errors clustered on individual level. High SES is defined as having at least one parent with an Abitur. Data: SOEP

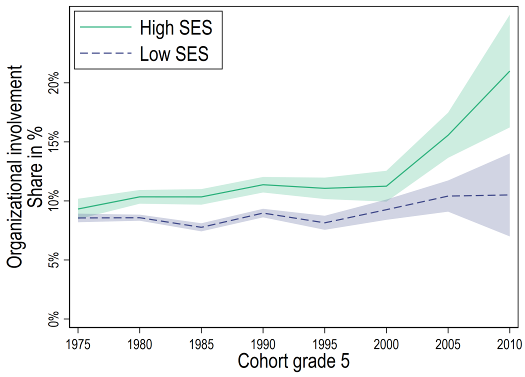
Figure A4.7 : General pattern of organizational involvement by cohort.



(a) School track



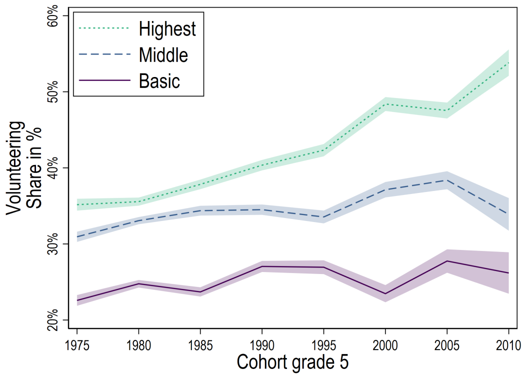
(b) Gender



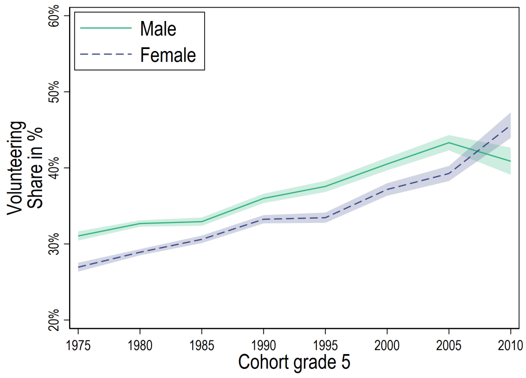
(c) SES

Notes: Organizational involvement by five-year cohort groups with a 95 percent confidence band. All plots conditional on age group, standard errors clustered on individual level. High SES is defined as having at least one parent with an Abitur. Data: SOEP

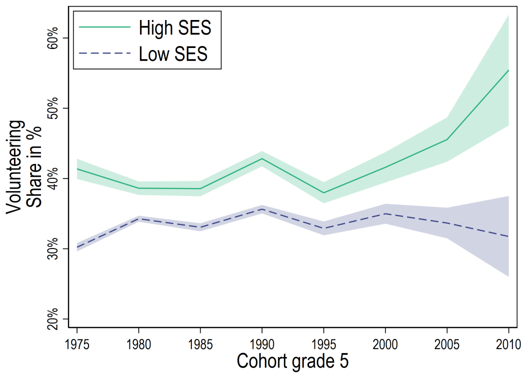
Figure A4.8 : General pattern of volunteering by cohort.



(a) School track



(b) Gender



(c) SES

Notes: Volunteering by five-year cohort groups with a 95 percent confidence band. All plots conditional on age group, standard errors clustered on individual level. High SES is defined as having at least one parent with an Abitur. Data: SOEP

Table A4.1 : Effect of civic education on civic engagement ignoring missing values (extensive margin)

	(1)	(2)	(3)	(4)
ATE	.186	.186	.184	.184
	(.058)	(.058)	(.058)	(.058)
Observations	161100	161100	161100	161100
Switchers	1306	1306	1306	1306
Stayers	2568	2568	2568	2568
StateByType FE	Yes	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes
State Controls	No	Yes	No	Yes
Indiv. Controls	No	No	Yes	Yes

Notes: This table presents the results from Equation 4.2 following De Chaisemartin and d'Haultfoeuille (2024) with treatment $1\{d > 0\}$. The dependent variable civic engagement is non-missing if at least one component variable is non-missing. Switchers are the treated groups that switch in or out of treatment and have a valid control group. Stayers are the valid control group with the same initial treatment dose whose treatment has not (yet) changed. State controls include the party in power at time of reform, compulsory school years, and G8. Individual controls include gender, individual born in Germany, and parental education. Standard errors are clustered at the state-school type level. Data: SOEP and own data on civic education.

Table A4.2 : Effect of civic education on civic engagement ignoring missing values (intensive margin)

	(1)	(2)	(3)	(4)
ATE	-.037	-.026	-.034	-.030
	(.030)	(.031)	(.031)	(.031)
Observations	161100	161100	161100	161100
Switchers	3611	3611	3611	3611
Stayers	27001	27001	27001	27001
StateByType FE	Yes	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes
State Controls	No	Yes	No	Yes
Indiv. Controls	No	No	Yes	Yes

Notes: This table presents the results from Equation 4.2 following De Chaisemartin and d'Haultfoeuille (2024) with treatment $1\{d > 0.7\}$. The dependent variable civic engagement is non-missing if at least one component variable is non-missing. Switchers are the treated groups that switch in or out of treatment and have a valid control group. Stayers are the valid control group with the same initial treatment dose whose treatment has not (yet) changed. State controls include the party in power at time of reform, compulsory school years, and G8. Individual controls include gender, individual born in Germany, and parental education. Standard errors are clustered at the state-school type level. Data: SOEP and own data on civic education.

Table A4.3 : Effect of civic education on political interest, democratic party identification, organizational involvement, and volunteering ignoring missing values (extensive margin)

	(1)	(2)	(3)	(4)
VARIABLE	Pol. Interest	Dem. Party Ident.	Org. Involv.	Volunt.
ATE	.384	-.173	.366	.359
	(.021)	(.107)	(.078)	(.066)
Observations	153084	149963	89780	89837
Switchers	1242	1211	776	774
Stayers	2504	2492	1369	1372
StateByType FE	Yes	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes
State Controls	Yes	Yes	Yes	Yes
Indiv. Controls	Yes	Yes	Yes	Yes

Notes: This table presents the results from Equation 4.2 with treatment $1\{d > 0\}$. The dependent variable ignores missing values in any other component variable. Switchers are the treated groups that switch in or out of treatment and have a valid control group. Stayers are the valid control group with the same initial treatment dose whose treatment has not (yet) changed. State controls include the party in power at time of reform, compulsory school years, and G8. Individual controls include gender, individual born in Germany, and parental education. Standard errors are clustered at the state-school type level. Data: SOEP and own data on civic education.

Table A4.4 : Effect of civic education on political interest, democratic party identification, organizational involvement, and volunteering ignoring missings (intensive margin)

	(1)	(2)	(3)	(4)
VARIABLE	Pol. Interest	Dem. Party Ident.	Org. Involv.	Volunt.
ATE	.014	-.120	.029	.020
	(.042)	(.048)	(.046)	(.044)
Observations	153084	149963	89780	89837
Switchers	3452	3389	2099	2099
Stayers	25595	25157	15313	15337
StateByType FE	Yes	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes
State Controls	Yes	Yes	Yes	Yes
Indiv. Controls	Yes	Yes	Yes	Yes

Notes: This table presents the results from Equation 4.2 with treatment $1\{d > 0.7\}$. The dependent variable ignores missing values in any other component variable. Switchers are the treated groups that switch in or out of treatment and have a valid control group. Stayers are the valid control group with the same initial treatment dose whose treatment has not (yet) changed. State controls include the party in power at time of reform, compulsory school years, and G8. Individual controls include gender, individual born in Germany, and parental education. Standard errors are clustered at the state-school type level. Data: SOEP and own data on civic education.

