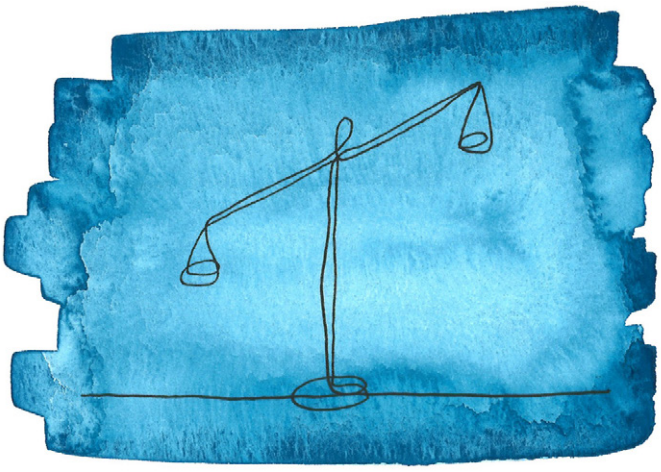
ACTA UNIVERSITATIS GOTHOBURGENSIS
GOTHENBURG STUDIES IN EDUCATIONAL SCIENCES 475



Teacher sorting and the opportunity gap

A cross-national investigation of institutional differentiation and educational equity

Leah Natasha Glassow

UNIVERSITY OF GOTHENBURG

Teacher sorting and the opportunity gap

ACTA UNIVERSITATIS GOTHOBURGENSIS
GOTHENBURG STUDIES IN EDUCATIONAL SCIENCES 475

Teacher sorting and the opportunity gap

A cross-national investigation of institutional differentiation and educational equity

Leah Natasha Glassow

UNIVERSITY OF GOTHENBURG

ISBN 978-91-7963-127-7 (printed)
ISBN 978-91-7963-128-4 (pdf)
ISSN 0436-1121

The publication is also available in full text at:
http://hdl.handle.net/2077/74350

Doctoral dissertation in Education at the Department of Education and Special Education, University of Gothenburg

Subscriptions to the series and orders for individual copies sent to: Acta Universitatis Gothoburgensis, PO Box 222, SE-405 30 Göteborg, Sweden or to acta@ub.gu.se

Cover artist: Lea Eldstål-Ahrens

Photographer: Jacob Pettersson

Print:
Stema Specialtryck AB, Borås, 2023

SVANENMÄRKET
Trycksak
3041 0234

Abstract

Title:	Teacher sorting and the opportunity gap: a cross-national investigation of institutional differentiation and educational equity
Author:	Leah Natasha Glassow
Language:	English
ISBN:	978-91-7963-127-7 (printed)
ISBN:	978-91-7963-128-4 (pdf)
ISSN:	0436-1121
Keywords:	teacher sorting, teacher quality, equity, student achievement, TIMSS, TALIS, educational policy

Inequitable access to teacher competence ('teacher sorting' or the teacher 'opportunity gap') is increasingly the focus of international educational bodies worldwide but is still relatively underexplored empirically. The overarching purpose of this doctoral thesis is to investigate the relationship between teacher sorting and educational inequity from a cross-national perspective, while empirically addressing theoretical questions related to social reproduction and inequality of educational opportunities in school systems. A final aim is to provide empirically grounded policy recommendations related to the findings. With these aims in mind, the constituent studies in the thesis cover several facets of the phenomenon of teacher sorting: the magnitude and development cross-nationally over the past two decades, the impact on inequity in student test scores, as well as the associated institutional features.

The data come from international large-scale assessments such as the Trends in International Mathematics and Science Study (TIMSS, 1999-2019) and the Teaching and Learning International Survey (TALIS, 2018), and include 32 and 46 education systems, respectively, with a special focus on mathematics and science teachers. The main analytical approaches include descriptive statistical methods, panel data regressions with country fixed effects, and hierarchical generalized linear modelling.

The dissertation is comprised of four empirical studies. Study I investigates the magnitude of teacher sorting cross-nationally as well as its development since 1999. Results show that the magnitude of inequity varies by the country and teacher qualification in focus. Few countries show widening inequities in the teacher qualification gaps. Study II investigates the impact of teacher sorting on

mathematics achievement inequity and finds that more pronounced sorting by specialization exacerbates inequity in student achievement, and that this finding remains marginally significant after controlling for increasing socioeconomic school segregation. Studies III and IV investigate policy- and institution-level correlates of teacher sorting and teacher turnover, respectively. The results of Study III show a general pattern of mixed results related to stratification, accountability, autonomy, and competition, depending on the teacher quality indicator in focus. National economic development level as well as school competition were positively related to the slope on more than one occasion, however. Study IV found a more pronounced relationship between teacher turnover intentions and classroom SES in school systems with more widespread use of external accountability practices with student performance data.

The results point to several key conclusions. First, there was evidence of inequity in teacher sorting across many educational systems to varying degrees. The patterns varied depending on how teacher qualifications and socioeconomic status were measured as well as how students were grouped. Next, the studies provided mixed results regarding school autonomy, accountability, competition and stratification, indicating that the determinants of socioeconomic teacher sorting do not easily generalize according to cross-national patterns. School competition was the single system-level variable to be associated with both qualifications. Despite this, performance data-based accountability (teacher appraisal) was consistently associated with higher turnover intention rates in low-SES settings. Appraisal of teacher performance for those working in lower-SES classrooms should rely on metrics other than performance data and should be conducted by those with appropriate knowledge of the school context. With respect to inequity in student outcomes, socioeconomic teacher sorting by specialization was found to have a modest effect. In most cases, priority should be given to democratizing access to teachers with appropriate content knowledge, but educational systems must go beyond providing socioeconomically disadvantaged students with teachers with basic qualification levels. While incentivizing the most competent teachers to work in socioeconomically disadvantaged settings is an ongoing challenge for many educational systems, building upon the content knowledge of underqualified mathematics teachers currently working in hard-to-staff settings is a worthwhile endeavor. Last, while reducing teacher sorting by specialization is likely to alleviate some degree of inequity in educational outcomes, it is not a panacea in the wider context of rising income inequality and social segregation in many educational systems.

Contents

Study I - IV

List of tables and figures

To Jacob

Acknowledgements

My sincerest thanks go first to my supervisors. Dr. Kajsa Yang Hansen has been my main supervisor. Thank you for the many hours you spent with me on data and methodological issues. I felt completely supported throughout these last several years. Dr. Jan-Eric Gustafsson has been my co-supervisor. I am extremely grateful for your breadth of expertise and knowledge, and continued support throughout these last four (and a half) years. Smart academics are a dime a dozen, but kind academics are less common, and I have been blessed with not one but two incredibly talented supervisors who are also very humble and kind.

This thesis was part of the OCCAM project, which was an invaluable part of my training and competence development. I was hosted by the FUR research group at Gothenburg University. Through these environments, I had the privilege to learn from world class researchers and professors and meet wonderful friends. Dr. Monica Rosén in particular contributed to our OCCAM supervision meetings. Dr. John Jerrim, the co-author for my first study, was a fantastic mentor throughout my secondment hosted by UCL. Another special thank you goes to my dear friend Emilie Franck from OCCAM who was a collaborator and co-author on Study III and sounding board throughout the PhD. Emilie, I am so grateful for you. I would also like to thank Dr. Erika Majoros, my OCCAM counterpart at GU, who has been a special addition to my academic and personal life. We learned from each other along the way, cried together, laughed together, and I will always be grateful for your wisdom, friendship, and approach to life.

I would also like to thank the discussants in my planning, mid, and final seminars, Dr. Trude Nilsen, Dr. Stefan Johansson, and Dr. Sigrid Blömeke, for their detailed and thoughtful commentary which greatly contributed to the thesis. I also extend my gratitude to the defence opponent, Dr. Leonidas Kyriakides. There is also Lea Eldstål-Ahrens who executed the cover art flawlessly.

I started thinking seriously that I might want to do a PhD in 2014 and worked towards that goal until 2018 when I accepted this position. However, the decades of schooling before 2014 also got me here, and I owe those years to my family. My Mum and Dad greatly emphasized education at home, tirelessly supported my goals and believed in me. My mum, Karen Glassow, has especially inspired me. A

teacher herself, she has had immeasurable impacts on her students. This must have osmotically influenced me and my interest in education. One student in particular has said that her classroom was the only place he felt at home[1]. My Swedish family has also greatly contributed to this doctoral project. Claes and Annika, this quite literally would not have been possible without you and your support. Daniel, Julia, Kate and Johan, I cherish our conversations, laughs, and belief in one another. Thank you to my Toronto support network, A/C/T/T/M, and too many others to name.

Finally, there is one person, to whom this thesis is dedicated, who has been my biggest champion, my best friend, my rock, and that is my husband, Jacob Pettersson. There should be a support group for spouses of people doing PhDs. Jacob, you have endured the late nights, the tears, the stress, the doubt, and you have believed in me and helped me to believe in me. Finishing my PhD while postpartum (not to mention during a global pandemic) has been a balancing act for our family. This has been my dream since we have been together, and you have helped me achieve it in more ways than I can count. Thank you.

Gothenburg, December 2022

[1] https://www.cbc.ca/listen/live-radio/1-82-here-and-now-toronto/clip/15945761-i-just-called-say-thank-you-cameron-bailey

Chapter 1 Introduction

This doctoral thesis began with a basic wish to understand a relatively simple question: do all students, regardless of socioeconomic background, have access to teachers of similar competence, and does this matter for their academic achievement? Which has since evolved to include a follow-up question: what can education systems do to democratize such access? To answer these two questions, I turned to the cross-national and cross-temporal information provided by the international large-scale assessment datasets. First, the Trends in International Mathematics and Science Study (TIMSS), and next, the Teaching and Learning International Survey (TALIS). In exploring these two fundamental questions, I have had to grapple with and pin down concepts such as how to validly measure teacher competence across socioeconomic groups and cultural contexts, how to conceptualize educational (in)equity, and how to differentiate statistical associations from causal descriptions and explanations. This dissertation therefore deals in some capacity with all of these—among many other—concepts. In what follows I present an introduction to the topic in general as well as the overarching aims and purpose of the project.

There is a persistent link between socioeconomic status (SES)—the relative social, economic, and cultural position of students—and their performance on tests which aim to assess their ability levels in mathematics, science, reading, and other subjects (Chmielewski, 2019; Coleman et al., 1966). This 'achievement gap' is present in virtually all countries participating in the large-scale assessment studies. To better position the full impact of this claim, it is worth highlighting the purpose and development of many educational systems across the world to begin with. Since at least 1948, the most prominent international bodies have declared education as a 'human right' (United Nations, 1948). This movement towards the democratization of education occurred in the case of more liberal welfare state systems as well as more traditionally socially democratic systems, such as the Nordic countries (Arnesen & Lundahl, 2006), but is less obvious in systems which are not explicitly meritocratic (Chmielewski, 2019). Nevertheless, there appears to be a movement towards focusing more on equity in education globally (OECD, 2018a). Strong schooling has been consistently positioned as a sort of 'equalizer'

between individuals of different backgrounds, and crucial for countries wishing to take a preventative approach to economic inequality (Hanushek & Woessmann, 2015; Solga, 2014). For countries, strong schooling is associated with a productive workforce and economic growth; and for individuals, upward social mobility, higher levels of adult numeracy and literacy, and greater outcomes in health and wellbeing (Gustafsson, 2016; Hanushek & Woessmann, 2015). However, since at least the publication of the Coleman Report (1966), theoretical and empirical criticisms of the view of schooling as the 'great equalizer' have been accumulating (Bourdieu, 1977; Lavrijsen & Nicaise, 2016; Strello, Strietholt, Steinmann, & Siepmann, 2021). Studies have documented inequitable allocation of funding and school resources, differential effects of reforms such as school choice or tracking, and the growing issue of neighbourhood and social segregation in Europe and beyond (Burger, 2019; Marcinczak et al., 2016). Meanwhile, the proliferation of neoliberal ideology and international large-scale assessments have led to widespread global educational reforms prioritizing accountability and results, at times neglecting issues of equity (Biesta, 2007; Lee & Wong, 2003). The conclusion of many, therefore, is that schooling and school systems not only reproduce societal inequities but exacerbate them.

The persistence of the socioeconomic achievement gap has forced researchers and policymakers alike to face the reality that the myriad of educational reforms and amendments of the past decades have not adequately addressed the problem. Consequently, they have been forced to look elsewhere. Recently, certain international bodies such as the Organization for Economic Cooperation and Development (OECD, 2018b) have turned to teacher competence as one such potentially crucial resource which has been inequitably allocated across students. This phenomenon has been referred to as teacher sorting, allocation, and distribution.

Teacher competence is consistently cited as one of the few important educational inputs which can be altered that is particularly consequential for disadvantaged students (Darling-Hammond, 2000; Goe, 2007; Hattie, 2003; Nye, Hedges & Konstantopoulos, 2004; Rivkin, Hanushek & Kain, 2005). This is especially true in mathematics and science (Baumert et al., 2010; Goe, 2007), as students tend to do most of their learning in these subjects at school. While much of the research has focused on individual countries such as the United States, cross-national comparative studies have underscored the importance of teacher competence as well (Gustafsson & Nilsen, 2016). The consensus is that teachers matter, but there remains some trepidation as to which teacher and teaching

characteristics are most consequential for student outcomes. The importance of certain characteristics may vary across educational contexts (Jackson, Rockoff & Staiger, 2014; Rockoff, Kane, Jacob, & Staiger, 2011). There are several decades of research attempting to measure and define the concept of a 'competent teacher,' and a split in opinions regarding appropriate methods for such investigations (Bitler, Corcoran, Domina & Penner, 2019; Hanushek & Rivkin, 2012). Some of these disagreements may have to do with the fact that educational research is inherently multidisciplinary. While certain approaches may value causation above all else, others may prioritize construct validity and the equivalence of measurement properties across contexts.

Despite the long-documented importance of teachers, we know very little about how they are 'sorted' across socioeconomic groups, and even less about how such sorting differs across countries. On the one hand, there is the issue of attracting teachers to lower-SES settings, and on the other, there is the issue of *retaining* them. These two dimensions, attraction and retention, are investigated throughout the thesis via the average qualification levels across students as well as an indicator of turnover intentions. This problem has just recently come more into focus for the international testing bodies (OECD, 2018b). Akiba, Letendre and Scribner, who coin the term teacher 'opportunity gap' and find that teacher competence (measured by qualifications) is inequitably distributed across socioeconomic groups in many countries, first examined the phenomenon in 2007. Since then, however, less than a handful of studies have taken a cross-national approach to the issue, although many from the United States have long engaged with the question (Clotfelder, Ladd & Vigdor, 2005; Darling-Hammond, 2004a). There is therefore first and foremost a need to document the extent of this problem across national educational contexts. Next, there are unanswered empirical questions about whether such sorting really matters for student achievement.

In addition to this, there is a parallel line of questioning which is relevant to the problem of sorting. Namely, which institutional determinants are relevant? This area of research again comes mostly out of the United States. There is evidence that teachers with higher competence levels tend to seek out jobs in more 'desirable' settings (that is, in classrooms or schools with a higher share of socioeconomically advantaged students, Jackson, 2009). There are also labour market constraints, such as more job openings in socioeconomically disadvantaged schools (Sims & Allen, 2018). In regards to reforms and policy, the discussion has generally been limited to factors at the state or local level, focused on teacher

training, evaluation, retention, and pay (Jackson et al., 2014). There are many unanswered questions about the role of institutional policies in the sorting of teachers across settings. Educational reforms such as the No Child Left Behind (2001) act in the United States have included dimensions related to accountability with the explicit intention of equalizing access to teacher quality but often education reforms lack sufficient within-country variability. While some recent work has examined the relationship between teacher sorting and institutional characteristics such as school autonomy and tracking (Han, 2018; Luschei & Jeong, 2019), there is a need to replicate these findings in more settings and with additional characteristics related to teacher competence, as well as other institutional features.