Table A4.5 : Effect of civic education on civic engagement, robustness for state approximation (extensive margin)

	(1)	(2)	(3)	(4)
ATE	.110	.110	.107	.106
	(.062)	(.062)	(.062)	(.062)
Observations	77270	77270	77270	77270
Switchers	675	675	675	675
Stayers	1265	1265	1265	1265
StateByType FE	Yes	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes
State Controls	No	Yes	No	Yes
Indiv. Controls	No	No	Yes	Yes

Notes: This table presents the results from Equation 4.2 following De Chaisemartin and d'Haultfoeuille (2024) with treatment $1\{d > 0\}$ excluding approximation with current state. Switchers are the treated groups that switch in or out of treatment and have a valid control group. Stayers are the valid control group with the same initial treatment dose whose treatment has not (yet) changed. State controls include the party in power at time of reform, compulsory school years, and G8. Individual controls include gender, individual born in Germany, and parental education. Standard errors are clustered at the state-school type level. Data: SOEP and own data on civic education.

Table A4.6 : Effect of civic education on civic engagement, robustness for state approximation (intensive margin)

	(1)	(2)	(3)	(4)
ATE	-.036	-.027	-.019	-.014
	(.031)	(.031)	(.027)	(.027)
Observations	77270	77270	77270	77270
Switchers	1865	1865	1865	1865
Stayers	13017	13017	13017	13017
StateByType FE	Yes	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes
State Controls	No	Yes	No	Yes
Indiv. Controls	No	No	Yes	Yes

Notes: This table presents the results from Equation 4.2 following De Chaisemartin and d'Haultfoeuille (2024) with treatment $1\{d > 0.7\}$ excluding approximation with current state. Switchers are the treated groups that switch in or out of treatment and have a valid control group. Stayers are the valid control group with the same initial treatment dose whose treatment has not (yet) changed. State controls include the party in power at time of reform, compulsory school years, and G8. Individual controls include gender, individual born in Germany, and parental education. Standard errors are clustered at the state-school type level. Data: SOEP and own data on civic education.

Table A4.7 : Effect of pure civic education on civic engagement.

	(1)	(2)	(3)	(4)
ATE	-.016	-.027	-.007	-.017
	(.043)	(.041)	(.041)	(.039)
Observations	79604	79604	79604	79604
Switchers	1028	1028	1028	1028
Stayers	6484	6484	6484	6484
StateByType FE	Yes	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes
State Controls	No	Yes	No	Yes
Indiv. Controls	No	No	Yes	Yes

Notes: This table presents the results from Equation 4.2 following De Chaisemartin and d'Haultfoeuille (2024) with treatment $1\{d > 0\}$ considering only civic education when it is not combined with other subjects. Switchers are the treated groups that switch in or out of treatment and have a valid control group. Stayers are the valid control group with the same initial treatment dose whose treatment has not (yet) changed. State controls include the party in power at time of reform, compulsory school years, and G8. Individual controls include gender, individual born in Germany, and parental education. Standard errors are clustered at the state-school type level. Data: SOEP and own data on civic education.

Table A4.8 : Drop one state at a time (extensive margin)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
State dropped	SH	HH	NI	HB	NW	HE	RP	BW	BY	SL
ATE	-.156	.240	.177	.240	.237	.240	.236	.246	.308	.193
	(.038)	(.062)	(.067)	(.062)	(.062)	(.062)	(.062)	(.062)	(.072)	(.063)
Observations	76642	79292	69227	78817	57288	72376	74412	65250	64213	78919
Switchers	100	678	136	678	678	678	678	678	578	642
Stayers	1185	1284	1264	1295	678	1162	1192	1019	74	1302
StateByType FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Indiv. Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table presents the results from Equation 4.2 following De Chaisemartin and d'Haultfoeuille (2024) with treatment $1\{d > 0\}$. Each column drops one state from estimation. Switchers are the treated groups that switch in or out of treatment and have a valid control group. Stayers are the valid control group with the same initial treatment dose whose treatment has not (yet) changed. State controls include the party in power at time of reform, compulsory school years, and G8. Individual controls include gender, individual born in Germany, and parental education. Standard errors are clustered at the state-school type level. Data: SOEP and own data on civic education.

Table A4.9 : Drop one state at a time (intensive margin)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
State dropped	SH	HH	NI	HB	NW	HE	RP	BW	BY	SL
ATE	-.016	-.014	-.026	-.013	-.025	-.034	-.046	.017	-.009	-.009
	(.025)	(.029)	(.026)	(.030)	(.038)	(.028)	(.036)	(.048)	(.037)	(.029)
Observations	76642	79292	69227	78817	57288	72376	74412	65250	64213	78919
Switchers	1891	1870	1158	1858	1798	1812	1663	1344	1747	1877
Stayers	13291	12674	12015	10591	1798	11108	10324	10506	11451	12441
StateByType FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Indiv. Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table presents the results from Equation 4.2 following De Chaisemartin and d'Haultfoeuille (2024) with treatment $1\{d > 0.7\}$. Each column drops one state from estimation. Switchers are the treated groups that switch in or out of treatment and have a valid control group. Stayers are the valid control group with the same initial treatment dose whose treatment has not (yet) changed. State controls include the party in power at time of reform, compulsory school years, and G8. Individual controls include gender, individual born in Germany, and parental education. Standard errors are clustered at the state-school type level. Data: SOEP and own data on civic education.

Bibliography

Acemoglu, Daron, Nicolas Ajzenman, Cevat Giray Aksoy, Martin Fiszbein, and Carlos A Molina (2021). *(Successful) democracies breed their own support*. Tech. rep. National Bureau of Economic Research.

Acemoglu, Daron, David Autor, Jonathon Hazell, and Pascual Restrepo (2022). "Artificial Intelligence and Jobs: Evidence from Online Vacancies". *Journal of Labor Economics* 40(S1), S293–S340.

Ahlskog, Rafael (2021). "Education and voter turnout revisited: Evidence from a Swedish twin sample with validated turnout data". *Electoral Studies* 69, 102186.

Almlund, Mathilde, Angela L. Duckworth, James Heckman, and Tim Kautz (2011). "Personality Psychology and Economics". *Handbook of the Economics of Education*. Ed. by Eric A. Hanushek, Ludger Woessmann, and Stephen Machin. Vol. 4. North Holland, 1–181.

Alscher, P., U. Ludewig, and N. McElvany (2022). "Civic Education, Teaching Quality and Students' Willingness to Participate in Political and Civic Life: Political Interest and Knowledge as Mediators". *Journal of Youth and Adolescence* 51(10). Epub 2022 Jun 1, 1886–1900.

Altinok, Nadir, Claude Diebolt, and Jean-Luc Demeulemeester (2014). "A New International Database on Education Quality: 1965–2010". *Applied Economics* 46(11), 1212–1247.

Angrist, Noam, Simeon Djankov, Pinelopi K. Goldberg, and Harry A. Patrinos (2021). "Measuring Human Capital Using Global Learning Data". *Nature* 592, 403–408.

Angrist, Noam, Harry A. Patrinos, and Martin Schlotter (2013). *An Expansion of a Global Data Set on Educational Quality*. Policy Research Working Paper 6536. Washington, DC: World Bank.

Apfeld, Brendan, Emanuel Coman, John Gerring, and Stephen Jessee (2022). "Education and social capital". *Journal of Experimental Political Science* 9(2), 162–188.

Arold, Benjamin, Ludger Woessmann, and Larissa Zierow (2022). "Can Schools Change Religious Attitudes? Evidence from German State Reforms of Compulsory Religious Education".

Arold, Benjamin W (2024). "Evolution vs. creationism in the classroom: The lasting effects of science education". *The Quarterly Journal of Economics*, qjae019.

Arouri, Mohamed, Cuong Nguyen, and Adel Ben Youssef (2015). "Natural Disasters, Household Welfare, and Resilience: Evidence from Rural Vietnam". *World Development* 70, 59–77.

ASER (2019). *Annual Status of Education Report (Rural) 2018*. Tech. rep. New Delhi: ASER Centre.

Ashenfelter, Orley C and David Card (1984). *Using the longitudinal structure of earnings to estimate the effect of training programs*.

Athey, Susan and Guido W Imbens (2022). "Design-based analysis in difference-in-differences settings with staggered adoption". *Journal of Econometrics* 226(1), 62–79.

Aucejo, Esteban M. and Teresa Foy Romano (2016). “Assessing the Effect of School Days and Absences on Test Score Performance”. *Economics of Education Review* 55, 70–87.

Auh, Jun Kyung, Jaewon Choi, Tatyana Deryugina, and Tim Park (2022). *Natural Disasters and Municipal Bonds*. Tech. rep. w30280. National Bureau of Economic Research.

Autor, David, David Figlio, Krzysztof Karbownik, Jeffrey Roth, and Melanie Wasserman (2016). “School Quality and the Gender Gap in Educational Achievement”. *American Economic Review* 106(5), 289–295.

– (2019). “Family Disadvantage and the Gender Gap in Behavioral and Educational Outcomes”. *American Economic Journal: Applied Economics* 11(3), 338–381.

Autor, David H. (2019). “Work of the Past, Work of the Future”. *AEA Papers and Proceedings* 109, 1–32.

Autor, David H., Frank Levy, and Richard J. Murnane (2003). “The Skill Content of Recent Technological Change: An Empirical Exploration”. *Quarterly Journal of Economics* 118(4), 1279–1333.

Azevedo, João Pedro, Diana Goldemberg, Silvia Montoya, Reema Nayar, Halsey Rogers, Jaime Saavedra, and Brian William Stacy (2021a). *Will every child be able to read by 2030? Defining learning poverty and mapping the dimensions of the challenge*. Policy Research Working Paper 9588. Washington, D.C.: World Bank Group.

Azevedo, João Pedro, Amer Hasan, Diana Goldemberg, Koen Geven, and Syedah Aroob Iqbal (2021b). “Simulating the potential impacts of COVID-19 school closures on schooling and learning outcomes: A set of global estimates”. *The World Bank Research Observer* 36(1), 1–40.

Banerjee, Abhijit V, Shawn Cole, Esther Duflo, and Leigh Linden (2007). “Remedying education: Evidence from two randomized experiments in India”. *The Quarterly Journal of Economics* 122(3), 1235–1264.

Barro, Robert J (1991). “Economic growth in a cross section of countries”. *The Quarterly Journal of Economics* 106(2), 407–443.

Barro, Robert J. and Jong Wha Lee (2015). *Education Matters: Global Schooling Gains from the 19th to the 21st Century*. Oxford: Oxford University Press.

Baumert, Jürgen and Anke Demmrich (2001). “Test Motivation in the Assessment of Student Skills: The Effects of Incentives on Motivation and Performance”. *European Journal of Psychology of Education* 16(3), 441–462.

Becker, Gary S (1962). “Investment in human capital: A theoretical analysis”. *Journal of Political Economy* 70(5, Part 2), 9–49.

Bertrand, Marianne and Jessica Pan (2013). “The Trouble with Boys: Social Influences and the Gender Gap in Disruptive Behavior”. *American Economic Journal: Applied Economics* 5(1), 32–64.

Bietenbeck, Jan, Marc Piopiunik, and Simon Wiederhold (2018). “Africa’s Skill Tragedy: Does Teachers’ Lack of Knowledge Lead to Low Student Performance?” *Journal of Human Resources* 53(3), 553–578.

Bilinski, Alyssa and Laura A Hatfield (2018). “Seeking evidence of absence: Reconsidering tests of model assumptions”. *arXiv preprint arXiv:1805.03273*.

Bils, Mark and Peter J. Klenow (2000). “Does Schooling Cause Growth?” *American Economic Review* 90(5), 1160–1183.

Black, Sandra E., Paul J. Devereux, and Kjell G. Salvanes (2007). “From the Cradle to the Labor Market? The Effect of Birth Weight on Adult Outcomes”. *The Quarterly Journal of Economics* 122(1), 409–439.

Blanden, Jo, Matthias Doepke, and Jan Stuhler (2023). “Educational Inequality”. *Handbook of the Economics of Education*. Vol. 6. Elsevier, 405–497.

Block, Michelle L. and Lilian Calderón-Garcidueñas (2009). “Air Pollution: Mechanisms of Neuroinflammation and CNS Disease”. *Trends in neurosciences* 32(9), 506–516.

BMFSFJ (2020). *16. Kinder-und Jugendbericht. Förderung demokratischer Bildung im Kindes-und Jugendalter*.

Borge, Julie Ane Odegaard (2017). “Tuning in to formal politics: Mock elections at school and the intention of electoral participation among first time voters in Norway”. *Politics* 37(2), 201–214.

Börsch-Supan, Axel (2000). “A Model under Siege: A Case Study of the German Retirement Insurance System”. *Economic Journal* 110(461), F24–F45.

Borusyak, Kirill, Xavier Jaravel, and Jann Spiess (2024). “Revisiting event-study designs: robust and efficient estimation”. *Review of Economic Studies*, rdae007.

Boustan, Leah Platt, Matthew E. Kahn, Paul W. Rhode, and Maria Lucia Yanguas (2020). “The Effect of Natural Disasters on Economic Activity in US Counties: A Century of Data”. *Journal of Urban Economics* 118, 103257.

Bowen, Daniel H and Brian Kisida (2020). “Never again: The impact of learning about the Holocaust on civic outcomes”. *Journal of Research on Educational Effectiveness* 13(1), 67–91.

Bronfenbrenner, Urie (1981). “Children and Families: 1984?” *Society* 18(2), 38–41.

Callahan, Rebecca M, Chandra Muller, and Kathryn S Schiller (2008). “Preparing for citizenship: Immigrant high school students’ curriculum and socialization”. *Theory & Research in Social Education* 36(2), 6–31.

Callaway, Brantly, Andrew Goodman-Bacon, and Pedro HC Sant’Anna (2024). *Difference-in-Differences with a Continuous Treatment*. Tech. rep. w32117. National Bureau of Economic Research.