To sum up, there is an ongoing interest in empirically evaluating the 'critical' view of schooling and school systems (Downey & Condron, 2016) or the line of questioning which examines whether schooling perpetuates inequities. More specifically, there is a need to employ innovative measures and rigorous empirical research methods to contribute to the scarce body of cross-national literature examining the effect of and factors associated with the socioeconomic sorting of teacher competence.

Research aims and outline

The primary aim of the thesis is to close the gap in the cross-national literature regarding the extent of inequitable teacher sorting and its impact on student achievement inequity. There is an equally important secondary aim, which is to provide school leaders and policymakers with information about institutional features associated with inequities in the attraction and retention of teachers in certain educational settings, and to make policy recommendations based on these findings.

The thesis is comprised of an integrating essay along with four empirical studies. The studies can be seen as two sets of pairs which are related. The first two studies (I and II) deal with the primary aim listed above and the second two studies (III and IV) deal with the secondary aim. They correspond to the following overarching research questions:

1. Study I: To what extent do education systems display teacher sorting across socioeconomic student groups, and has this changed over the past two decades?

2. Study II: Does the sorting of teacher qualifications by student socioeconomic background exacerbate the relationship between student socioeconomic status and student mathematics achievement?
3. Study III: Which institutional features are associated with inequitable teacher sorting?
4. Study IV: Are student performance-data driven accountability practices, such as performance-based teacher appraisal, associated with increases in teacher turnover intentions?

Each of the research questions are answered through their corresponding empirical studies in the results and discussion section.

Study I utilizes TIMSS data (1999-2019) to investigate the magnitude and development of mathematics teacher qualification 'opportunity gaps' over a twenty-year period across 32 educational systems. It is a descriptive study, which tracks first which education systems exhibit teacher sorting in the TIMSS 2019 cycle, and next where among the education systems are sorting inequalities trending upwards or downwards since 1999.

Study II also uses TIMSS data (1999-2019) and builds on the system-level measures of sorting inequity conceptualized in the first study. It extends the line of questioning to examine the effect of teacher sorting on mathematics achievement inequity. The study employs a panel regression approach with country fixed effects for 32 education systems. Given the assumption of no time-invariant confounders in such models, the analysis also considers the influence of changes in socioeconomic school segregation on achievement inequity.

Study III uses TALIS data from the 2018 cycle to investigate the associations between institutional features and teacher sorting across 46 education systems. Institutional features examined include between school tracking, ability grouping within schools, school autonomy, data-driven school accountability, school competition, and teacher shortages at the national level. The study employs three-level generalized hierarchical linear modelling with a random slope.

Study IV also uses TALIS data from the 2018 cycle across 46 education systems to estimate whether the relationship between teacher turnover and classroom socioeconomic status varies as a function of system-level teacher appraisal practices. More specifically, it examines whether using student test scores for teacher evaluation (by either external authorities or a school management team) is associated with turnover intentions in low-SES classrooms. The study employs three-level generalized hierarchical linear modelling with a random slope.

The measures related to student achievement deal strictly with the mathematics domain. Therefore, in the first two studies, only TIMSS mathematics teachers are included. The final two studies are more inclusive and examine teachers from a variety of educational backgrounds but do not include student achievement data. Each of the studies deals with the comparison of many educational systems, and therefore much emphasis was placed on the comparability of measures across educational settings.

In the integrating essay, the conceptual framework and simultaneous literature review consists of three chapters. The chapter 'Theoretical points of departure' outlines the theoretical basis and framework for the investigation. The following chapter 'An overview of the phenomenon of teacher sorting' presents the problem of teacher sorting in full as well as a literature review on the topic. The last chapter of the background section focuses on 'institutional determinants of teacher sorting' and introduces the country-level features of interest. Following these three chapters comes the section on 'Methodology,' where data, variables, and methods are presented. The 'Methodology' chapter also outlines challenges and threats related to validity and reliability of the measures and statistical analyses. Thereafter comes the section 'Results and discussion,' which outlines the findings from each empirical study and discusses them with additional detail and context, as well as how they relate to one another. The 'Concluding remarks' chapter relays a number of methodological challenges, limitations, and directions for future research. Last, a Swedish summary is presented as well as each of the four empirical studies in full.

Guide for readers

Table 1 Guide for readers

Studies		Data	Explanatory variables	Dependent variable
1	Is inequitable teacher sorting on the rise? Cross-national evidence from 20 years of TIMSS	TIMSS 1999, 2003, 2007, 2011, 2015, 2019	Student SES	Teacher qualifications
2	Does socioeconomic sorting of teacher qualifications exacerbate mathematics achievement inequity? Panel data estimates from 20 years of TIMSS	TIMSS 1999, 2003, 2007, 2011, 2015, 2019	Teacher qualification gaps	Mathematics achievement inequity
3	Institutional characteristics moderating the relationship between classroom socioeconomic composition and teacher qualifications: Evidence from 46 education systems in TALIS 2018	TALIS 2018	School stratification, autonomy, accountability, competition, HDI	Teacher qualification-classroom SES slope
4	Inequitable teacher turnover and performance-based appraisal: A global trend?	TALIS 2018	School accountability	Teacher turnover intentions-classroom SES slope

Chapter 2 Theoretical points of departure

Developments in the educational effectiveness research field

While determining the quality of educational systems is at the center of much political discourse today, there is a long history of research concerned with educational effectiveness (Creemers, 2005; Scheerens & Bosker, 1997). The roots of educational effectiveness can be traced back to the Coleman report (1966) as well as a well-known study by Jencks et al. (1972). Both studies aimed to examine the amount of variation in student test scores which could be attributed to school-related factors alone outside of the home environment. Both studies, according to Creemers (2005), come to a similar conclusion, which is that school-related factors accounted for a surprisingly low proportion of the outcome variation after individual student factors such as family background were under control. This shift in perspective, according to Creemers (2005), occurs around the time that policymakers and scholars started to take note of the failure of large-scale compensatory educational programs, such as the Headstart program in the United States. Since around this time, different methodological and theoretical attempts at chronicling educational and school effectiveness factors have developed around the world (Creemers & Schaveling, 1985; Kyriakides, 2007; Kyriakides, Creemers & Charambolous, 2018; Lundgren, 1972; Rutter et al., 1979; Scheerens & Bosker, 1997).

The studies isolated several factors shown to be correlated with student outcomes, including the ability composition of students in the school, the school climate, teacher characteristics and behavior, and strong school leadership (Creemers, 2005; Creemers & Kyriakides, 2007). However, the earlier attempts were mired with several methodological flaws. Creemers (2005) emphasizes that the field in general has been heavily focused on methodology with a naturalistic point of departure, and that much of the studies tend to be 'atheoretical' in nature (p. 4). Since then, it has developed to include more sound theoretical dimensions

as well as methodological improvements. Kyriakides et al. (2018) differentiate between two main effectiveness dimensions. The first, related to quality—having to do with average test scores—and the second, related to equity—having to do with the compensatory power of schools and school systems to mitigate the link between student outcomes and their socioeconomic backgrounds. The survey items included in the international large-scale assessments have evolved to reflect this diversity in potential outcomes. Another criticism of the educational effectiveness field highlights the disconnect between documenting effectiveness and contributing to school and school system improvement (Biesta, 2007). This has mainly to do with ecological validity. That is, how generalizable are the findings from educational effectiveness studies across different educational contexts? While much research has remained within national borders, the growth of cross-national tests and comparisons of education systems has largely to do with answering this question.

Policy evaluation with international large-scale assessment data

Cycle-based international large-scale assessment studies such as the Trends in International Mathematics and Science Study (TIMSS) began in the latter half of the 20th century, with TIMSS occurring every four years since 1995, with earlier studies such as FIMS and SIMS beginning in the 1960s. The assessments focus on a variety of age groups, most commonly fourth and eighth grade or the country's equivalent grades. TIMSS questionnaires are based on subject matter taught within a country's curriculum. In 2000, the Organization for Economic Co-operation and Development (OECD) conducted its first cycle-assessment called the Programme for International Student Assessment (PISA), happening subsequently every three years since. In contrast to TIMSS, for example, PISA is based on global skill-level benchmarks for 15-year-old students. In this way, countries compare the indirect effects of their education systems. The OECD studies have evolved to include more domains and reach more audiences, such as the Teaching and Learning International Survey (TALIS), which began sampling teachers in 2008 and happens every five years. A key feature of the international assessments are the background questionnaires which survey relevant teacher and school information. A growing number of countries are participating in international assessments, with the fewest number of participating countries in Africa (Kamens & McNeely, 2010). In recent cycles, the largest studies have over

60 participating education systems. In theory, TIMSS and PISA differ somewhat in their basic frameworks. However, the studies use similar sampling methods and psychometric techniques, and appear to include similar questions in the tests (Wu, 2009). TIMSS and PISA results are highly correlated at the country level (Hanushek & Woessmann, 2011), suggesting the measurement of a common array of skills.

There is a growing interest in using international large-scale assessment data for policy evaluation and study (Strietholt, Gustafsson, Rosèn & Bos, 2014). The differentiation across educational systems provides opportunities for comparing counterfactual outcomes and educational inputs across a range of economic, social, and cultural contexts. International large-scale assessments have therefore evolved into much-debated global arbiters of educational reform. For example, the role of knowledge capital is now somewhat ubiquitously present across the different international testing bodies (Hanushek & Woessmann, 2015). This has pushed education systems across the globe towards a greater emphasis on educational monitoring, accountability and cross- country comparison (Baird et al., 2016). Because of their widespread reach, international large-scale assessments have been heavily criticized as having a homogenizing effect on global curricula, and an outsized focus on measurement and results (Biesta, 2009; Johansson, 2016; Komatsu & Rappleye, 2017; Meyer & Benavot, 2013; Nóvoa & Yariv-Marshal, 2003).

Despite the continued debate, international large-scale assessments offer unique opportunities for policy assessment (Strietholt et al., 2014). Because of the similarity in sampling and psychometric techniques across the tests, this information can be used to make valid comparisons providing measurement equivalence is reached. They are also generally representative of the entire educational system due to the sampling structure. The tests also have a longitudinal design at the country level (Gustafsson, 2013), and can be used as measures of achievement trends within countries and may be the only available information on such trends for some educational systems (Johansson, 2016). There are also examples of researchers using innovative methods to determine the relative effectiveness of educational systems, given their level of economic development and access to other resources (Lenkeit & Caro, 2014). Perhaps the most important benefit of cross-national comparisons and large-scale assessment relates to the fact that many educational policies and system-level or 'institutional' features have limited variation within countries (Hanushek & Woessmann, 2011; 2014) This means that there is almost no way to evaluate the impact of such features without

comparing their differences across educational systems. This 'institutional differentiation' is therefore crucial to determining the ecological validity of certain findings.

Opportunity to learn and the opportunity gap

The overarching theoretical framework of the thesis comes from the 'opportunity to learn' (OTL) model first introduced by Carroll (1963). The OTL model provides an overview of the many factors that can affect student learning outcomes at all levels within the school system, and which can be either equitably or inequitably distributed. The model has since expanded to include not only the time dimension, but other dimensions of opportunity, such as factors related to the school context and resources (McDonnell, 1995; Schmidt, Raizen, Britton, Bianchi, & Wolfe, 1997; Schmidt, Zoido & Cogan, 2013). The OTL model also reflects the 'nested' structure of educational systems, and the relationships between the different levels. The expanded OTL model includes a contextual domain as well as the intended domain at the institutional level (see 'model of potential educational experiences' by Schmidt et al., 1997), which builds upon the original OTL framework to include an emphasis on the instructional/teacher qualification dimension, as well as the relationship between student characteristics and test score outcomes (Schmidt et al., 1997; 2013). A visual display of this model is presented in Figure 1.

Figure 1 portrays the relationship between the 'intended' educational opportunities and policies within an educational curriculum, the 'implemented' or enacted educational curriculum, and the 'attained' curriculum, concerned with the actual outcomes of students. Regardless of the intended features of an education system, there can be a mismatch between the quality of the implemented educational inputs between students of different socioeconomic background due to their socioeconomic positioning (Schmidt & McKnight, 2012). For the purposes of this thesis, this can be observed between students in the same classrooms, between classrooms within schools, between schools, or at the national or district level across groups of students based on other background characteristics. This inequality of opportunity, particularly as it relates to teachers and their qualifications and competence levels, has been called the teacher 'opportunity gap' (Akiba et al., 2007).

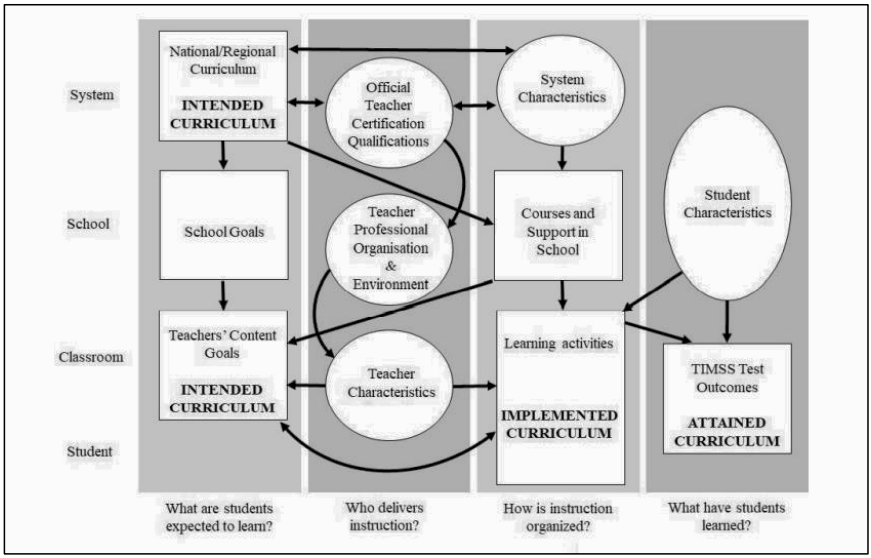


Figure 1 Theoretical model of educational experiences (Schmidt et al., 1997)

As can be seen in Figure 1, factors at the system level necessarily constrain the implemented level; for example, the degree of economic development in a system may moderate an educational policy and may also have a direct effect on school resources. Teacher sorting, however, can be measured at the system-level as differences in average teacher qualifications between students, schools, or classrooms. The thesis examines teacher sorting as both an explanatory variable and as well as a dependent variable. Sorting is therefore somewhere between the implemented curriculum and the system-level intended characteristics, as it both constrains the learning activities of students and formulates an institutional feature of education systems. While most empirical studies tend to focus on the student-level (i.e., the attained curriculum) as an outcome, this thesis focuses on both the attained curriculum (in the achievement inequity domain, or more specifically the relationship between SES and student test performance) and the implemented curriculum in the form of teacher sorting. In this way, the thesis uses the 'equity' outcome domain of effectiveness (Kyriakides et al., 2018), particularly in Study II, displayed in the relationship between socioeconomic status and TIMSS test outcomes in Figure 1, as well as the relationship between 'Teacher characteristics' and an additional 'student characteristics' at the classroom level (Studies III and IV) as dependent variables.