Callaway, Brantly and Pedro HC Sant’Anna (2021). “Difference-in-Differences with Multiple Time Periods”. *Journal of Econometrics* 225(2), 200–230.

Campbell, Angus, Philip E. Converse, Warren E. Miller, and Donald E. Stokes (1960). *The American Voter*.

Campbell, David E and Richard G Niemi (2016). “Testing civics: State-level civic education requirements and political knowledge”. *American Political Science Review* 110(3), 495–511.

Chetty, Raj, John N Friedman, Nathaniel Hilger, Emmanuel Saez, Diane Whitmore Schanzenbach, and Danny Yagan (2011). “How does your kindergarten classroom affect your earnings? Evidence from Project STAR”. *The Quarterly Journal of Economics* 126(4), 1593–1660.

Ciccone, Antonio and Elias Papaioannou (2009). “Human Capital, the Structure of Production, and Growth”. *Review of Economics and Statistics* 91(1), 66–82.

Claassen, Christopher, Kathrin Ackermann, Eri Bertsou, Lucas Borba, Ryan E Carlin, Amnon Cavari, Sirianne Dahlum, Sergiu Gherghina, Darren Hawkins, Yphtach Lelkes, et al. (2023). “Conceptualizing and measuring support for democracy: A new approach”.

Comber, L. C. and John P. Keeves (1973). *Science Education in Nineteen Countries: An Empirical Study*. Stockholm: Almquist and Wiksell.

Currie, Janet, Eric A. Hanushek, E. Megan Kahn, Matthew Neidell, and Steven G. Rivkin (2009). “Does Pollution Increase School Absences?” *The Review of Economics and Statistics* 91(4), 682–694.

Currie, Janet and Maya Rossin-Slater (2013). “Weathering the Storm: Hurricanes and Birth Outcomes”. *Journal of Health Economics* 32(3), 487–503.

Cutter, Susan L., Lindsey Barnes, Melissa Berry, Christopher Burton, Elijah Evans, Eric Tate, and Jennifer Webb (2008). “A Place-Based Model for Understanding Community Resilience to Natural Disasters”. *Global Environmental Change* 18(4), 598–606.

Dahlgaard, Jens Olav (2018). “Trickle-Up Political Socialization: The Causal Effect on Turnout of Parenting a Newly Enfranchised Voter”. *American Political Science Review* 112(3), 698–705.

Dalton, Russell J (2016). “Party identification and its implications”. *Oxford Research Encyclopedia of Politics*.

Das, Jishnu and Tristan Zajonc (2010). “India Shining and Bharat Drowning: Comparing Two Indian States to the Worldwide Distribution in Mathematics Achievement”. *Journal of Development Economics* 92(2), 175–187.

De Chaisemartin, Clément and Xavier d’Haultfoeuille (2020). “Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects”. *American Economic Review* 110(9), 2964–96.

– (2024). “Difference-in-differences estimators of intertemporal treatment effects”. *Review of Economics and Statistics*, 1–45.

Dee, Thomas S. (2004). “Are there civic returns to education?” *Journal of Public Economics* 88(9), 1697–1720.

Deimel, Daniel (2023). *Kontexte der Genese politischer Partizipationsbereitschaft*. Springer.

Deimel, Daniel, Katrin Hahn-Laudenberg, Johanna F Ziemes, and Hermann J Abs (2024). “Civic education and social interactions at school as drivers of intended electoral participation: Similarities and differences in four European school systems”. *Education, Citizenship and Social Justice* 19(1), 37–57.

Deimel, Daniel, Bryony Hoskins, and Hermann J Abs (2021). “How do schools affect inequalities in political participation: Compensation of social disadvantage or provision of differential access?" *Civic Learning for Alienated, Disaffected and Disadvantaged Students*. Routledge, 6–26.

Deming, David J (2023). "Multidimensional human capital and the wage structure". *Handbook of the Economics of Education*, 469.

Deryugina, Tatyana (2017). "The Fiscal Cost of Hurricanes: Disaster Aid versus Social Insurance". *American Economic Journal: Economic Policy* 9(3), 168–198.

Detjen, Joachim (2013). *Politische Bildung: Geschichte und Gegenwart in Deutschland*. Oldenbourg Wissenschaftsverlag.

Deuchert, Eva and Christina Felfe (2015). "The Tempest: Short-and Long-Term Consequences of a Natural Disaster for Children's Development". *European Economic Review* 80, 280–294.

Di Pietro, Giorgio (2018). "The Academic Impact of Natural Disasters: Evidence from L'Aquila Earthquake". *Education Economics* 26(1), 62–77.

Dompnier, Nathalie and Raul Magni Berton (2012). "How durably do people accept democracy? Politicization, political attitudes and losers' consent in France". *French Politics* 10(4), 323–344.

Donbavand, Steven and Bryony Hoskins (2021). "Citizenship Education for Political Engagement: A Systematic Review of Controlled Trials". *Social Sciences* 10(5), 151.

Dubeck, Margaret M. and Amber Gove (2015). "The Early Grade Reading Assessment (EGRA): Its Theoretical Foundation, Purpose, and Limitations". *International Journal of Educational Development* 40, 315–322.

Educational Initiatives (2012). *Student Learning Study: Status of Student Learning across 18 States of India in Urban and Rural Schools*. Tech. rep. Educational Initiatives.

Emmers, Dorien, Qi Jiang, Hao Xue, Yue Zhang, Yunting Zhang, Yingxue Zhao, Bin Liu, Sarah-Eve Dill, Yiwei Qian, Nele Warrinnier, Hannah Johnstone, Jianhua Cai, Xiaoli Wang, Lei Wang, Renfu Luo, Guirong Li, Jiajia Xu, Ming Liu, Yaqing Huang, Wenjie Shan, Zhihui Li, Yu Zhang, Sean Sylvia, Yue Ma, Alexis Medina, and Scott Rozelle (2021). "Early Childhood Development and Parental Training Interventions in Rural China: A Systematic Review and Meta-Analysis". *BMJ Global Health* 6(8), e005578.

Engzell, Per, Arun Frey, and Mark D Verhagen (2021). "Learning Loss Due to School Closures during the COVID-19 Pandemic". *Proceedings of the National Academy of Sciences* 118(17), e2022376118.

Fahle, Erin M, Belen Chavez, Demetra Kalogrides, Benjamin R Shear, Sean F Reardon, and Andrew D Ho (2021). *Stanford Education Data Archive (Version 4.1)*. Tech. rep.

Filmer, Deon, Amer Hasan, and Lant Pritchett (2006). *A Millennium Learning Goal: Measuring Real Progress in Education*. Working Paper 97. Washington, DC: Center for Global Development.

Filmer, Deon, Halsey Rogers, Noam Angrist, and Shwetlena Sabarwal (2020). "Learning-adjusted years of schooling (LAYS): Defining a new macro measure of education". *Economics of Education Review* 77, 101971.

Foster, Andrew D. and Mark R. Rosenzweig (1996). “Technical Change and Human-Capital Returns and Investments: Evidence from the Green Revolution”. *American Economic Review* 86(4), 931–953.

Freyaldenhoven, Simon, Christian Hansen, and Jesse M Shapiro (2019). “Pre-event trends in the panel event-study design”. *American Economic Review* 109(9), 3307–3338.

Fuchs-Schündeln, Nicola and Matthias Schündeln (2015). “On the endogeneity of political preferences: Evidence from individual experience with democracy”. *Science* 347(6226), 1145–1148.