Measuring socioeconomic status

One of the most frequently used student-level covariates in models with international large-scale assessment data is socioeconomic status (SES). A longstanding construct in the educational and psychological literature, the SES scale positions students along a spectrum related to their social, cultural, and economic resources (also known as 'forms of capital,' Bourdieu, 1986). Bourdieu (1986) emphasized the importance of non-material forms of capital, in the form of access to familiarity with certain types of knowledge and individuals and goes so far as to say that 'the scholastic yield from educational action depends on the cultural capital previously invested by the family' (p. 244). SES is a relative construct which necessarily requires degrees of stratification in the levels of advantage and disadvantage afforded to each group based on their social positioning.

A host of contextual questions measure socioeconomic status in the international large-scale assessment questionnaires. They generally include indicators related to parental educational attainment for both the mother and the father, as well as indicators of parental income and occupation. There is a large body of literature confirming the relevance of this tripartite construct of SES and its reflection of a student's status along the socioeconomic spectrum (Sirin, 2005: Yang & Gustafsson, 2004). International large-scale assessments have incorporated another measure related to educational resources or 'home possessions' which will typically include books, a study desk, or a computer (Sirin, 2005). While related, these indicators measure different dimensions of socioeconomic status corresponding to the forms of capital outlined by Bourdieu (1986). The relationship between socioeconomic status and student achievement can be measured at the student level or at the school level by aggregation or other ways of measuring school socioeconomic composition, such as the perception of the principal.

The correlation between SES and academic achievement is perhaps the longest standing finding in the educational literature (Chmielewski, 2019; Coleman et al., 1966; White, 1982). It has been shown that by the end of secondary school, disadvantaged pupils are behind their most affluent peers in their academic abilities by one to three years of schooling (Jerrim, 2012). However, there is considerable variation in the strength of the association between student background and their academic performance (Broer, Bai & Fonseca, 2018; Sirin, 2005; White, 1982). In a large meta-analysis, Sirin (2005) found that correlations between SES and

achievement at the student level had a mean of around 0.23 to 0.29. Harwell et al. (2017) find a correlation of 0.22, concluding that SES explains just under 5% of the variation in student achievement. According to Sirin (2005), when aggregated school-level, the mean correlation was around 0.60. There is some controversy about aggregating student SES to the school level, however. First, there is the issue of measurement error, in that school SES takes over part of the effect of individual SES (Marsh et al., 2009). Next there are the so-called 'phantom-effects,' which point to the fact that without controls for prior student ability, the association between SES and achievement at the school level includes this ability variation as well as the peer effects (Burger, 2019). However, some studies have controlled for prior ability and found that students with equal socioeconomic positions perform much worse if they attend a school with a high proportion of socioeconomically disadvantaged students as compared to a more affluent school (Palardy, 2013).

In a large study examining the association of student background on educational test scores across educational systems, Freeman, Machin and Viarengo (2011) find that in general countries tend to raise average achievement levels via improvements of those at the bottom of the achievement distribution. Another approach to this question comes from Woessmann (2004), who measures the degree of 'educational opportunities' afforded to students by determining first how much their socioeconomic background contributes to their educational outcomes, and then taking what variation left over as space for 'educational opportunity' and finds the largest effect of SES among European countries in England, Scotland, and Germany. Woessmann (2004) also notes that without a control for prior ability, there is also no way to estimate whether the impact of socioeconomic background on achievement differs as a function of student ability.

The measurement of socioeconomic status has attracted much debate having to do with the validity and reliability of the construct, while another degree of complexity is introduced with cross-national and cross-temporal comparison. There has been a wave of skepticism regarding the 'number of books at home' indicator. While a host of studies have recommended the number of books in the home as an important proxy for cultural capital as well as a comparable indicator across countries (Brese & Mirazchiyski, 2013; Hanushek and Woessmann, 2011), others doubt its usefulness. Both Engzell (2018) and Jerrim and Micklewright (2014) discuss the issue of parent-child agreement, where student and parent reporting of socioeconomic indicators do not match up. This is particularly true for younger students in grade 4 as opposed to grade 8. Engzell (2018) also finds that low-achieving students in grades 4 or 5 tend to underestimate the number of

books they have at home. Other lines of criticism have to do with the validity and comparability of these measures across different cultural contexts and across time (Pokropek, Borgonovi, & McCormick, 2017; Rutkowski & Rutkowski, 2013). For instance, having one or more parents with a bachelor's degree may be very rare in certain education systems and commonplace in others and the substantive significance of for example having a computer may change, particularly over longer time ranges. Traynor and Raykov (2013) find that the household possession items in PISA include such a degree of measurement error that they are questionable indicators of wealth. Despite these concerns, Chmielewski (2019) finds that the association between the number of books in the home and other indicators of socioeconomic status (such as parental education or occupation) have been increasing over time. Chmielewski (2019) and Reardon (2011) both advocate for a percentile method whereby when considering SES across countries and time student SES is estimated within each country and cycle. Therefore, even as the 'absolute' meaning of the positions may change, the comparisons are still based on the position of students relative to all others in their country and year cohort. The last line of criticism questions the SES-achievement model altogether and proposes a 'cognitive ability/genetic transmission' model instead (Marks & O'Connell, 2021). Recent examinations of genetic determinants of long-term behavioural outcomes however conclude that 'consistent evidence for both selection (from us) and social causation (in the larger literature) means that policies and interventions will need to target resources at both people and place to be effective' (Belsky et al., 2019, p. 582).

Taken together, the criticisms of the measurement of socioeconomic status warrant emphasizing that SES is an imperfect measure which does not capture all sources of individual differences in student achievement.

Social reproduction in schools and school systems

From the emergence of educational effectiveness research in the latter half of the 20th century came questions regarding the role of the school and school system in reproducing or exacerbating societal inequities. Many of these concerns are encapsulated by Bourdieu's (1977) social and cultural reproduction theory. Via the concept of *habitus,* Bourdieu describes how institutional structures such as class relations can have physical embodiments. Habitus therefore acts as a link between 'social cause and social effect' (Nash, 1990, p. 434). Bourdieu situates the school

as the primary place in which habitus is generated. The main argument purports that schooling benefits a small number of individuals who are ready for learning and the schooling process and neglects the remaining students. The students who exhibit this readiness are by no coincidence often from middle- or upper-class backgrounds. The forms of capital afforded to such students by their families may be economic in the form of monetary investments to their education, such as extra lessons, private schooling, or schooling in socially segregated upper-class areas, cultural in the form of exposure to certain types of knowledge (i.e., art, speech, and mannerisms), and social in the form of connections and network. Bourdieu (1977) asserts that the school system is controlled by the dominant cultural classes, and that the habitus of other students is seen as a deficit. There is therefore a distinction to be made with respect to student outcomes which are generated by differences in class and cultural habitus and those that are generated by disadvantages for certain students. This is not such an easy task, as it has been shown repeatedly that students from working class socioeconomic backgrounds have different preferences for their educational outcomes and goals (Willis, 1977). The designs of educational systems also reflect *habitus*, as they are individual views concerning social welfare and individualism made collective (Lavrijsen & Nicaise, 2016). This dialectical relationship between structures and individuals is a central theme in Bourdieu's theory of social reproduction.

How is habitus enacted? A plethora of empirical studies have documented persistent social inequities in schools at individual and collective levels. Chmielewski (2019) highlights several possible reasons for the widening SES achievement gap, including increasingly unequal parental investments in education, a more diverse student population in school, and rising income inequality. Additional ways in which habitus is reflected through schooling and school organization are through the psychological processes involved in decision making and human resource governance. Why are certain schools 'desirable' as places of employment and others are not? Which students are deemed 'more likely' to succeed on the part of school leaders, and how does this affect the way school resources are distributed? How do various educational reforms—such as increased teacher monitoring—interact with these processes? Certain reforms (i.e., the movement towards performance-based accountability) have neglected the dimension of equity altogether (Lee & Wong, 2003), and others have simply prioritized efficiency as in the case of tracking and ability grouping (Hanushek & Woessmann, 2006; Strello et al., 2021). Both Jerrim (2012) and Gutiérrez et al. (2020) conclude that educational inequity has remained largely unchanged and is

the result of structural elements in education systems, such as residential segregation or long-standing organizational elements of education systems.

As has been shown, this somewhat pessimistic view can be traced back to the Coleman report of 1966 and still pervades much educational policy debate today. However, there is also much empirical evidence demonstrating that schooling does indeed matter (Dupriez & Dumay, 2006; Hanushek & Kain, 1972) and can play either a compensatory function and mitigate the relationship between socioeconomic status and student performance or vice versa. Gustafsson, Nilsen and Yang Hansen (2018) outline two ways in which this can occur. First, educational factors may have differential effects on student outcomes depending on their position along the socioeconomic spectrum; and second, through a 'correlation between SES and educational factors' (p. 17). This latter reason is perhaps most reflective of decision making and the 'embodiment of structures' on the part of educational actors such as parents, principals, teachers, and policymakers. It is, however, less straightforward to connect this capacity for change and individual and collective agency to Bourdieu's work. Is habitus deterministic? Bourdieu has little to say about institutional change. Referred to as an 'internalized structure,' it is distinct from 'objective structures' such as educational institutions or political regimes and can in turn influence these objective structures while similarly reorganizing itself. In addition, while habitus predisposes individuals to certain behaviors, it is in itself unpredictable (Nash, 1990). Circularity and determinism are in any case central critiques of Bourdieu's work (Jenkins, 1982).

Empirically, school factors which are significantly related to stronger or weaker SES-achievement relationships demonstrate that educational inputs and policy amendment can in some cases effectively address inequity and in others reproduce social inequities. One of the most frequently cited 'interventions'—class size—has been supported by mixed evidence (Ecalle, Magnan & Gilbert, 2006; Nye, Konstantopoulos & Hedges, 2004). There is other evidence for the importance of a supportive school climate, school emphasis on academic success, and school safety (Gustafsson et al., 2018; Kyriakides, Creemers, Antoniou & Demetriou, 2010). Time-on-task has also been shown to be inequitably distributed in favour of higher-SES students (Burger, 2016; Gustafsson et al., 2018; Rolfe, Strietholt & Yang Hansen, 2021).

Despite these interventions, educational inequity has persisted and is even on the rise (Chmielewski, 2019). Some researchers argue that it persists because scholars and policymakers have failed to capture its most important causes in terms

of school inputs (Sims & Allen, 2018). The quality of instruction and competence of teachers is one area which has been repeatedly come into focus cross-nationally, especially in recent years (OECD, 2018b; OECD, 2022). This differential access to teachers is the main focus of the thesis and will be presented in detail in Chapter 3.

Equality, sufficiency, and priority

By now, it is accepted that schooling matters, but how should educational resources be most effectively allocated? Policymakers and scholars have moved away from attempting to ensure equality of educational outcomes towards a focus on equality of educational opportunities (Roemer, 1998). Equality of opportunity, however, is rarely specified beyond the notion of equal educational inputs. This question has to do with the 'distributive intuitions' of societies (Brighouse & Swift, 2006, p. 471), ranging from absolute equality to sufficiency (all have 'enough') and priority (giving priority to the 'worse off'). This is also discussed in a similar fashion in terms of equity as 'inclusion' or equity as 'fairness', respectively (Kyriakides et al., 2018). Despite this, equality in the context of positional goods makes little sense as the absolute meaning of the good depends on its position along the slope. The priority argument for teaching quality, for example, disservices those with the 'good' in the first place and favors those without. A priority or 'fair' distribution of teachers would not only be captured by a null relationship between SES and teacher qualifications but an inverse one, where students with lower socioeconomic status have teachers with higher qualification levels. With a strictly equal distribution of teacher resources, on the other hand, both the relative and absolute value of the good is lost, as when all are afforded the same resources their possession no longer yields a competitive advantage.

Decisions about the allocation of educational inputs—often described as the 'paramount' positional good—depend on the good in question (Brighouse and Swift, 2006). In most cases, if priority is the preferred distribution, policymakers cannot rely on growth alone. For example, increasing the total share of certified teachers in the job market does not guarantee that the most disadvantaged students will have access to them. To achieve a prioritarian distribution, according to Brighouse and Swift (2006), 'the only way is to give less to some' (p. 475). There is a difference however between restricting higher SES students from having teachers with, for instance, master's degrees, and ensuring that lower-SES students have teachers with basic teaching qualifications such as the appropriate education

and certification. Such an approach is where equality, priority and sufficiency meet and is termed the 'threshold of adequacy' (Brighouse & Swift, 2006, p. 480). Adequacy posits that for example students, regardless of socioeconomic background, will have teachers with at least the same basic qualifications thereby 'leveling down' both the absolute and relative value of this good, while at the same time ensuring that everyone has 'enough.' In this context, 'equity' or 'leveling down' works as a function of what is considered 'enough' for those at the lower end of the socioeconomic spectrum. There is another angle where these perspectives overlap which favours more explicitly the priority or 'fairness' view. This has to do with the value of educational inputs relative to other inputs related to socioeconomic positioning. Here, a strictly prioritarian distribution of teacher qualifications (i.e., higher qualifications favoring low-SES students) would yield equality and sufficiency in competition, as it is assumed that the advantages afforded to higher-SES students by their background would compensate for this type of inequality.

Equality of opportunity, therefore, is at times not distinguished between the 'adequacy' and 'inclusion' approach, or the 'priority' and 'fairness' approach. There is still not a clear consensus about the best way forward. Anderson (2007) argues in favour of a sufficientarian standard whereby all individuals are guaranteed an education which prepares them to complete a four-year post-secondary degree and thereby access to the 'elite' sector of society. This is, however, a way of leveling down as Anderson (2007) argues that academic achievement beyond these basic qualifications should be discounted in favour of other procedural forms of knowledge with the end goal of 'social integration' (p. 617). Brighouse and Swift (2008), in turn, position prioritarian forms of allocation to provide fair competition in their view of equity as 'fairness.' For example, they argue that weighted funding-per-student in favour of socioeconomically disadvantaged students may attract better teachers to schools and neighborhoods with more challenging working conditions. In this way, it is a type of compensation rather than an actual advantage.

Brighouse, Ladd, Loeb and Swift (2016) utilize the aforementioned distinctions and provide a framework for decision-makers in educational policy. They stress that educational policymakers operate within societies and decisions and educational governance are often contingent upon this fact. They write, 'if they [policymakers] could change residential segregation in the United States, their decisions about how to fund schools might change' (p. 5). Other scholars note that focusing on such non-educational policies may be more effective (Solga, 2014). In

any case, policymakers must decide which aspects they want to prioritize. For example, complete centralization of teacher allocation may guarantee equal distribution of teacher credentials across schools at the cost of individual choice and autonomy on the part of teachers. Such a compromise may have stronger negative unintended consequences in societies where individual autonomy is prized. Ultimately, when educational policymakers choose either an adequacy, equality, or priority approach, such values and trade-offs should be made visible in the decision-making process (Brighouse et al., 2016).

Up until now, only theoretical ideals regarding the allocation of educational resources have been discussed. In practice, few, if any, educational systems provide higher quality education in the form of teacher competence to individuals of lower socioeconomic status. It should be noted that the allocation of human resources, as in the case of teacher allocation, adds a degree of complexity which is not present to the same extent in the allocation of material resources. In this way, the thesis argues that the first step is indeed adequacy, and only then is it practical to discuss more prioritarian forms of resource allocation.

Chapter 3 An overview of teacher sorting

The inequitable distribution of teacher competence has increasingly been touted as a key educational input which has been neglected in terms of its ability to account for variation in educational outcomes, and according to the OECD (2018b), is still a challenge for many education systems around the globe. There has been some progress—over the past several decades, countries have invested in attracting more people to the teaching workforce (OECD, 2005, 2018b), and many countries have smaller student-teacher ratios in disadvantaged schools (OECD, 2018b). Despite this, it is now well-understood that increasing the quantity of teachers is not a solution in itself, as teachers do not necessarily change their classroom practices in smaller classrooms (Finn, Pannozzo & Achilles, 2003). Principals of disadvantaged schools are still more likely to report that a lack of qualified teachers hinders learning (OECD, 2018b).

Currently, hypotheses regarding the mechanisms of teacher sorting are still developing (Luschei & Jeong, 2019). Overall, the underlying mechanisms having to do with teacher self-selection can be encapsulated by rational choice theory, where individual decisions stem from one's best interest (Kahneman & Tversky, 1979), with some exceptions. In general, teachers tend to care more about working conditions than pay after entering the field (Bacolod, 2007). The general hypothesis underlying the idea of within-school teacher sorting is that in more socioeconomically heterogeneous schools which practice ability grouping (explicitly or otherwise), teachers with higher competence and qualification levels may be afforded more choice over which classrooms they wish to teach. Principals or school actors may also prioritize raising the achievement levels of the classes with higher-ability students and therefore also have a hand in the allocation process (Boyd et al., 2008). As ability is correlated with socioeconomic background, this is likely to result in socioeconomic sorting of teachers across students.

Hypotheses for why teachers may be inequitably distributed across schools are many. There may be differences in ideological motivations between certain teachers. Indeed, there is evidence to suggest that newer teachers tend to more

frequently report wishing to 'make a difference' in students' lives (Sims and Allen, 2018). They may therefore seek out positions in otherwise hard-to-staff settings. Teacher labour market conditions should also be taken into account. Schools with a high proportion of socioeconomically disadvantaged students tend to have more job openings and less applicants (Engel, Jacob, & Curran, 2014; Jackson et al., 2014). In addition to this, inefficiencies in hiring practices (i.e., less aggressive recruiting or inopportune hiring times) may contribute to the disparities in teacher competence across schools (Lankford, Loeb & Wyckoff, 2002). Early career teachers may also seek out jobs in lower-SES schools with the explicit intention of changing schools as soon as it becomes feasible. In some countries, teachers are afforded a 'seniority score' which gives them more leverage in the job seeking process and working in lower-income schools is one way to positively influence such scores (Özoğlu, 2015). Others may leave due to poor working conditions and burnout (Valli & Buese, 2007).

Measuring teacher competence

Underpinning the discussion of teacher sorting is the problem of defining teacher quality and teacher competence. In short, how do we identify a competent or highly qualified teacher? Some of the earliest studies on the topic were conducted by economists (Hanushek, 1971; Murnane, 1975), where the tradition remains strong today (Hanushek, Piopiunik, & Wiederhold, 2014; Sancassani, 2021). Another strand of research has come from the educational scientists (Baumert et al., 2010; Nilsen & Gustafsson, 2016). Since then, the growing availability of datasets linking teacher and student data has led to a proliferation of research examining teacher effects. Despite this, there is no general consensus in the literature about which teacher characteristics matter most. This makes it difficult to identify competent teachers and provide concrete recommendations for school leaders and policy makers for how to better support student learning in low-SES settings. The lack of consensus about how to define and measure teacher competence is somewhat paradoxical, as teachers and teaching quality are repeatedly cited as the most important school-level inputs for students (Chetty, Friedman & Rockoff, 2014; Goe, 2007; Hanushek & Rivkin, 2012; Hattie, 2003; Hattie, 2009; Nye, Konstantoupolous & Hedges, 2004; Rivkin, Hanushek & Kain, 2005; Slater, Davies & Burgess, 2012; Wayne & Youngs, 2003). A meta-analysis by Hattie (2003) purports that 30% of the variance in student achievement can be attributed to teachers, while a within-teacher approach by Nye et al. (2004) finds a

variance component of around 10%. A recent study by Sancassani (2021) uses cross-national data and finds an effect size of 3% of a standard deviation for the specialization qualification alone. Conceptualizations of teacher quality and different methodologies yield different results, and it is important to take into account such measures and methodologies when determining the overall influence of a particular factor related to teacher quality. The terms teacher *competence*, *quality*, and *effectiveness* are often used interchangeably within the literature. They refer to the ability of teachers to influence student outcomes positively or negatively but can be measured in different ways.

Goe (2007) provides a useful framework for conceptualizing the different dimensions of teacher quality, comprised of an input, output, and process dimension. Inputs generally comprise of immutable teacher characteristics such as ethnicity and gender, but also their beliefs about student learning, confidence and preparedness levels in teaching, for example. Goe (2007) notes that the evidence supporting factors such as self-efficacy and preparedness is rather weak. The input dimension also includes teacher qualifications such as their licensure test scores, educational attainment, experience, specialization and pedagogical training. The process dimension is concerned with the classroom; namely, instructional practices such as classroom management and cognitive activation (Goe, 2007; Kyriakides, Christoforou and Charambolous, 2013). Last, the output dimension pertains to value-added measures of teacher competence which are based on variations in student test scores before and after their exposure to a certain teacher, or when teachers can be identified in multiple school or classroom settings.

Much of the teacher competence literature thus far has focused on teacher qualifications as observable proxies for teacher quality, with mixed results. Teacher experience is perhaps the most well-cited input. The hypothesis suggests that there is a 'learning-by-doing' effect, where teachers improve substantially after a few years on the job (Rice, 2003). This was first introduced in the American literature and has been confirmed in a number of meta-analyses and observational studies which show both correlations and causal estimates (Goe, 2007; Rice, 2003; Rockoff, 2004; Rivkin et al., 2005; Wayne & Youngs, 2003). However, newer evidence suggests that the returns to experience continue much farther into a teacher's career, as far as 10 or even 19 years on the job (Papay & Kraft, 2013; Toropova, Johansson & Myrberg, 2019; Wiswall, 2013). To make matters more complicated, other recent studies have called into question the importance of teacher experience (Isenberg et al., 2021; Sancassani, 2021). There is also evidence for the importance of specialization or having a university level degree in the

subject matter taught (Baumert et al., 2010; Darling-Hammond, 2012; Goe, 2007; Hill, Rowan & Ball, 2005; Johansson & Myrberg, 2019; Sancassani, 2021; Wayne & Youngs, 2003). The reasons for this are somewhat obvious, but it is assumed that a teacher will have less domain-specific knowledge about a subject without subject-specific training. Other studies have failed to support the importance of subject specialization (Harris & Sass, 2011; Rockoff et al., 2011), but in general, there are fewer studies disconfirming its importance as compared to those disconfirming the importance of experience. Several other qualifications have been found to be important for student outcomes, including cognitive skills (which are sometimes measured by licensure test scores) (Hanushek et al., 2014) and pedagogical certification (Clotfelter et al., 2006; Goldhaber & Brewer, 2000).

The teacher sorting literature tends to focus on the impact of teachers with lower qualification levels, as it is claimed that low-SES students are more likely to encounter such teachers. This also limits the endogeneity issue related to changeable teacher characteristics which are bound to vary across teaching contexts. Because of limitations for cross-country comparison and the quality of certain indicators in the large-scale assessment data, experience and specialization are the dimensions focused upon in this thesis. Although Jackson et al. (2014) note that some researchers were better able to expose teacher effectiveness using composite measures of teacher quality, such measures based on teacher qualifications are often not empirically well-founded and may be less useful for policy recommendations.

Despite the controversy, there is a degree of consensus in the literature regarding the subject-specific nature of teacher competence. Extensive subject matter training as well as opportunities for field training have been outlined as crucial elements of a successful teacher preparation program (Darling-Hammond, 2000; 2012). The proposed mechanisms underlying this process may have to do with the degree of subject knowledge and pedagogical knowledge a teacher possesses. Shulman (1986; 1987) introduced the concept of content knowledge (CK) and pedagogical content knowledge (PCK) some decades ago. According to this model, CK relates to the degree of competency, familiarity, and knowledge about a certain subject, and PCK pertains to the ability of teachers to effectively communicate this knowledge to students. General pedagogical knowledge, on the other hand, refers to classroom practices such as management, cognitive activation; and a general knowledge dimension includes knowledge of student learning propensities and the role of educational settings. The importance of both teacher qualifications and PCK and CK has been found to be especially important

in mathematics and to a lesser degree science (Goe, 2007; Wayne & Youngs, 2003). This may be because the highly technical nature of mathematics and science require competent teaching to be effectively transmitted, and unlike other subjects, mathematics and science are more likely to be learned mostly in school (Nye et al., 2004). Additionally, mathematics and science curricula continue to become more homogenized around the globe and are therefore more easily comparable across countries. The concepts of both content knowledge pedagogical content knowledge have since been applied to the mathematics domain (Hill, Ball, & Schilling, 2008). A well-known study documenting the impact of CK and PCK comes from Baumert et al. (2010) and a German sample, who note that while content knowledge is important, PCK is what really makes the difference for student learning gains. They also find that while CK was related to the alignment of tasks to the curriculum goals, it was unrelated to both individual learning support and cognitive activation.

Two recent cross-national studies find a link between teacher qualifications and instructional quality. This 'mediating role' of instructional quality between teacher qualifications and student outcomes has been proposed by several researchers (Darling-Hammond, 2006; Kunter et al., 2013). Using TIMSS 2011 data and multilevel structural equation models, Blömeke, Olsen, and Suhl (2016) found that teacher experience was significantly related to instructional quality in a pooled model with 47 countries. They also found support for subject major and level of education in several countries, but the differing directions of these findings made them difficult to generalize either way. In another study focused on Nordic countries (Denmark, Finland, Sweden and Norway), Nilsen, Scherer and Blömeke (2018) confirm that in certain countries education level and specialization (in science) are significantly related to student outcomes via instructional quality. However, these findings are not ubiquitous across countries in the sample.

Taken together, the contradictory nature of the literature shows that the cross-national comparisons involving the construct of teacher competence and quality are not so straightforward. While many of the studies on teacher competence, effectiveness and quality are conducted in the United States, other studies from around the world also show mixed results. This may be because the impact of teachers may vary across grades and subjects, as well as across countries (Blömeke & Delaney, 2012; Blömeke & Olsen, 2019; Jackson et al., 2014). In general, however, there is a consensus that qualifications matter most at the secondary level (Goe, 2007; Baumert et al., 2010), and that regional patterns may be seen in the relevance of teacher qualifications across countries (Blömeke & Olsen, 2019). In

certain studies, qualifications such as experience, certification, and specialization are described as the most validly comparable across countries (Akiba et al., 2007). It may be the case that they are in fact the most conservative indicators of teacher quality to compare, but they are not without their own set of potential pitfalls. Blömeke and Olsen (2019) write, 'specialization in a subject could be well understood in one country but not another. The relation may be weaker in the latter case although the actual relevance could be the same' (p. 179). Despite these limitations, the persistent murkiness in the teacher qualification literature signals a need for continued focus on the topic.

A final facet of teacher quality related to this thesis pertains to the differential effects of teacher quality and competence on achievement as a function of student SES. Teachers (and educational inputs in general) are hypothesized to matter more for students of lower socioeconomic status due to the fact that they have less educational supports outside of school (Darling-Hammond, 2000). The focus of this thesis in general is not on the contribution of teacher quality and qualifications to overall student achievement, but on whether inequities in access to similarly qualified teachers widen the socioeconomic achievement gap and perpetuate educational inequity. It is mostly for this reason that the most conservative measures of quality are used, such as exposure to novice teachers, exposure to teachers with no mathematics education specialization.

Review of past literature on teacher sorting

Investigation of teacher sorting has a longer history in the United States as compared to elsewhere in the world. Some of the earliest studies happened more than 20 years ago (Betts, Rueben & Danenberg, 2000; Lankford, Loeb & Wyckoff, 2002). Lankford et al. (2002) use New York City data to describe the variation in teacher credentials across schools. They note that urban schools, or schools with a high proportion of low-income and 'non-white' students have a much higher share of teachers with lower qualifications by almost every single measure, including experience, level of education, pedagogical certification, and exam scores. Teachers tended to quit these schools at a higher rate as well. They also note that variation in the distribution of teacher attributes has remained stable for 15 years at the time the paper was published (from 1985-2000). In another study, Clotfelter, Ladd and Vigdor (2006) examine the extent to which non-random matching of teachers to students generated by the sorting process biases estimates of the relationship between teacher characteristics and student achievement. Using

data from North Carolina, USA, they first find that teachers with more experience, degrees from more competitive colleges, and more advanced levels of education tend to teach in schools with a higher proportion of affluent students. They also find that this phenomenon extends to inequities between classrooms within-schools. Clotfelter et al. (2006) conclude that licensure test scores and experience are robust determinants of student achievement. Findings in the US confirm that such inequities are somewhat ubiquitous across the country, and extend to racial disparities as well (Clotfelter, Ladd, & Vigdor, 2005; DeAngelis, Presley & White, 2005; Feng, 2010; Goldhaber, Lavery & Theobald, 2015; Grissom, Kalogrides & Loeb, 2015; Jackson, 2009; Kalogrides, Loeb, & Beteille, 2013).

There is some evidence to suggest that teacher sorting in the USA has actually declined as a result of policy changes, including the No Child Left Behind Act (2001), as well as lower requirements and costs for entering the teaching profession (Boyd, Lankford, Loeb, Rockoff & Wyckoff, 2008). To this end, Boyd et al. (2008) find that such changes account for an increase in student outcomes and recommend that low-SES schools hire teachers with better qualifications such as test scores or pedagogical training. Other studies have not seen such pronounced declines. Using data from Washington and North Carolina, Goldhaber, Quince and Theobald (2018) find that largely, opportunity gaps related to most teacher qualifications have remained broadly stable over time and even that the 'novice teacher gap' has grown over time. Taken together, these findings make it difficult to form a conclusion about the trends in teacher sorting, even within a single country.

A handful of national studies from other countries have also raised the alarm. Allen, Burgess and Mayo (2012) use data from England and study the relationship between teacher turnover and socioeconomically disadvantaged schools and find that turnover is partly due to the low 'attractiveness' of disadvantaged schools, and partly due to job opportunities and labour market conditions between neighborhoods. Sims and Allen (2018) calculate the opportunity gap by school deprivation quintile (schools with the highest concentration of students with free school meals) and find a significant difference in teaching quality between the highest and lowest deprivation schools. This pattern was also seen within schools as well, where more experienced teachers were often allocated to the highest-performing classes.

Using data from Chile, Meckes and Bascopé (2012) examine the distribution of novice teachers, and find that teachers who performed better on their exit exams are more likely to be hired in socioeconomically advantaged settings. They

also note that in Chile, teachers in disadvantaged schools tend to have lower mobility rates, a pattern which seems to diverge from characteristics of teacher mobility in the US and England. In general, the few (English language) studies that exist on the topic demonstrate that teacher sorting tends to be quite high in Latin America (Luschei, 2012; Luschei & Carnoy, 2010).

In Sweden, Hansson and Gustafsson (2016) show on average about a 5 percentage point different between the proportion of uncertified teachers between the most and least socioeconomically disadvantaged schools. Other sources find preliminary evidence of inequities in Turkey, Australia, Canada, Dominican Republic, Kosovo, Jordan, China, Czech Republic, Iceland, Ireland, Luxembourg, New Zealand, Norway, the Netherlands, Singapore, Taipei, Trinidad and Tobago, Slovenia, Spain, Switzerland and the UAE (Bonesrønning, Falch & Strøm, 2005; OECD, 2018b; Özoğlu, 2015). Some of this evidence comes from the OECD (2018b) on the degree to which principals of different schools report a lack of teaching staff.