Fuente, Ángel de la and Rafael Doménech (2024). “Cross-country data on skills and the quality of schooling: A selective survey”. *Journal of Economic Surveys* 38(1), 3–26.

Fuller, Sarah C. (2014). “The Effect of Prenatal Natural Disaster Exposure on School Outcomes”. *Demography* 51(4), 1501–1525.

Gagel, Walter (2005). *Geschichte der politischen Bildung in der Bundesrepublik Deutschland 1945-1989/90*. Springer.

Gao, Qiufeng, Huan Wang, Fang Chang, Hongmei Yi, and Yaojiang Shi (2021). “Reading Achievement in China’s Rural Primary Schools: A Study of Three Provinces”. *Educational Studies* 47(3), 344–368.

Gardner, John (2022). “Two-Stage Differences in Differences”. *Working Paper*.

Gneezy, Uri, John A. List, Jeffrey A. Livingston, Xiangdong Qin, Sally Sadoff, and Yang Xu (2019). “Measuring Success in Education: The Role of Effort on the Test Itself”. *American Economic Review: Insights* 1(3), 291–308.

Gökbudak, Mahir and Reinhold Hedtke (2018). “Ranking Politische Bildung 2017”.

Goldin, Claudia and Lawrence F Katz (2020). *The Incubator of Human Capital: The NBER and the Rise of the Human Capital Paradigm*. Tech. rep. w26909. National Bureau of Economic Research.

Goodman, Joshua (2014). *Flaking out: Student Absences and Snow Days as Disruptions of Instructional Time*. Tech. rep. w20221. National Bureau of Economic Research.

Goodman-Bacon, Andrew (2021). “Difference-in-Differences with Variation in Treatment Timing”. *Journal of Econometrics* 225(2), 254–277.

Green, Donald P, Peter M Aronow, Daniel E Bergan, Pamela Greene, Celia Paris, and Beth I Weinberger (2011). “Does knowledge of constitutional principles increase support for civil liberties? Results from a randomized field experiment”. *The Journal of Politics* 73(2), 463–476.

Gruijters, Rob J. (2020). *How (Un)Representative Are China’s Stellar PISA Results?* In NORRAG Blog.

Gust, Sarah, Eric A. Hanushek, and Ludger Woessmann (2024). “Global Universal Basic Skills: Current Deficits and Implications for World Development”. *Journal of Development Economics* 166, 103205.

Hall, Robert E. and Charles I. Jones (1999). “Why Do Some Countries Produce So Much More Output Per Worker Than Others?” *Quarterly Journal of Economics* 114(1), 83–116.

Hamm, Ingrid, Anna Engelke, Maja Finke, Diana Kinnert, Thomas Krüger, Mirko Meyerding, Armin Nassehi, Andreas Schleicher, Linda Teuteberg, and Ludger Wößmann (2023). *Mehr und besser: Vorschläge für eine Demokratiebildung von morgen*. Report of the Hertie Commission Democracy and Education.

Hanushek, Eric A (1986). “The economics of schooling: Production and efficiency in public schools”. *Journal of economic literature* 24(3), 1141–1177.

– (2020). “Education production functions”. *The Economics of Education*. Elsevier, 161–170.

Hanushek, Eric A and Dennis D Kimko (2000). “Schooling, labor-force quality, and the growth of nations”. *American Economic Review* 90(5), 1184–1208.

Hanushek, Eric A., Lavinia Kinne, Philipp Lergetporer, and Ludger Woessmann (2022). “Patience, Risk-Taking, and Human Capital Investment across Countries”. *Economic Journal* 132(646), 2290–2307.

Hanushek, Eric A., Jens Ruhose, and Ludger Woessmann (2017a). “Economic Gains from Educational Reform by US States”. *Journal of Human Capital* 11(4), 447–486.

– (2017b). “Knowledge Capital and Aggregate Income Differences: Development Accounting for U.S. States”. *American Economic Journal: Macroeconomics* 9(4), 184–224.

Hanushek, Eric A. and Ludger Woessmann (2008). “The Role of Cognitive Skills in Economic Development”. *Journal of Economic Literature* 46(3), 607–668.

– (2011). “How Much Do Educational Outcomes Matter in OECD Countries?” *Economic Policy* 26(67), 427–491.

– (2012a). “Do Better Schools Lead to More Growth? Cognitive Skills, Economic Outcomes, and Causation”. *Journal of Economic Growth* 17(4), 267–321.

– (2012b). “Schooling, Educational Achievement, and the Latin American Growth Puzzle”. *Journal of Development Economics* 99(2), 497–512.

– (2015a). *The Knowledge Capital of Nations: Education and the Economics of Growth*. Cambridge, MA: MIT Press.

– (2015b). *Universal Basic Skills: What Countries Stand to Gain*. Paris: Organisation for Economic Co-operation and Development.

– (2016). “Knowledge Capital, Growth, and the East Asian Miracle”. *Science* 351(6271), 344–345.

– (2020a). “A Quantitative Look at the Economic Impact of the European Union’s Educational Goals”. *Education Economics* 28(3), 225–244.

– (2020b). *The Economic Impacts of Learning Losses*. Paris: Organisation for Economic Co-operation and Development.

Harka, Elona and Lorenzo Rocco (2022). “Studying more to vote less. Education and voter turnout in Italy”. *European Journal of Political Economy* 72, 102115.

Helbig, Marcel and Rita Nikolai (2015). *Die Unvergleichbaren: der Wandel der Schulsysteme in den deutschen Bundesländern seit 1949*. Julius Klinkhardt.

Hener, Timo, Helmut Rainer, and Thomas Siedler (2016). "Political socialization in flux?: linking family non-intactness during childhood to adult civic engagement". *Journal of the Royal Statistical Society Series A: Statistics in Society* 179(3), 633–656.

Herrlitz, Hans-Georg, Wulf Hopf, Hartmut Titze, and Ernst Cloer (2005). *Deutsche Schulgeschichte von 1800 bis zur Gegenwart: Eine Einführung*. 4., überarb. u. aktual. Aufl. Weinheim: Juventa Verlag, 292.

Hoskins, Bryony, Lihong Huang, and Cecilia Arensmeier (2021). "Socioeconomic inequalities in civic learning in Nordic Schools: Identifying the potential of in-school civic participation for disadvantaged students". *Northern lights on civic and citizenship education: A cross-national comparison of Nordic data from ICCS*, 93–122.

Hoskins, Bryony, Jan Germen Janmaat, Christine Han, and Daniel Muijs (2016). "Inequalities in the education system and the reproduction of socioeconomic disparities in voting in England, Denmark and Germany: The influence of country context, tracking and self-efficacy on voting intentions of students age 16–18". *Compare: A Journal of Comparative and International Education* 46(1), 69–92.

Hoxby, Caroline M (2021). "Advanced cognitive skill deserts in the United States: their likely causes and implications". *Brookings Papers on Economic Activity* 2021(1), 317–351.

Hoyos, Rafael de, Ricardo Estrada, and María José Vargas (2021). "What Do Test Scores Really Capture? Evidence from a Large-Scale Student Assessment in Mexico". *World Development* 146, 105524.

Huneke, Samuel (2024). "How Germany's Political Stability May Be Fueling the Rise of the Far Right". *Time Magazine*. Made by History.