Despite its presence in nationally focused studies, cross-national comparative research on the topic of teacher sorting is scarce. In an international comparison using TIMSS 2003 data, Akiba et al. (2007) compute the difference in percentage of low- and high-SES students with access to qualified teachers. They show that in 2003 the international mean was a 2.5% difference, and the countries with the highest opportunity gaps include Saudi Arabia, Jordan, Singapore, Hong Kong, United States, Taiwan, Chile, and Syria. Akiba et al. (2007) found that teacher qualification gaps had no influence on SES-achievement gaps across countries. Other studies have focused largely on teacher distribution in developing countries (Chudgar & Luschei, 2016; Luschei, Chudgar & Rew, 2013). They note that teachers of marginalized children in these settings are disproportionately 'male, young, and inexperienced' (p. 16). Similar to the Akiba et al. (2007) finding, Chudgar and Luschei (2011) found no evidence for the importance of teacher qualifications as a function of student socioeconomic background. In a more recent study, using PISA 2009 data from 65 countries, Chiu (2015) finds that in schools with teachers with a higher share of university degrees, SES-gaps were larger, arguing that students with more cultural capital benefit from teachers with more university training. Han (2018) also examined inequity in teacher sorting across countries, but found a connection between school autonomy, teacher distribution, and educational inequity. Last, using a composite measure of teacher quality comprised of level of education, experience, and self-efficacy, Luschei and

Jeong (2019) find that within-school sorting is more prevalent in higher income countries.

There are several ways of measuring teacher sorting. Most studies compute exposure rates of different socioeconomically diverse groups of students or schools to various teacher characteristics and qualifications (Akiba et al., 2007; Goldhaber et al., 2015; Sims & Allen, 2018). Others use outcome-based measures, or teacher 'value-added' measures (Goldhaber et al., 2015; Rivkin, et al., 2005; Hanushek & Rivkin, 2012). Value-added measures of teacher effectiveness and teacher sorting are not possible to employ based on the international large-scale assessment data, as they require repeated observations of individual students or schools. Clotfelter, Ladd and Vigor (2007) conceptualized the teacher opportunity gap as percentage of students eligible for free lunches, as well as the percentage of non-white students, and found considerable differences in access to teacher qualifications across groups based on these categories. Luschei and Jeong (2019) measure the teacher opportunity gap as the intra-class correlation coefficient, or the proportion of variation in a composite measure of teacher quality attributable to the school level. They also examine the association between teacher characteristics and classroom and school socioeconomic composition.

A large majority of the teacher sorting literature focuses on teacher qualifications, as opposed to teaching processes such as classroom management, or cognitive activation. This is in part because such processes are amenable to the classroom environment. For example, if teaching in low-SES classrooms is more challenging than in higher-SES classrooms, it would be no surprise that observed levels of certain teaching practices are lower in these settings. Such contextual differences make it difficult to say anything about variations in teacher competence across settings. Teacher hiring is also generally based on observable characteristics and qualifications (Engel & Finch, 2014).

Teacher working conditions

Parallel to the body of work examining the overall impact of teacher quality on student achievement is a dimension of the teacher sorting literature concerned with school and system processes related to teacher mobility factors. It can be classified into two dimensions: attraction and retention. Part of the problem has to do with incentivizing teachers to apply and seek out jobs in socioeconomically disadvantaged schools or classrooms, and the other part has to do with keeping teachers in such settings or the profession altogether. Much of this work has been

limited to investigating local practices of hiring, evaluation, and pay in the USA (Jackson et al., 2014), but a growing international literature is investigating the relevance of school system or institutional features (Han, 2018; Luschei & Jeong, 2019; Smith & Holloway, 2020). While salary has been shown to matter in certain settings, pay differentials between schools are generally not enough to incentivize teacher mobility (Bacolod, 2007; Clotfelter, Glennie, Ladd, & Vigdor, 2008; Clotfelter, Ladd & Vigdor, 2011; Jackson et al., 2018; Sims & Allen, 2018). In general, mobility issues are linked to differences in working conditions across educational settings.

The working conditions of teachers are changing around the globe. First, there are changing regulations and barriers to entry into the teaching profession. In some regions in the world, entry-requirements are getting stricter, with the introductions of mandatory licensure exams and more comprehensive teacher education programs (Meckes & Bascopé, 2012), but there is an ongoing debate about how strict such barriers to entry and quality monitoring should be (Darling-Hammond, 2004b). There are also ongoing issues related to teacher shortages but increasing enrollment rates into teacher education programs does not seem to solve this issue, as there is a high degree of teacher turnover out of the profession among new recruits (Feng, 2010; Ingersoll, 2017). The problem of attraction and retention is particularly pronounced in mathematics and science teachers (Ingersoll & May, 2012; Sibieta, 2018).

Issues related to attraction and retention have been found in Sweden, Canada, the USA, Australia, Chile, and England, and many other places in the world (Allen et al., 2018; Carver-Thomas & Darling-Hammond, 2017; Den Brok et al., 2017; Elacqua, Hincapie, Martinez, 2019; Ingersoll, 2002; Karsentil & Collin, 2013; Lindqvist, Nordänger & Carlsson, 2014). Such issues may be related to the worsening societal status of teachers (Craig, 2017; Darling-Hammond, 2017; Lindqvist et al., 2014) alongside increasingly stringent monitoring and an emphasis on performance and competition (Ball, 2003; Ingersoll et al., 2016; Perryman, Ball, Maguire & Braun, 2011). Although some evidence suggests that teacher stress simply mirrors trends in declining wellbeing in most professions (Jerrim, Sims, Taylor & Allen, 2021), the bottom line is that teachers are over-worked and undervalued, leading to a high degree of stress and burnout (Perryman et al., 2011). A high teacher attrition rate hinders school climate and the success of students (Hanushek et al., 2016). This may be because the school is forced to hire individuals with lower competency or qualification levels, the staff are not used to

working alongside each other, or the school simply cannot fill the open spots and is forced to enlarge classes (Ingersoll, 2001; Sorensen & Ladd, 2020).

The problems of attraction and retention are particularly pronounced in socioeconomically disadvantaged schools and classrooms (Allen et al., 2012; Boyd et al., 2011; Jackson, 2009). Such 'hard-to-staff' settings are disproportionately affected by poor working conditions. This includes infrastructure issues, the safety and cleanliness of school facilities and grounds, a lack of administrative support, larger class sizes, insufficient educational resources for students, and a lack of cohesion among school staff (Horng, 2005). Teachers may also be prone to the 'white-flight' syndrome (Jackson, 2009), and mirror the changing ethnic demographics of neighborhoods and schools.

Chapter 4 Institutional correlates of teacher sorting

Institutional (system-level) educational features can magnify or mitigate socioeconomic dispersion in test scores (Van de Werfhorst & Mijs, 2010). The educational systems included in the thesis reflect a range of economic development and inequality levels as well as institutional structures. A main hypothesis of the thesis is that such characteristics will moderate the extent to which socioeconomic background of students and teacher qualifications are correlated. While there is some precedent in the teacher sorting literature for this type of investigation (Han, 2018; Luschei & Jeong, 2019), cross-national studies are for the most part scarce. As mentioned in Chapter 2, Hanushek and Woessmann (2014) outline several advantages of studying features at the institutional level, mostly to do with the variation in policies which is much larger across countries, as well as the handling of selection effects when data are aggregated to the system-level. However, these advantages also introduce a host of challenges with regards to comparison which are discussed in the methodology section.

There is one area which is not examined in the thesis, however, which has to do with entrance requirements and quality of teacher education programs (Wang et al., 2003; Schwille et al., 2013). Past research has shown that in some educational systems, entrance requirements are becoming selective, while the opposite is true elsewhere (Alatalo, Hansson & Johansson, 2021; Park, 2019). Such variables are not available from the international large-scale assessments, and there are likely selection effects which characterize teachers currently in the profession. Moreover, there is a need for shorter-term solutions to teacher sorting as changing the attractiveness of the teaching profession is a long-term project (William, 2010).

The underlying mechanisms of teacher sorting—namely, self-selection based on working conditions—are assumed as the primary drivers of teacher sorting even with the inclusion of policy and institutional context. Some of the institutional features and policies investigated are hypothesized as having indirect effects on teacher sorting via their influence on student socioeconomic composition in schools and classrooms, as educational policies which affect the composition of

students are likely to affect the distribution of teachers (Jackson, 2009). This is the case for within- and between- school tracking, school competition and privatization, and socioeconomic school segregation. Others, on the other hand, are hypothesized as having more direct effects on teacher sorting, such as school autonomy over staffing or school accountability practices. The following sections examine each of these features and the corresponding literature in closer detail.

School tracking, competition, and segregation

A host of research has confirmed that between school tracking exacerbates educational and social segregation (Hanushek & Woessmann, 2006; Strietholt et al., 2019; Strello et al., 2021; Van de Werfhorst & Mijs, 2010). Mons (2007) provides a useful cross-national classification system describing the ways in which heterogenous student ability is handled. Education systems such as Germany, Austria and Hungary separate students into different tracks from an early age, while countries like Canada, the United States, and the United Kingdom practice ability grouping within schools beginning at the secondary level, for instance (Mons, 2007). In school systems which practice mixed ability grouping, schools may have similar teacher qualification profiles between schools but differing qualification profiles across classrooms.

Though much of the justifications for tracking lean on educational efficiency and the grouping of students with similar abilities, in practice, tracking from an early age tends to deepen inequities already present in society as the choice of educational track depends highly on student socioeconomic background (Burger, 2019). In such a way, it would be very unexpected if more selective school systems were better able to equitably allocate teachers. A similar justification for within-school ability grouping follows. A recent cross-national review by Luschei and Jeong (2019) found a significant link between the proportion of cross-school tracking and lower average years of teacher education in a school. They find no relationship between tracking and average years of teacher experience or average self-efficacy levels of teachers. In systems with less within-school grouping, on the other hand, low-SES classes have teachers with higher levels of experience. The results of Luschei and Jeong (2019) are the sole cross-national study to date which has examined how student grouping is associated with the distribution of teachers. In addition, they are highly dependent on teacher qualification, emphasizing the need to examine a wider sample of education systems as well as more teacher qualifications.

In the case that between-school tracking is uncommon, schools may be segregated socially due to other reasons, such as divides along neighborhood or geographic lines or school choice and competition. Thus far, most school systems around the world have failed to adequately address the issue of social segregation, which remains at the very least unchanged and may actually be increasing (Chmielewski, 2019; Gutiérrez et al., 2020; Marczinzak et al., 2016). This may be due to rising income inequality or increasingly diverse population demographics in certain countries (Chmielewski, 2019). Kirabo Jackson (2009) exploits the end of the student 'bussing' policy in Charlotte-Mecklenburg North Carolina (USA) in 2002. After a longstanding 'school integration' policy which required schools to have a racial composition similar to the district average, the end of bussing resulted in fast demographic changes reflecting neighborhood demographic composition. These changes provide an opportunity to examine the relationship between student composition (albeit racial) and teacher quality which are not compromised by unobserved characteristics of the schools themselves. Jackson (2009) finds that schools which had an influx of black students resulted in a decrease of qualified teachers in terms of their experience and certification levels, stating that 'one can reject the hypothesis that the correlation between student demographics and teacher quality is merely an artifact of geography or residential segregation' (p. 249). Such conclusions have been made elsewhere in the world as well. In Norway, Bonesrønning, Falch and Strøm (2005) use panel data and document a relationship between excess demand for certified teachers and share of minority students.

Other policies which may stratify students according to social class include school competition, choice and marketization. While some evidence demonstrates that increased competition in the form of greater school choice or a larger private school sector may have positive impacts on schools via the pressure to improve resources, indirectly influencing working conditions and thereby teacher competence (Hanushek & Rivkin, 2003), others show that SES achievement gaps are larger in countries which have a higher share of private schools and degree of school competition and conclude that increased opportunity of choice may reinforce inequity (Strietholt et al., 2019). Yang Hansen and Gustafsson (2016) conclude that school choice reforms introduced in the Swedish education system are an important determinant of increased socioeconomic school segregation over the past two decades. Ultimately, however, the extent to which school choice, competition and privatization are associated with inequitable educational outcomes is mixed (Strietholt et al., 2019). More explicit investigations of teacher

distribution and school competition may deliver clarity regarding these continually mixed effects.

Staffing autonomy and school accountability

One of the institutional features most often investigated with respect to teacher sorting is school autonomy over staffing (Han, 2018; Luschei, Chudgar & Rew, 2013; Luschei & Jeong, 2019). It is continuously asserted that more centralized school autonomy mitigates the teacher self-selection process (Han, 2018; Kang & Hong, 2008; Luschei & Jeong, 2019). Luschei, Chudgar and Rew (2013) as well as Kang and Hong (2008) propose that one of the main reasons for the equitable allocation of teachers in South Korea (as compared to for instance, Mexico), is the centralized hiring and teacher rotation policies that are a widespread practice in the country. More centralized staffing systems which allocate teachers into settings based via central authorities can limit opportunistic behavior on the part of school leaders and teachers. This hypothesis has been investigated by a number of studies with mixed results. The findings tend to vary depending on the indicator of teacher quality in focus. Han (2018) finds that higher levels of autonomy are associated with a higher share of certified teachers in the school, but no relationship to the educational attainment of teachers. Luschei and Jeong (2019) continue this pattern of mixed findings and find no SES-based school level differences in average teacher quality in education systems with higher levels of autonomy. Across-classrooms in more autonomous systems, they find that low-SES classrooms have teachers with less experience, but also teachers with higher levels of self-efficacy. The OECD (2018b) also finds that principals are less likely to report teacher shortages in more autonomous systems.

There may be several explanations for this pattern of mixed findings. First, central teacher assignment may only assign teachers to schools and not classes within schools, leaving room for teachers to exert seniority and preference over the classes in which they teach. Second, even fully centralized systems may allow for selection mechanisms on the part of teachers. Özoğlu (2015) writes of the Turkish case, 'teachers are allocated to schools centrally through either initial assignments of novice teachers or seniority-based transfers' (p. 19). Teachers accrue higher seniority scores faster by working in lower-socioeconomic contexts and can later transfer to higher-SES schools. According to Özoğlu (2015), this is likely an important reason why teachers are so unequally distributed in the country. Last, Luschei and Jeong (2019) theorize that the degree of unequal teacher

distribution in Korea, long touted as the model teacher allocation education system due to the centralized allocation of teachers, may be due to the increasing social inequality in the country (p. 571) but do not elaborate on how this may occur. Given the mixed state of the research, it is expected that further investigations of autonomy can provide more definitive evidence.

Similarly, school accountability has been linked to teacher sorting in a number of national studies and at least one cross-national study (Luschei and Jeong, 2019). There are many dimensions of accountability, but the use of student test results is one such domain that has been shown to impact educational inequity as well as teacher sorting (Boyd et al., 2008; Feng et al., 2016; Strietholt et al., 2019). Here, the results are again mixed. Some research has shown that increased school accountability may incentivize school leaders to allocate the best teachers to students more likely to perform better. It may also worsen teacher job satisfaction and working conditions due to increased stress and administrative burdens (Feng et al., 2016; Hanushek & Rivkin, 2003; Jerrim & Sims, 2021). In a recent cross-national study, Luschei and Jeong (2019) confirmed the anti-compensatory function of accountability in relation to sorting with regards to publicly posting achievement data, but a compensatory function of accountability with regards to keeping track of data.

Increasing evaluation and monitoring is frequently proposed as a type of 'quality control' in education systems. Some research suggests that it is predominantly novice and lower-quality teachers who tend to move schools (or away from the teacher workforce) as a result of such increased accountability mechanisms (Hanushek & Rivkin, 2010), others show such accountability 'shocks' affect more effective teachers (Feng, Figlio & Sass, 2010). There is also some doubt about the design of such evaluation programs, with 'vague standards, poor evaluation instruments, a lack of time, and a school culture that discourages critical feedback and negative evaluation ratings' (Donaldson, 2009, p. 2). Allen et al. (2012) advocate for an approach which focuses more on the development of teaching practices and coaching. Others support this conclusion and emphasize that the increasing focus on standards and educational outputs may have unintended consequences for lower-SES students and teachers in such settings (Darling-Hammond, 2004b). Unchecked accountability may incentivize principals and school leaders to allocate the best teachers to those they deem most likely to reach certain cutoffs. As argued by Lee and Wong (2003), evaluation and accountability programs tend to neglect the equity dimension altogether in favour of raising average performance across schools. Whether school accountability

fulfils its intended purpose and improves transparency and quality in the education system or further incentivizes opportunistic behavior on the part of parents, schools, and teachers remains an open question. This thesis intends to expand on this work by using other measures of teacher qualifications and performance data-based accountability as it directly relates to teachers via their formal appraisal and the resulting changes in teacher mobility.

National economic development level

When socioeconomic school segregation is not included as a control variable, a final institutional feature of interest pertains to the level of economic development in a country. Generally, it is not possible to include both in cross-national models, as economic development level is negatively correlated with social and educational segregation. This is mainly due to the fact that more economically developed societies tend to ascribe lower importance to social background (Gustafsson, Nilsen & Yang Hansen, 2018; Ferreira & Gignoux, 2014). Several cross-national comparative studies have highlighted the importance of considering this context when comparing education systems. Lenkeit and Caro (2014) compare educational rankings before and after accounting for the development level of countries, and find that after taking into account economic context, certain lower-performing countries move to the top of the 'effectiveness' ranking (such as Indonesia) and other education systems move to the bottom of the effectiveness scale (i.e., Sweden). They also write, 'historically, the choice of reference society was guided by geographical, cultural and political relations between nation states' (p. 148), but that large-scale assessments have linked these 'reference societies' mainly to which nations top the test score rankings and have high levels of economic prosperity. In another cross-national study investigating this issue, Hanushek et al. (2013) show that school autonomy and decentralization function differently depending on the economic development level of a country. In relation to teacher sorting, Luschei and Jeong (2019) use GNI as a system-level control and find that cross-school sorting is more prevalent in lower-income countries, and cross-classroom sorting is more prevalent in higher-income countries. Given the dearth of research on teacher sorting from a cross-national perspective, there is no clear hypothesis about its relationship to economic development level. Nevertheless, it is assumed that such characteristics are at least important to have under control. Education systems with higher levels of economic development may have inherently better working conditions (Schwille et al., 2013). In Studies III and IV, such possible

effects are captured by the Human Development Index (HDI) provided by the United Nations which is a geometric mean of normalized indices of life expectancy, educational attainment, and per capita income (UNDP, 2022). Study II controls for time-invariant features of educational systems by the inclusion of country fixed effects.

Chapter 5 Methodology

Data

Analyses performed in the thesis are based on data from two international large-scale assessment studies. First, the Trends in International Mathematics and Science Study (TIMSS) conducted by the IEA, including six waves across 20 years, from 1999 up until the most recent wave at the time of writing in 2019. For the purposes of this thesis, only data from grade eight students and mathematics teachers are included. Second, data are taken from the most recent (at the time of writing) Teaching and Learning International Survey (TALIS) conducted by the OECD in 2018 sampling teachers of lower secondary students, or 'population 2'. TIMSS is conducted every four years, while TALIS is conducted every five years. Each of the datasets have advantages related to their specific aims and foci.

Because of the sampling structures and strata of both TIMSS and TALIS, when adequate participation rates are reached, the data are representative at the country level. Both TIMSS and TALIS have a nested structure. In TALIS, teachers and classrooms comprise level 1, nested within schools (level 2), nested within education systems (level 3). TIMSS has an additional student level. While the TIMSS and TALIS reports describe in detail much of the findings, they generally do not attempt to explain the variation in outcomes or make inferences about directions of such effects (Gustafsson & Rosèn, 2014). This so-called 'secondary analysis' is mostly left to individual countries and researchers (Gustafsson & Rosèn, 2014).

TIMSS

TIMSS is the only international large-scale assessment study with many countries to link student and teacher data, and teachers and teaching are a prominent focus of the study. TIMSS focuses on mathematics and science, and therefore includes teachers of mathematics as well as teachers of science and the relevant subdomains. The TIMSS assessment framework is based upon curricula of the various countries included and is created in collaboration with research coordinators in every country (Mullis et al., 2016). Because of the breadth of the

TIMSS questionnaire, students are not tested on the complete test, but rather receive a subset of items. To generate individual student test scores (also known as plausible values), the item Response Theory (IRT) psychometric approach is used (Lord & Novick, 1968). Alongside the test items are survey questions related to student background, such as their motivation, beliefs about learning, as well as factors such as gender, age, and information about their home educational resources and family socioeconomic background.

TIMSS employs a two-stage stratified sampling design, which samples schools according to previously determined strata proportional to their size as well as whole classrooms within the schools to cover a range of nationally representative educational contexts. TIMSS has minimum participation requirements for a country to be included, which calls for a minimum of 150 schools to be sampled per grade, and a minimum of 4000 students (Mullis et al., 2016). If countries do not have adequate participation rates, they can be excluded from the results. However, the criteria in certain cases relating to the strata are relatively opaque and defined by each participating education system, but often are related to average achievement levels, socioeconomic composition, location, and school type (Mullis et al., 2016).

Teachers of the assessed classes in TIMSS are required to respond to a questionnaire, along with the principals of each school. Teacher response rates vary across countries (Mullis et al., 2016), with between 1000-8000 teachers per country sampled. There can also be more than one teacher sampled per student, leading to the necessitation of using the appropriate teacher weights. TIMSS provides survey weights for students, teachers, schools, and countries, which account for the sampling structure of the study (Mullis et al., 2016). Survey weights for students and schools are inversely proportional to their probability of being sampled, while country (or 'senate' weights) account for the differing sample sizes between participating education systems.

TALIS

The OECD TALIS study focuses not on student outcomes (although just over a handful of studies participate in the PISA-TALIS link, an effort to connect the TALIS data to outcomes in student performance) but on teacher outcomes. TALIS focuses on collecting contextual data from teachers and principals, related to the characteristics of their target classrooms and schools, the working environment, their job satisfaction, personal characteristics, qualifications and

teaching processes. For the purpose of this thesis, only teachers from ISCED level 2 (lower-secondary or grade eight) are included in the analyses.

TALIS employs a similar two-stage stratified sampling procedure as TIMSS. Within each education system, 200 schools are drawn with a probability proportional to their size. Within each of these schools, 20 teachers are randomly selected. The OECD also has requirements for participation, with 75% of schools and 75% of teachers required to complete the survey (OECD, 2019). TALIS provides weights to account for the sampling structure for teachers and schools but does not provide weights to account for using multiple countries in the same analysis (such as the senate weights from TIMSS).

Other data sources

Two variables related to stratification (educational tracking between schools and ability grouping within schools) are taken from PISA data from the OECD in cycles 2018 (and 2015 for 2 education systems). These variables were not available in the TALIS dataset from 2018. The OECD uses a similar sampling procedure as TALIS, but the data are not meant to represent teachers in a country (as in TALIS) but rather students in school at the grade eight level.

In certain cases, to control for the social and economic circumstances in a country, the Human Development Index (HDI) is used. HDI is computed based on a longevity/health dimension (life expectancy at birth), an educational dimension (expected years of schooling and mean years of schooling), and a standard of living dimension (gross national income per capita). For more information on how this scale is constructed, see UNDP (2020).

Variables

The studies draw from the student, teacher, and principal questionnaires from the large-scale assessment studies. More information regarding how each of these questionnaires are constructed can be found in the Technical Reports from the IEA and the OECD (IEA, 2019; OECD, 2019).

Student-level variables

Socioeconomic status and student-level controls

In Studies I and II, socioeconomic status is operationalized as a composite measure of the number of books in the home and parental level of education. To generate individual SES scores for each student, a factor analysis is conducted with

a polychoric correlation matrix. The factor analyses are conducted within each country-by-year unit, so scores for a student (and their subsequent designation as in the top or bottom SES groups) are relative to their position in the country-year cohort, not to scores across the whole sample of countries and years.

TIMSS produces a 'Home Educational Resources' scale using the aforementioned variables as well as home possessions. However, this scale is not consistently included in the questionnaires between 1999 and 2019, limiting its usefulness to analyses across this period. Home possessions are not included in the country-year SES scales as there are even more doubts about their comparability over time (Pokropek et al., 2017).

Following Chmielewski (2019), we employ a percentile method for each country-year. This allows us to 'compare students at the top and bottom relative position within a socioeconomic distribution, even as the absolute meanings of these positions change' (Chmielewski, 2019, p. 525). In this way, socioeconomic status is conceptualized as a positional good with relative rather than absolute significance. We take students who score below the 33rd and above the 66th percentile of the SES scale to represent those in low- and high-SES families. We use thirds instead of quartiles so as to ensure maximum statistical power and minimum sampling error for each country-year estimate. In a next step, we examine whether pooled the gaps widen at more extreme ends of the socioeconomic status scale and compare the results to 90th and 10th percentiles.

Parental education is an ordered categorical variable, with (1) some primary or lower secondary, (2) lower secondary, (3) upper secondary, (4) post-secondary, non-tertiary, (5) short-cycle tertiary, (6) bachelor's or equivalent, (7) postgraduate degree. After the year 2011, another category was added to indicate differences between postgraduate and doctoral degrees, and we have made them into one category for all cycles in the study, reducing the categories to seven. The number of books in the home is an ordered categorical variable from (1) 0-10, (2) 11-25, (3) 26-100, (4) 101-200, (5) more than 200 books.

In Study II, additional student background controls are introduced, including language spoken at home (coded as 1 if a student speaks a language other than the test language at home), gender (coded 1 if the student is female), foreign-born status (coded as 1 if the student was born outside the country of testing), and age.

Mathematics achievement

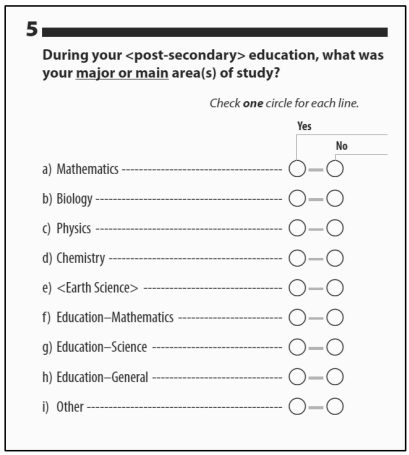
In Study II, mathematics achievement is introduced as the main dependent variable, interacted with the within country-year socioeconomic status factor

scores. Mathematics achievement has long been recognized as the subject domain the most affected by schooling inputs and resources (Murnane, 1974). It is also frequently positioned to be the subject domain most validly comparable across educational contexts, particularly in its relation to teachers (Akiba et al., 2007; Baumert et al., 2010; Goe, 2007). The international mathematics scale has a mean of 500 and a standard deviation of 100. TIMSS produces five plausible values which estimate a range of student scores given their background and responses to given items (Wu, 2005). The plausible values are taken into account using pooling approaches outlined by Rubin (1987) and Gonzalez (2014).

Teacher/classroom-level variables

Teacher qualifications

In Studies I and II, teachers are coded as 'novice' or 'out-of-subject'. Teachers with 5 or less years of experience are coded as 'novice'. The category of 'out-of-subject' teacher is created based on specialization-related information collected from the teacher questionnaires in each cycle. Teachers reported on which subject they studied in their post-secondary educations. Teachers could choose 'mathematics' or 'mathematics education', as well as other subject areas such as science or history. Teachers citing 'math' or 'mathematics education' or both thus may have some level of formal specialization in mathematics. We therefore designate a teacher as an 'out-of-subject' mathematics teacher if they choose neither 'mathematics' nor 'mathematics education.' See Figure 2 for this measure.



5

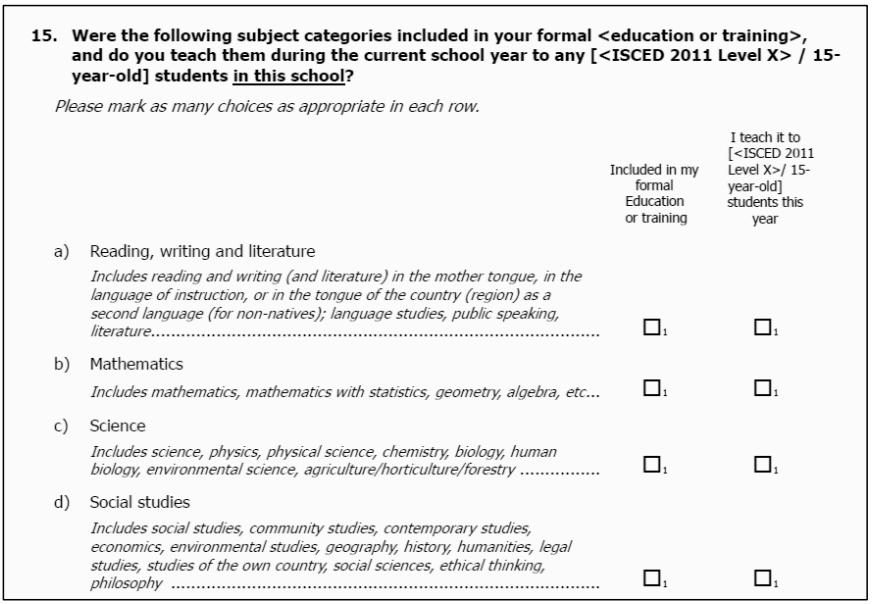
During your <post-secondary> education, what was your major or main area(s) of study?

*Check **one** circle for each line.*

	Yes	No
a) Mathematics	○	○
b) Biology	○	○
c) Physics	○	○
d) Chemistry	○	○
e) <Earth Science>	○	○
f) Education–Mathematics	○	○
g) Education–Science	○	○
h) Education–General	○	○
i) Other	○	○

Figure 2 Specialization Item in the TIMSS Teacher Questionnaire

In Studies III and IV (using TALIS data), the same classification of novice teacher is employed, whereas a specialized teacher is considered to be 'out-of-subject' if they are currently teaching a subject in which the content was not the focus of their degree. In this way, both of the specialization measures focus on 'content knowledge' and not 'pedagogical content knowledge' (Baumert et al., 2010; Schulman, 1986), while TIMSS makes this distinction more explicit. See Figure 3 for an example of this measure.



15. Were the following subject categories included in your formal <education or training>, and do you teach them during the current school year to any [<ISCED 2011 Level X> / 15-year-old] students in this school?

Please mark as many choices as appropriate in each row.