Ichino, Andrea and Rudolf Winter-Ebmer (2004). "The long-run educational cost of World War II". *Journal of Labor Economics* 22(1), 57–87.

Irmert, Natalie, Kaja Mummert, Rita Nikolai, Lukas Mergele, and Larissa Zierow (2023). "Wer leitet die deutschen Bildungsministerien? Ein neuer Datensatz zur Analyse von biografischen Merkmalen von Bildungsministern der Bundesländer, 1950-2020". *Zeitschrift für Politikwissenschaft* 33(3).

Jahan, Ferdous A, Nicole Zviedrite, Hongjiang Gao, Faruque Ahmed, and Amra Uzicanin (2022). "Causes, Characteristics, and Patterns of Prolonged Unplanned School Closures Prior to the COVID-19 Pandemic—United States, 2011–2019". *PLoS One* 17(7), e0272088.

Jamison, Dean T. and Lawrence J. Lau (1982). *Farmer Education and Farm Efficiency*. Baltimore, MD: The Johns Hopkins University Press.

Jamison, Dean T. and Peter R. Moock (1984). "Farm Education and Farm Efficiency in Nepal: The Role of Schooling, Extension Services and Cognitive Skills". *World Development* 12(1), 67–86.

Janmaat, Jan Germen and Bryony Hoskins (2022). "The changing impact of family background on political engagement during adolescence and early adulthood". *Social Forces* 101(1), 227–251.

Janmaat, Jan Germen, Tarek Mostafa, and Bryony Hoskins (2014). "Widening the participation gap: The effect of educational track on reported voting in England". *Journal of Adolescence* 37(4), 473–482.

Jaume, David and Alexander Willén (2019). "The Long-Run Effects of Teacher Strikes: Evidence from Argentina". *Journal of Labor Economics* 37(4), 1097–1139.

Jennings, M Kent, Laura Stoker, and Jake Bowers (2009). "Politics across generations: Family transmission reexamined". *The Journal of Politics* 71(3), 782–799.

Jerch, Rhiannon, Matthew E. Kahn, and Gary C. Lin (2023). "Local Public Finance Dynamics and Hurricane Shocks". *Journal of Urban Economics* 134, 103516.

Jung, Jilli and Maithreyi Gopalan (2023). "The Stubborn Unresponsiveness of Youth Voter Turnout to Civic Education: Quasi-experimental Evidence from State-Mandated Civics Tests". *Educational Evaluation and Policy Analysis*, 01623737231195887.

Kahn-Lang, Ariella and Kevin Lang (2020). "The promise and pitfalls of differences-in-differences: Reflections on *16 and pregnant* and other applications". *Journal of Business & Economic Statistics* 38(3), 613–620.

Kalla, Joshua and Ethan Porter (2022). "Can the Political Ambition of Young Women Be Increased? Evidence from U.S. High School Students". *Quarterly Journal of Political Science* 17(2), 259–281.

Kalsnes, Bente (2023). "Fake news and digital politics". *Handbook of Digital Politics*. Edward Elgar Publishing, 432–471.

Karbownik, Krzysztof and Anthony Wray (2019). "Long-Run Consequences of Exposure to Natural Disasters". *Journal of Labor Economics* 37(3), 949–1007.

Kerbel, Barbara (2016). "Wie der sozialistische Staat die Bildungseinrichtungen prägte". *BPB* 2, 2017.

Kinne, Lavinia (2023). *Good or bad news first? The effect of feedback order on motivation and performance*. Tech. rep. 396. ifo Working Paper.

Kitanova, Magdelina (2020). "Youth political participation in the EU: Evidence from a cross-national analysis". *Journal of Youth Studies* 23(7), 819–836.

Klauber, Hannah, Felix Holub, Nicolas Koch, Nico Pestel, Nolan Ritter, and Alexander Rohlf (2024). "Killing prescriptions softly: Low emission zones and child health from birth to school". *American Economic Journal: Economic Policy* 16(2), 220–248.

Koch, Alexander, Julia Nafziger, and Helena S. Nielsen (2015). "Behavioral economics of education". *Behavioral Economics of Education* 115, 3–17.

Kostelka, Filip and André Blais (2021). "The Generational and Institutional Sources of the Global Decline in Voter Turnout". *World Politics* 73(4), 629–667.

Kotschy, Rainer and Uwe Sunde (2022). "Does Demography Determine Democratic Attitudes?"

Krueger, Alan B. and Mikael Lindahl (2001). "Education for Growth: Why and for Whom?" *Journal of Economic Literature* 39(4), 1101–1136.

Kuhn, Hans-Werner (2013). *Politische Bildung in Deutschland: Entwicklung - Stand - Perspektiven*. Springer-Verlag.

Lavy, Victor (2015). "Do Differences in Schools' Instruction Time Explain International Achievement Gaps? Evidence from Developed and Developing Countries". *The Economic Journal* 125(588), F397–F424.

Lawrence, Michael, Thomas Homer-Dixon, Scott Janzwood, Johan Rockstöm, Ortwin Renn, and Jonathan F. Donges (2024). "Global polycrisis: the causal mechanisms of crisis entanglement". *Global Sustainability* 7, e6.

Lazear, Edward P. (2001). "Educational Production". *The Quarterly Journal of Economics* 116(3), 777–803.

Lee, Hanol and Jong Wha Lee (2020). *Educational Quality and Disparities in Income and Growth across Countries*. Mimeo.

Lee, Jong–Wha and Robert J Barro (2001). "Schooling quality in a cross–section of countries". *Economica* 68(272), 465–488.

Lee, Jong-Wha and Hanol Lee (2016). "Human capital in the long run". *Journal of Development Economics* 122, 147–169.

Lim, Stephen S., Rachel L. Updike, Alexander S. Kaldjian, Ryan M. Barber, Krycia Cowling, Hunter York, Joseph Friedman, R. Xu, Joanna L. Whisnant, Heather J. Taylor, Andrew T. Leever, Yesenia Roman, Miranda F. Bryant, Joseph Dieleman, Emmanuela Gakidou, and Christopher J. L. Murray (2018). "Measuring Human Capital: A Systematic Analysis of 195 Countries and Territories, 1990-2016". *The Lancet* 392(10154), 1217–1234.

Lindgren, Karl-Oskar, Sven Oskarsson, and Mikael Persson (2019). "Enhancing Electoral Equality: Can Education Compensate for Family Background Differences in Voting Participation?" *American Political Science Review* 113(1), 108–122.

Loveless, Tom (2014). *How Well Are American Students Learning? With Sections on the PISA-Shanghai Controversy, Homework, and the Common Core*. The 2014 Brown Center Report on American Education. Washington, DC: Brookings Institution.

– (2017). *How Well Are American Students Learning? With Sections on the Latest International Test Scores, Foreign Exchange Students, and School Suspensions*. The 2017 Brown Center Report on American Education. Washington, DC: Brookings Institution.

Lutz, Wolfgang, Raya Muttarak, and Erich Striessnig (2014). "Universal Education Is Key to Enhanced Climate Adaptation". *Science* 346(6213), 1061–1062.

Mankiw, N. Gregory, David Romer, and David Weil (1992). "A Contribution to the Empirics of Economic Growth". *Quarterly Journal of Economics* 107(2), 407–437.

Manning, Nathan and Kathryn Edwards (2014a). "Why Has Civic Education Failed to Increase Young People's Political Participation?" *Sociological Research Online* 19(1), 1–12.

Manning, Nathan and Kathy Edwards (2014b). “Does civic education for young people increase political participation? A systematic review”. *Educational review* 66(1), 22–45.

Marcotte, Dave E. (2007). “Schooling and Test Scores: A Mother-Natural Experiment”. *Economics of Education Review* 26(5), 629–640.

Marquis, Christopher and András Tilcsik (2013). “Imprinting: Toward a Multilevel Theory”. *Academy of Management Annals* 7(1), 195–245.

Masozera, Michel, Melissa Bailey, and Charles Kerchner (2007). “Distribution of Impacts of Natural Disasters across Income Groups: A Case Study of New Orleans”. *Ecological economics* 63(2-3), 299–306.

Meyer, Andrew G (2017). “The impact of education on political ideology: Evidence from European compulsory education reforms”. *Economics of Education Review* 56, 9–23.

Milligan, Kevin, Enrico Moretti, and Philip Oreopoulos (2004). “Does education improve citizenship? Evidence from the United States and the United Kingdom”. *Journal of public Economics* 88(9-10), 1667–1695.

Mincer, Jacob (1974). “Schooling, Experience, and Earnings. Human Behavior & Social Institutions No. 2.”

Moore, Mark A., Anthony E. Boardman, Aidan R. Vining, David L. Weimer, and David Greenberg (2004). “’Just Give Me a Number!’ Practical Values for the Social Discount Rate”. *Journal of Policy Analysis and Management* 23(4), 789–812.

Mullis, Ina V.S., Michael O. Martin, Pierre Foy, Dana L. Kelly, and Bethany Fishbein (2020). *TIMSS 2019 International Results in Mathematics and Science*. Chestnut Hill, MA: Boston College, TIMSS & PIRLS International Study Center, and International Association for the Evaluation of Educational Achievement (IEA).

National Bureau of Statistics of the People’s Republic of China (2010). *National Statistics Yearbook 2010*.

– (2015). *One-Percent National Sample Census*.

– (2016). *The 2016 Report on China’s Migrant Population Development*. Beijing: China Population Publishing Co.

Nelsen, Matthew D. (2021). “Teaching Citizenship: Race and the Behavioral Effects of American Civic Education”. *The Journal of Race, Ethnicity, and Politics* 6(1), 157–186.

Neundorf, Anja, Richard G Niemi, and Kaat Smets (2016). “The compensation effect of civic education on political engagement: How civics classes make up for missing parental socialization”. *Political Behavior* 38, 921–949.

Nguyen, Cuong Viet and Nguyet Minh Pham (2018). “The Impact of Natural Disasters on Children’s Education: Comparative Evidence from Ethiopia, India, Peru, and Vietnam”. *Review of Development Economics* 22(4), 1561–1589.

Noury, Abdul and Gerard Roland (2020). “Identity Politics and Populism in Europe”. *Annual Review of Political Science* 23(Volume 23, 2020), 421–439.

OECD (2016). *Skills Matter: Further Results from the Survey of Adult Skills*. Paris: Organisation for Economic Co-operation and Development.

– (2018a). *PISA for Development: Results in Focus*. Results in Focus 91. Paris: Organisation for Economic Co-operation and Development.

– (2018b). *PISA-D in-School Assessment Technical Report*. Tech. rep. Paris: Organisation for Economic Co-operation and Development.

– (2019). *PISA 2018 Results (Volume I)*. Paris: Organisation for Economic Co-operation and Development.

– (2020a). “How’s Life? Measuring Well-being”. *OECD Publishing, Paris*.

– (2020b). *PISA for Development: Out-of-School-Assessment*. Results in Focus 110. Paris: Organisation for Economic Co-operation and Development.

– (2020c). *PISA-D out-of-School Assessment Technical Report*. Tech. rep. Paris: Organisation for Economic Co-operation and Development.

Opper, Isaac M., R. Jisung Park, and Lucas Husted (2023). “The Effect of Natural Disasters on Human Capital in the United States”. *Nature Human Behaviour*, 1–12.

Otache, Ifeanyi, Isaiah P. Alfa, and Yakubu Ali (2023). “Civic education and voting intentions among the electorate: the roles of attitudes towards voting behaviour and educational level”. *Transforming Government: People, Process and Policy* 17(3), 387–398.

Özek, Umut (2023). “Examining the educational spillover effects of severe natural disasters: The case of Hurricane Maria”. *Journal of Human Resources* 58(2), 421–451.

Patel, Dev and Justin Sandefur (2020). *A Rosetta Stone for Human Capital*. Tech. rep. 550. Washington, DC: Center for Global Development.

Persson, Mikael, Klas Andersson, Pär Zetterberg, Joakim Ekman, and Simon Lundin (2020). “Does deliberative education increase civic competence? Results from a field experiment”. *Journal of Experimental Political Science* 7(3), 199–208.

Persson, Mikael, Karl-Oskar Lindgren, and Sven Oskarsson (2016). “How does education affect adolescents’ political development?” *Economics of Education Review* 53, 182–193.

Pischke, Jörn-Steffen (2007). “The Impact of Length of the School Year on Student Performance and Earnings: Evidence from the German Short School Years”. *The Economic Journal* 117(523), 1216–1242.

Pleninger, Regina (2022). “Impact of Natural Disasters on the Income Distribution”. *World Development* 157, 105936.

Pritchett, Lant (2013). *The Rebirth of Education: Schooling Ain’t Learning*. Washington, DC: Center for Global Development.

Pritchett, Lant and Martina Viarengo (2023). “The Learning Crisis of Developing Country Elites: Lessons from PISA-D”. *World Bank Economic Review* 37(2), 177–204.

Putnam, Robert D (2000). *Bowling alone: The collapse and revival of American community*. Simon and Schuster.

Quintelier, Ellen (2015). “Engaging adolescents in politics: The longitudinal effect of political socialization agents”. *Youth & Society* 47(1), 51–69.

Quintelier, Ellen, Marc Hooghe, and Sofie Marien (2011). “The effect of compulsory voting on turnout stratification patterns: a cross-national analysis”. *International Political Science Review* 32(4), 396–416.

Quintelier, Ellen and Jan W Van Deth (2014). “Supporting democracy: Political participation and political attitudes. Exploring causality using panel data”. *Political Studies* 62(1_suppl), 153–171.

Reardon, Sean F., John P. Papay, Tara Kilbride, Katherine O. Strunk, Joshua Cowen, Lily An, and Kate Donohue (2019). “Can Repeated Aggregate Cross-Sectional Data Be Usedto Measure Average Student Learning Rates? A Validation Study of Learning Rate Measures in the Stanford Education Data Archive. CEPA Working Paper No. 19-08.” *Stanford Center for Education Policy Analysis*.

Reichert, Frank and Murray Print (2018). “Civic participation of high school students: The effect of civic learning in school”. *Educational Review* 70(3), 318–341.

Reinhardt, Sibylle (2017). “The Beutelsbach Consensus”. *Mending Walls: Historical, Socio? Political, Economic, and Geographical Perspectives*. Publisher: IAP, 25.

Remington, Patrick L, Bridget B Catlin, and Keith P Gennuso (2015). “The County Health Rankings: Rationale and Methods”. *Population Health Metrics* 13(1), 11.