	Included in my formal Education or training	I teach it to [<ISCED 2011 Level X>/ 15-year-old] students this year
a) Reading, writing and literature *Includes reading and writing (and literature) in the mother tongue, in the language of instruction, or in the tongue of the country (region) as a second language (for non-natives); language studies, public speaking, literature*	☐1	☐1
b) Mathematics *Includes mathematics, mathematics with statistics, geometry, algebra, etc...*	☐1	☐1
c) Science *Includes science, physics, physical science, chemistry, biology, human biology, environmental science, agriculture/horticulture/forestry*	☐1	☐1
d) Social studies *Includes social studies, community studies, contemporary studies, economics, environmental studies, geography, history, humanities, legal studies, studies of the own country, social sciences, ethical thinking, philosophy*	☐1	☐1

Figure 3 Specialization Item in the TALIS Teacher Questionnaire

In Studies III and IV, teacher variables are interacted with class socioeconomic composition variables. In TALIS (2018), teachers were asked to describe their 'target' classrooms (the first class the teacher taught in the school after 11 AM last Tuesday). They are to mark the percentage of students from socioeconomically disadvantaged homes. It ranges from None (1), 1-10% (2), 11%-30% (3), 31%-60% (4), and More than 60% (5). Other classroom controls included in Study III include the proportion of students with immigrant or migrant backgrounds, the proportion of students without the language of the test as the mother tongue, and

the proportion of students with special needs. Each of these variables are coded None (1), 1-10% (2), 11%-30% (3), 31%-60% (4), and More than 60% (5). Study IV extends the analysis from Study III to include teacher gender (coded 1 as female) and subject domain (math/science versus all other subjects). While Studies I and II focus on mathematics teachers alone, Studies III and IV focus on all teachers as well as mathematics and science teachers. Study IV controls for teacher's job satisfaction in the profession (T3JSPRO) and overall teacher self-efficacy (T3SELF), which are scales provided by TALIS. Unlike TIMSS, TALIS (2018) provides the results of measurement equivalence testing, and both of these scales reach measurement invariance at the metric level. For more information about how these scales are constructed, see the TALIS technical report (OECD, 2019a).

Teacher turnover intentions

Study IV uses the 'turnover intentions' of teachers as the main dependent variable of interest. This variable measures whether teachers would like to change schools should the opportunity arise. It is a 4-point Likert type scale from Strongly Disagree (1), Disagree (2), Agree (3), and Strongly Agree (4). To make the variable more validly comparable across countries, it is recoded as a dichotomous variable, between either agree or disagree to any degree.

School-level variables

In both TIMSS and TALIS, school socioeconomic composition is measured by asking the principals about the proportion of socioeconomically disadvantaged students in the school. In TALIS, it is coded as None (1), 1-10% (2), 11%-30% (3), 31%-60% (4), and More than 60% (5). In TIMSS, it is coded with just four categories instead of five: 0-10% (1), 11-25% (2), 26-50% (3), More than 50% (4). Similarly, both TIMSS and TALIS ask principals about the location of their schools. In TIMSS, if a school is located in a town with 15000 people or less, it is coded 1, as a rural or small-town school. In TALIS, if a school is designated in an area of 15,000 people or less, it is coded 1 as a rural or small-town school. In TALIS, school type is measured as 1 if the school is privately funded. TIMSS does not include consistent information related to school type between the years 1999 and 2019. Study II using TIMSS data also controls for whether the principal deems that a 'shortage of material resources' impacts the quality of instruction at the school level, with (1) Not affected, (2) Affected, (3) Affected a lot.

Other control variables at the school level for Study II from TIMSS include whether the school practices ability grouping (a dichotomous variable), as well as the average student-teacher ratio (reported by teachers aggregated to the school level). These three variables (shortage of material resources, ability grouping and average class size) are available in 5 out of 6 TIMSS cycles, and we impute the values within countries across years to be able to include them in the model.

System-level variables

System-level variables were taken from various international large-scale assessments and other sources. In some cases, the variables were available directly from certain data sets (i.e., between school tracking and HDI), but in many other instances the data came from student- or school- level variables aggregated to the country level. These lower-level variables had missing data which needed to be accounted for before aggregating the variables to the system level. This was done using multiple imputation which is further explained later in this chapter. For an overview of the data sources and coding of the system level variables, see Table 2.

Table 2 Coding and data sources of system-level variables

Variables	Aggregated from	Source	Coding
Exposure to teacher qualifications	Student-level	TIMSS	Low-SES exposure rate to novice and unspecialized teachers
Teacher sorting	Student-level	TIMSS	(Low-SES exposure rate to novice and unspecialized teachers) – (High-SES exposure rates to novice and unspecialized teachers)
Social segregation	Student-level	TIMSS	Intra-class correlation coefficient of student SES explained by the school level
HDI	No aggregation	UNDP	Continuous index supplied by UNDP (0-1)
Between-school tracking	No aggregation	PISA	Dummy variable for schools which practice tracking from ages 10-14
Within-school ability grouping	School-level	PISA	% of schools where principals answered 'yes' to school ability grouping
Accountability	School-level	TALIS	% of schools where teacher appraisal is based on student test scores
Autonomy	School-level	TALIS	% of schools solely responsible for teacher staffing
Competition	School-level	TALIS	% of schools with two or more competing schools
Teacher shortage	Teacher-level	TALIS	% of schools where principals report a shortage of qualified teachers hinders instruction

Descriptive statistics

Descriptive statistics and missing data percentages for each of the studies in the thesis can be found in Appendix A and B. In general, Table A shows very high amounts of missing data for the variables related to parental education. In general, the mean mathematics achievement pooled across the participating cycles and education systems is 491 TIMSS score points. The average teacher opportunity gap is 7 percentage points in the case of novice teachers and 4 percentage points in the case of out-of-subject teachers. The average proportion of student SES variance explained by the school clustering is 21.8%.

Table B shows that about 22% of teachers would like to change schools given the opportunity across all education systems. Approximately 18% of teachers were

in their first 5 years on the job, compared with 25% who were teaching a subject without a relevant university degree. These statistics are similar to those found from TIMSS data (Table A) describing the overall percentage of students with novice teachers (0.13) and without a subject matter specialization (0.17) in that the latter percentage is higher than the former. When the sample was narrowed to mathematics and science teachers, however, 9.5% were recently qualified and about 7% were teaching the subject without a relevant university degree.

Method

The thesis utilizes a variety of statistical methods from various disciplines within the field of economics and the educational sciences. The first two studies (I and II) utilize descriptive and econometric techniques such as panel regressions with fixed effects (Schlotter, Schwerdt & Woessman, 2014), the latter of which provides a framework for making causal interpretations of the relationship between teacher sorting and educational inequity. The second two studies (III and IV) use a multi-level modelling approach to investigate the relationships between institutional differentiation and different aspects of socioeconomic teacher sorting.

The analyses were conducted using a variety of analytical softwares. To link the original data across education systems and time, the IEA IBD Analyzer is used. The IDB Analyzer can be used for analysis as well as data preparation for IEA studies and other international large-scale assessments. The data are originally exported as SPSS files and then further exported to other softwares for analysis (RStudio and Mplus). Descriptive statistics were carried out using SPSS. Analyses in Studies I and II were conducted using RStudio (version 1.4.1103, 2009-2021). RStudio is a free programming language for statistical computing and graphics. RStudio allows for a wide range of statistical methods to be used, including missing data imputation, multilevel modelling, econometric analyses, and descriptive techniques used in this thesis. Analyses in Studies III and IV were conducted using the Mplus (Muthèn & Muthèn, 1998-2017) statistical software. Mplus allows for multi-level modelling with both latent and manifest variables.

Many of the concepts in educational sciences are abstract and therefore not observable through a single test item. In certain cases, several indicators are used to grasp an underlying construct, and in others, only a single indicator is used. This is primarily important for socioeconomic status. Although such operationalizations provide a good approximation of 'unobservable' phenomena, they will always include a degree of measurement error.

Descriptive statistics and confirmatory factor analysis

To construct the country-by-year composite measures of SES, confirmatory factor analysis was used with polychoric correlations between the number of books in the home, and the parental educational attainment of the student's mother and father. In this way, socioeconomic status is conceptualized as a single latent (unobservable) construct comprised of various related but separate dimensions (Brown, 2015). The score for each individual student reflects their relative position on the SES scale which is based on the correlations between the different observed factors. The socioeconomic status composite score used in Studies I and II was derived in part from theoretical concept of socioeconomic status (Bourdieu, 1986), and in part due to the availability and cross-temporal comparability of the indicators. TIMSS does not provide test items related to parental income or occupation, and the home possessions items are not consistent in their wording over time. In this case, a high factor loading suggests that a high proportion of variance in the indicator can be explained by the latent construct, whereas a low factor loading suggests the opposite. Factor loadings for latent socioeconomic status constructs for each education system and cycle are presented in Appendix C.

To determine whether inequities in the distribution of teachers by experience and specialization have widened over time, logistic regressions were used. Specifically, within each education system, the status of novice or out-of-subject teacher was regressed on student socioeconomic status as well as an interaction term comprised of socioeconomic status and time. If the interaction term coefficient between low SES students and time was positive and statistically significant, the gap was considered to be widening. To deal with the hierarchical nature of the data in Study I, standard errors are adjusted for clustering by school. This approach to error estimation will be further described in a later section in this chapter.

Panel data regressions with country fixed effects

As outlined by Strietholt et al. (2014), Cordero et al. (2017), Gustafsson (2013), and others, frustrations around the limitations of correlational analyses for the purposes of policy-related research have led to an uptick in the number of quasi-experimental and causal approaches in educational effectiveness research. One such approach is the so-called country 'fixed effects' approach with panel data. The general idea of this method is to use the subject units (i.e., the countries) as

their own controls and limit the variation to within-subject change in both the independent and dependent variables. This can be achieved due to the longitudinal nature of large-scale assessment data at the country-level (Gustafsson, 2013). The country fixed effects are introduced by either adding country dummies to the models or by using a 'within' estimator in statistical programs.

Panel regressions build upon traditional linear regression models and include a time-dimension at each level. By restricting the variation to within-units, all time-invariant biases are controlled for. Hanushek et al. (2013) cite the cultural importance of education, commitment of families to their children's education, and the state of development of societal and economic institutions) as examples of 'time-invariant' cultural and institutional factors. Other unmeasurable elements related to culture may also be in the mix. By positioning the independent variable at the institutional level, the so-called 'selection effects' from lower- levels are taken into account as these selection effects are likely to be consistent within countries. In the case of teacher sorting, this may include families choosing schools for certain reasons or teachers seeking out certain schools. Such concerns may bias estimates of the impact of teacher quality on student achievement at lower levels. The remaining concern is whether any time-varying factors are associated with the pattern of teacher sorting in a country. By including social segregation and overall exposure rates to novice and out-of-subject mathematics teachers as time-varying controls, the most obvious potential confounders are taken care of, but the analysis cannot guarantee that all potential confounders are excluded.

Causal approaches using international large-scale assessments are not without criticism. Rutkowski and Delandshere (2016) note that the conditions for making causal conclusions are rarely met by large-scale assessment data. The authors first delineate between causal explanations and causal descriptions. The former is focused on the mechanisms and conditions by which a relationship functions, and the latter describes the consequence of varying the cause on the effect. They use the validity framework of Shadish et al. (2002) and note that causal claims are affected by a number of factors related to validity and reliability. Specifically, 'the issue is whether the causal mechanisms or causal explanations for a particular phenomenon are comparable across contexts' (p. 5). Another study by Jerrim, Lopez-Agudo, Gutierrez, and Shure (2017) demonstrate that survey designs of large-scale assessments are sometimes at odds with causal estimation approaches, most importantly for analyses which are restricted to within-students.

A final concern of fixed effects panel models is the large reduction in variation imposed by limiting the estimation to within units (in our case countries) over time.

Mummolo and Peterson (2018) offer first a critique of the ways in which results from FE models have been misinterpreted, and second, a way forward for researchers. Estimates produced by fixed effects models assume a full unit increase in the treatment variable. However, such variation in the independent variable is rarely achieved when the analysis is restricted to within-units over time. Mummolo and Peterson (2018) therefore argue that the effect sizes should be calculated according to the within-unit standard deviation in the treatment variable (the standard deviation of the treatment variable after introducing the country and time fixed effects). These considerations are taken into account in the interpretations of results from Study II.

Multilevel modelling

This project makes use of multilevel models to investigate the relationship between factors at the country/system level and the teacher/classroom level using a cumulative model building process following the recommendations of Bryk and Raudenbush (2002) and Sommet and Morselli (2017). Multilevel models allow for cross-level interactions and random slopes, i.e., the assumption of heterogeneity of the level 1 effect as a function of the level 2 or 3 effect can thus be tested.

Multilevel modelling can examine the decomposition of variance in a variable across levels and is a useful way to construct measures of between- versus within-unit differences. The intra-class correlation coefficient (ICC) can be used in such an instance, as in the degree of variance in socioeconomic status of students which can be explained by the school level as in the case of Study II. A high value (closer to 1) would indicate a high degree of similarity between students within schools as opposed to between schools, whereas a lower value would indicate the opposite. The intra-class correlation coefficient in the dependent variable of interest can also be included to determine whether a multilevel framework is needed (LeBreton & Senter, 2008). A low ICC related to the country level proportion of variance explained in the dependent variable would indicate that the variable of interest varies more within countries than between countries. The formula is as follows:

$$\text{(1) ICC school} = \frac{var(u_{0j})+var(u_{0k})}{var(u_{0j})+var(u_{0k})+(\frac{\pi^2}{3})}$$

$$\text{(2) ICC country} = \frac{var(u_{0k})}{var(u_{0j})+var(u_{0k})+(\frac{\pi^2}{3})}$$

Where $var(u_{0j})$ = the school level variance, $var(u_{0k})$ = the country- level variance, and ($\pi^2/3$) = the assumed teacher level variance of 3.29, as logistic regression does not include a level 1 residual. Using the above formulas, we therefore calculate the ICCs for the dependent variables in Studies III and IV below.

Table 3 Intra-class correlation coefficients for dependent variables

Variable	School-level variance	Country-level variance
Tnew	.056	.033
Tnspec	.220	.235
Tchange	.081	.050

Table 2 shows relatively low proportions of variance explained at the country-level for novice teacher designation as well as turnover intentions. The recommendation for including a hierarchical modelling structure generally requires an ICC of .05 (LeBreton & Senter, 2008), but others find group-level relationships when only 1% of the variation is accounted for (Bliese, 1998). Since the dependent variable of interest for Studies III and IV is not the teacher qualifications themselves but the relationship to classroom SES (the slope), another important dimension is whether the slope varies significantly across the education systems. Indeed, this is confirmed by the modelling. Therefore, the estimation is conducted as a multi-level model with the low-ICC values in mind. On the other hand, the ICC values for teacher specialization qualification were much higher, with around 23% of the variance explained by the country-level. This suggests that the probability of receiving an 'out-of-subject' designation varies more between countries than the probability of having a novice teacher designation or the probability of wishing to change schools.

If the relationship between two level 1 variables varies significantly across clusters, this warrants the inclusion of a so-called 'random slope' in the modelling procedure (Raudenbush & Bryk, 2002). In Studies III and IV, the degree of variance across countries in the relationship between the socioeconomic composition of a teacher's classroom and their likelihood of changing schools or having a lower qualification level is tested. For example, in Study IV, the main research question is whether the relationship between socioeconomic status and teacher turnover intentions is moderated by school system-level accountability practices. An example of the model is depicted in Figure 4.