Resnjanskij, Sven, Jens Ruhose, Simon Wiederhold, Ludger Woessmann, and Katharina Wedel (2024). “Can mentoring alleviate family disadvantage in adolescence? A field experiment to improve labor market prospects”. *Journal of Political Economy* 132(3), 1013–1062.

Rinnooy Kan, Willemijn F, Anke Munniksma, Monique Volman, and Anne Bert Dijkstra (2023). “Practicing voice: student voice experiences, democratic school culture and students’ attitudes towards voice”. *Research Papers in Education*, 1–21.

Roth, Jonathan (2022). “Pretest with Caution: Event-Study Estimates after Testing for Parallel Trends”. *American Economic Review: Insights* 4(3), 305–322.

Rustad, Siri Aas (2024). “Conflict Trends: A Global Overview, 1946–2023”. *PRIO Paper*.

Sacerdote, Bruce (2012). “When the Saints Go Marching out: Long-term Outcomes for Student Evacuees from Hurricanes Katrina and Rita”. *American Economic Journal: Applied Economics* 4(1), 109–35.

Sampermans, Dorien, Ellen Claes, and Jan Germen Janmaat (2021). “Back on track? How civic learning opportunities widen the political knowledge gap in a tracked education system”. *School Effectiveness and School Improvement* 32(2), 241–259.

Sandefur, Justin (2018). “Internationally Comparable Mathematics Scores for Fourteen African Countries”. *Economics of Education Review* 62, 267–286.

Sander, Wolfgang (2005). “Die Bildungsstandards vor dem Hintergrund der politikdidaktischen Diskussion”. *Redaktionen Politische Bildung & kursiv-Journal für politische Bildung*

(Hrsg.), Bildungsstandards. Evaluation in der politischen Bildung (S. 38-54). Schwalbach/Ts.: Wochenschau.

Schneider, Mark (2019). *The Strange Case of 'China' and Its Top PISA Rankings – How Cherry-Picking Regions to Take Part Skews Its High Scores*. The74Million.org.

Schoner, Florian, Lukas Mergele, and Larissa Zierow (2024). "Grading student behavior". *Labour Economics*, 102570.

Schultz, Theodore W. (1961). "Investment in Human Capital". *The American Economic Review* 51(1), 1–17.

Sendzik, Norbert, Ulrike Mehnert, and Marcel Helbig (2024). "Feuerwehr der Demokratie? Politische Bildung als Unterrichtsfach an allgemeinbildenden Schulen der Sekundarstufe I in der Bundesrepublik Deutschland von 1949 bis 2019". *Working Paper*.

Siedler, Thomas (2010). "Schooling and citizenship in a young democracy: Evidence from postwar Germany". *Scandinavian Journal of Economics* 112(2), 315–338.

Smith, Adam B. (2022). "2021 U.S. Billion-Dollar Weather and Climate Disasters in Historical Context". *Technical Report*.

Sondheimer, Rachel Milstein and Donald P Green (2010). "Using experiments to estimate the effects of education on voter turnout". *American Journal of Political Science* 54(1), 174–189.

Stadelmann-Steffen, Isabelle and Linda Sulzer (2018). "Just another brick in the wall? The relationship between classroom-based political education and the political interest of young adults in Switzerland". *Journal of Youth Studies* 21(4), 550–573.

Steinmann, Isa, Rolf Strietholt, and Wilfried Bos (2014). *Linking International Comparative Student Assessment*. Tech. rep. LINCS Technical Report. International Association for the Evaluation of Educational Achievement (IEA).

Sun, Liyang and Sarah Abraham (2021). "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects". *Journal of Econometrics* 225(2), 175–199.

Syvertsen, Amy K, Michael D Stout, Constance A Flanagan, Dana L Mitra, Mary Beth Oliver, and S Shyam Sundar (2009). "Using elections as teachable moments: A randomized evaluation of the student voices civic education program". *American Journal of Education* 116(1), 33–67.

Theocharis, Yannis and Jan W Van Deth (2018). "The continuous expansion of citizen participation: A new taxonomy". *European political science review* 10(1), 139–163.

Tillmann, Klaus-Jürgen (2020). "Von der Lehrerbedarfsprognose zum Seiteneinstieg – bildungspolitische Anmerkungen zur gegenwärtigen Versorgungskrise". *DDS–Die Deutsche Schule* 112(4), 439–453.

UNESCO (2021). *Global Education Monitoring Report 2021/2: Non-State Actors in Education - Who Chooses? Who Loses?* Paris: United Nations Educational, Scientific and Cultural Organization.

UNESCO Institute for Statistics (2022). *Rosetta Stone Analysis Report: Establishing a Concordance between PASEC and TIMSS/PIRLS*. Montréal: UNESCO-UIS.

UNESCO, UNICEF and Weltbank (2021). *The state of the global education crisis: a path to recovery*. United Nations Educational, Scientific and Cultural Organization.

United Nations (2022). *World Population Prospects 2022*. New York: United Nations.

Verba, Sidney, Kay Lehman Schlozman, and Henry E Brady (1995). *Voice and equality: Civic voluntarism in American politics*. Harvard University Press.

Wang, Lei, Wenbin Min, Siqi Zhang, Yaojiang Shi, and Scott Rozelle (2018). "Math and Chinese-Language Learning: Where Are China's Vulnerable Subpopulations?" *Asian Survey* 58(5), 797–821.

Wedel, Katharina (2021). "Instruction time and student achievement: The moderating role of teacher qualifications". *Economics of Education Review* 85, 102183.

Weiss, Julia (2020). "What is youth political participation? Literature review on youth political participation and political attitudes". *Frontiers in Political Science* 2, 1.

Welch, Finis (1970). "Education in Production". *Journal of Political Economy* 78(1), 35–59.

Werner, Katharina and Ludger Woessmann (2023). "The legacy of COVID-19 in education". *Economic Policy* 38(115), 609–668.

Woessmann, Ludger (2005). "Educational production in Europe". *Economic Policy* 20(43), 446–504.

– (2016a). "The Economic Case for Education". *Education Economics* 24(1), 3–32.

– (2016b). "The Importance of School Systems: Evidence from International Differences in Student Achievement". *Journal of Economic Perspectives* 30(3), 3–32.

Wolfers, Justin (2006). "Did Unilateral Divorce Laws Raise Divorce Rates? A Reconciliation and New Results". *American Economic Review* 96(5), 1802–1820.

Wooldridge, Jeffrey M. (2021). "Two-Way Fixed Effects, the Two-Way Mundlak Regression, and Event Study Estimators".

World Bank (2018). *World Development Report 2018: Learning to Realize Education's Promise*. Washington, DC: World Bank.

– (2019). *World Development Report 2019: The Changing Nature of Work*. Washington, DC: World Bank.

Wu, Ge (2022). *Do Wildfilres Harm Student Learning?* Tech. rep. University of Tennessee, Department of Economics.

Zamarro, Gema, Collin Hitt, and Ildefonso Mendez (2019). "When Students Don't Care: Reexamining International Differences in Achievement and Student Effort". *Journal of Human Capital* 13(4), 519–552.

Zhao, Qiran, Xiaobing Wang, and Scott Rozelle (2019). "Better Cognition, Better School Performance? Evidence from Primary Schools in China". *China Economic Review* 55, 199–217.