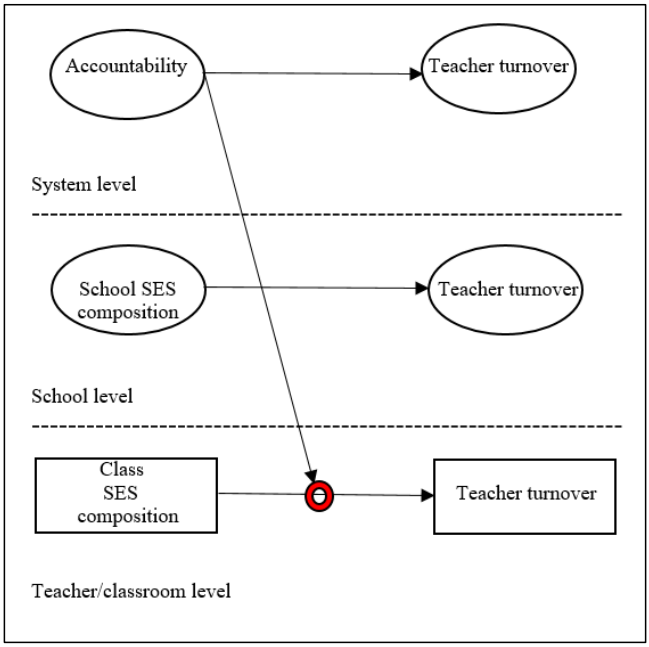


Figure 4 Three-level random slope model

In the above model, the random slope is portrayed as the regression of teacher turnover on the socioeconomic composition of the teacher's target classroom. First, it is necessary to determine whether this slope varies significantly across education systems. Next, the slope is regressed on the level of accountability in the school system. If the slope is positively related to the level of accountability, the effect can be described as 'anti-compensatory.' In other words, where system-level accountability is higher, there is a stronger relationship between teacher turnover intentions and classroom socioeconomic composition. The slope can also be regressed on other system-level variables not shown in the above figure such as the level of economic and social development of a country, or HDI. In this case, if the slope is positively related to HDI, we can conclude that inequitable socioeconomic teacher turnover would be more prevalent in more economically developed countries.

Estimating standard errors

Gonzalez (2014) outlines several sources of uncertainty in international large-scale assessments, which include translation of the instruments, administrative

conditions, scoring, data entry and processing, among others. While the international large-scale testing bodies account for these former types of uncertainty, two other sources of uncertainty, sampling error (related to the selection of respondents) and measurement error (related to the selection of items on the test) require an appropriate estimation of the standard errors via either hierarchical data modelling or cluster robust standard error estimation. Because students do not receive all items on a test, there is an additional degree of error which needs to be addressed. Similar to the ways in which Rubin (1974) suggests accounting for uncertainty in estimates with missing data, Gonzalez (2014) outlines that standard errors related to plausible values (when achievement data are used in the analysis) need to incorporate both the sampling and measurement error.

Another dimension of the error which must be accounted for relates to the clustered nature of the data. Statistical methods are often based on a 'simple-random' sampling technique, where the sampled units or observations are independent of one another, and therefore that the unexplained variance associated with their responses will be uncorrelated. In other cases, however, as in the case of international large-scale assessment data, the sampling procedure has an inherently nested structure. If analyses do not account for these relationships between students within schools and countries, the estimated standard errors will be underestimated and the statistical significance of certain variables may be overestimated. Using multilevel modelling accounts for this type of clustering and the standard errors are accurately estimated in such cases. In the case of the descriptive regressions or the panel models, errors can be clustered at various levels (i.e., class, school or country). The reasoning for including such 'robust' standard errors is similar to the motivation for multi-level modelling, that unobserved components in the outcomes (i.e., the error) are often similar between units (violating the assumption of heteroscedasticity) due to the sampling design (Abadie, Athey, Imbens & Wooldridge, 2017). Abadie et al. (2017) also note that clustering should be based on the level of treatment. In Study I, the standard errors are clustered at the country-level. This is the best option as it includes lower-level clustering in the large-scale assessment sampling design (i.e., school level), as well as the 'treatment' clustering of socioeconomic teacher sorting at the country-level. It also accounts for heterogeneity in the treatment effects by country. As Study I conducts within-country analyses, standard errors are clustered by school.

Missing data

Survey data is often vulnerable to issues related to attrition and incomplete responses on the questionnaires. In general, the characterization of missing data itself can be classified as random or systematic. It is relatively rare to have missing data which follows a completely random pattern—this pattern is called missing completely at random (MCAR). More often, the missing pattern of data can be related in a probabilistic fashion to certain observed variables such as student background or socioeconomic status; this pattern of missing data is denoted as missing at random (MAR) (Caro, 2018). The missing data in international large-scale assessments is generally considered to fall under this category of MAR as it is related to student background variables.

Missing data is handled in two ways. In most cases, the package 'mice' (Van Buuren & Groothuis-Oudshoorn, 2011) in RStudio is used, which stands for multivariate imputation by chained equation. MICE is a flexible iterative technique which allows for the specification of the imputation method to vary based on the variables based on the available data one 'chain' at a time. In the first two studies, the multiple imputation model is created based on student socioeconomic background variables (parental education, books in the home, home possessions), plausible values in both mathematics and science, school composition and location, and country and year dummies. Mplus, on the other hand, handles missing data through the full-information maximum likelihood (FIML) or a Bayesian method, depending on estimator used. In Studies III and IV, multiple imputation is used in certain cases, such as the within-country models as well as to generate the country-level aggregate measures. Mplus is used to fill in the leftover missing values (which are generally very low) during the multi-level analysis. To see the proportions of missing data for each variable, see Tables A and B in the Appendix.

Frequentist and Bayesian statistical approaches

There is a strong connection between multi-level modelling and Bayesian estimation (Kreft and De Leeuw, 1998), whereby when the number of higher-level units is small and the data unbalanced, there may not be enough information for correct estimations using frequentist approaches (Raudenbush & Bryk, 2002). Others insist on using Bayesian estimation methods for country-level effects (Bryan & Jenkins, 2016). Put very simply, the difference between Bayesian and frequentist approaches is that Bayesian estimation approaches take into account

the probability of hypotheses and data as prior conditions, whereas frequentist (i.e., maximum likelihood estimation, ML) approaches do not take into account the probability of the hypothesis. In other words, ML estimation views model parameters (i.e., variances) as constants and Bayesian estimation views model parameters as variables. Muthèn (2010) notes that Bayesian estimation is more useful for computational power, parameter estimates, and new model types. The 'priors' used in these models are diffuse or non- informative (Asparouhov & Muthèn, 2021). According to Muthèn (2010), when using diffuse priors, ML and Bayesian results are expected to be close in large samples.

The Bayesian estimator is used in Studies III and IV as specific by Mplus due to the categorical outcome variable, as the computation required is too heavy for ML estimation (Muthèn, 2010). According to Muthèn (2010), to estimate model fit, the posterior predictive p-value (PPP) should be above 0.05, indicating a non-significant Chi-square value. The PPP values for the random slope models are 0.75 (unspecialized teachers), 0.50 (novice teachers), and 0.58 in the case of teacher turnover intentions. A limitation of using the Bayesian estimation method is that the survey weights from TALIS are not able to be taken into account which is discussed in more detail in the articles. The interpretation of the estimates

Measurement invariance

In many cross-national studies, comparisons of latent constructs in particular require the investigations of measurement equivalence (or 'invariance') across groups. Latent variables should have the same meaning across groups. This can happen at the configural level, where the structure of latent measurement models is equivalent, the metric level, where the factor loadings across latent variables are equivalent, or the scalar level, where intercepts of latent variables are equivalent. In most cases, only the metric level of invariance is reached (Nilsen & Gustafsson, 2016). Studies I and II make use of the latent construct of 'socioeconomic status,' and the factor loadings are presented in Table C in the Appendix. However, because these are constructed within each country-year cohort traditional measurement invariance group testing is not applicable or particularly relevant. As the analysis is conducted within each country, only the positional meaning of SES is of relevance. Studies III and IV utilize overall job satisfaction and teacher self-efficacy in the models, both of which reach the level of metric invariance reported by TALIS. All other variables included in the analyses are manifest variables.

Validity and reliability

Large-scale international assessments employ a host of quality assurance practices in order to ensure statistical validity and country-level reliability in their findings. However, the tests have come under criticism for their inherent limitations in a number of areas concerning the integrity of constructs and measurements as well as the effect of sampling decisions across educational systems (Anders et al., 2021; Jerrim & Micklewright, 2014). Large-scale testing bodies such as the IEA and the OECD develop the surveys on the basis of theoretical frameworks by a host of experts in education and conduct field tests of the surveys to test out various issues related to construct validity before the widespread administration of the tests.

Validity

Shadish, Cook and Campbell (2002) define validity as 'the approximate truth of an inference or a knowledge claim' (p. 33) and differentiate between construct validity, internal/external validity, and statistical validity. Internal validity relates to the nature and direction of the relationships between variables. For instance, to what degree can researchers claim that one variable (X) influences another (Y) in a particular direction? Throughout the thesis, unless otherwise specified (as in the case of the panel models), relationships are considered to be correlational, and few inferences are made about the direction of such relationships. Internal validity, however, played a large part in focusing on teacher qualifications as the main indicators of teacher quality and competence, as they are not amenable to classroom or school context in the way that instructional practices would be, for instance. A similar line of thinking explains the dependent variable (the random slope) depicting the relationship between classroom socioeconomic composition and teacher qualifications in Study III. In Studies III and IV, no claims are made with respect to the direction of the relationships between institutional variables and socioeconomic teacher sorting as the findings are based on associations. There may be selection effects at the system level, for example, which constrain the interpretation of the findings.

Statistical conclusion validity is related to the concept of internal validity, and deals with the appropriate use of methods to make certain inferences and interpretations of the findings; for instance, ensuring that causal inferences are not being made from correlational analyses. As has been discussed throughout the text, observational data presents a host of challenges for making causal inferences. Introducing fixed effects and panel data techniques is one way to reduce these

limitations. However, there are still threats to the statistical validity of the conclusions from Study II despite the presence of the fixed effects. Namely, there is a threat regarding other time-varying characteristics which may be associated with socioeconomic teacher sorting and which are not included in the modelling. It is for this reason that the study focuses on making a causal description as outlined by Rutkowski and Delandshere (2016) rather than a causal explanation.

Construct validity, a long-discussed concept (Cronbach and Meehl, 1955; Shadish et al., 2002) pertains to how accurately a construct is conceptualized, measured and portrayed. For example, the exclusion of parental income and occupation from the TIMSS questionnaires may present a credible threat to the construct validity of socioeconomic status in Studies I and II. Another related threat to construct validity may be the accuracy of student reports of their parents' educational attainment levels and the number of books in their home (Engzell, 2019; Jerrim and Micklewright, 2014). One of the major reasons for separating teacher qualifications in the construction of teacher qualification gaps in Studies I and II as well as the sorting measures in Studies III and IV has to do with construct validity. Although many studies combine teacher qualifications into a composite indicator of teacher quality (i.e., Luschei & Jeong, 2019), there is limited evidence to suggest that these variables are correlated to a high degree. Moreover, there is even less reason to believe that they would be correlated in a similar fashion across countries. Study I shows that patterns of socioeconomic teacher sorting are highly dependent on the qualification in focus, composite measures of teacher competence therefore have limited value for policymakers and school leaders for a host of reasons.

A final but equally important threat to construct validity is the potential for misinterpretation and a lack of nuance introduced by aggregate or higher-level measures of characteristics of school systems common in large-scale assessment research. The example of centralization of the Turkish education system outlined Chapter 4 is a good example of one such potential threat to construct validity. Although Turkey is one of the education systems with the lowest levels of school autonomy over hiring (Özoğlu, 2015), there is still a degree to which teacher preferences are taken into account when teachers are assigned to schools. Are such preferences taken into account in other centralized staffing systems, such as Japan or South Korea? These construct validity threats also constrain the findings in Studies III and IV in terms of their policy implications.

Reliability

The question of whether educational measures are reproducible across contexts is known as 'reliability' (in the case of Shadish et al., 2002, this is related to external validity). Aside from threats to construct validity, threats to reliability are one of the most pressing concerns related to large-scale international assessments. International large-scale assessments employ sampling procedures which in theory allow for inferences to be made in a country- representative fashion (Mullis et al., 2016; OECD, 2019). However, recent work has illuminated inconsistencies in the inclusion and exclusion criteria for certain students across countries (Anders et al., 2021). If a country is systematically excluding a lower-achieving segment of the student population while other countries include such students, comparisons made regarding the effectiveness of educational systems will indeed be heavily biased. As the thesis relies exclusively on international large-scale assessment data, its reliability exists as a direct function of the sampling integrity of the international testing bodies in each education system and over each cycle.

Another complexity related to reliability is introduced when student data are linked to teachers, as in the case of TIMSS. In order for TIMSS data to be country-representative, inferences must be made at the student level (Rutkowski, Gonzalez, Joncas, & von Davier, 2010). However, because the construct of teacher sorting necessarily requires the use of teacher data, the construct is prone to threats to external validity based on sampling error at both the student and teacher levels. Finally, the teacher and principal perceptions of socioeconomic status composition of the school and target classrooms are not able to be verified with student-level data and this limitation may impact the replicability of the findings in future studies.

To sum up, the threats to reliability in the thesis have mostly to do with the TIMSS and TALIS sampling procedures. Although TIMSS samples multiple teachers, they typically sample just one classroom per school. In Study II, a cross-school socioeconomic segregation measure is used (the intra-class correlation coefficient), and this in certain cases may confound the school and classroom levels. This is only an issue, however, in the case that one classroom is sampled per school *and* if this school has a heterogenous socioeconomic student composition and practices ability grouping. This concern is discussed in more depth in Study II, along with a summary of the number of classrooms per school sampled in each education system and cycle. The uncertainty introduced by the

sampling structure is reflected in the estimated standard errors, which are in many cases quite large and emphasized in each of the four studies.

Ethical considerations

Data used in this thesis are publicly available by the IEA and the OECD. Data are fully anonymized, and there is no threat to uncovering school identities based on the information provided in the studies. Due to this, there was no requirement for ethical authorization by the University of Gothenburg or any other ethics committee.

The main ethical issues pertaining to the thesis are related to the potential political influence of the study conclusions on students, teachers, and educational systems. Nóvoa and Yariv-Marshal (2003), Biesta (2015), Johansson (2016), and Komatsu and Rappleye (2017) are just a few of the scholars who underscore the pitfalls related to the political influence of large-scale assessments. These have to do with the homogenization of global educational curricula due to the large-scale assessments, the proliferation of 'teaching to the test,' the drive towards neoliberalism and an outsized focus on measurement and comparison and educational 'monitoring' at the expense of individual student differences and strengths.

The influence of large-scale assessments and secondary analyses of the studies is especially problematic if incorrect interpretations of the findings are made and causal inferences are made when they are not warranted (Gustafsson, 2008). The studies have therefore attempted to be transparent and up front about their limitations and methodological approaches in terms of the possibility of policy recommendations. With these cautions in mind, the following section presents and interprets results from the four empirical studies.

Chapter 6 Results and Discussion

In this section the findings from the four empirical studies are summarized and discussed. The results are presented in order of the research questions provided in Chapter 1, (i) *To what extent do education systems display teacher sorting across socioeconomic student groups, and has this changed over the past two decades?* (ii) *Does sorting of teacher qualifications by student socioeconomic background exacerbate mathematics achievement inequity?* (iii) *Which institutional features are associated with inequitable teacher sorting?* And (iv) *Is student performance-based teacher appraisal associated with increases in teacher turnover intentions in low-SES contexts?* Summaries of the four studies answer these questions and an integrated discussion follows.

Teacher sorting across countries and time (Study I)

Teacher qualification gap magnitudes

Socioeconomic sorting of teacher competence and qualifications is repeatedly touted as a key determinant of the socioeconomic achievement gap (OECD, 2018b; Sims and Allen, 2018). The first step was to determine which out of 32 education systems displayed significant levels of socioeconomic teacher sorting in the TIMSS 2019 cycle as well as when the data were pooled across all cycles from 1999, 2003, 2007, 2011, 2015 and 2019. By taking advantage of the TIMSS dataset linking students to teachers, it was possible to investigate this question for each country participating in at least 4 TIMSS cycles and make inferences at the level of students. The teacher qualification gaps were calculated by the difference in exposure rates to novice or out-of-subject teachers between top and bottom SES tertiles for each country and year.

The results show that novice teacher sorting is more pronounced than sorting by mathematics educational specialization. Across 33/66 SES groups, Turkey, Tunisia, Iran, Indonesia and Morocco show statistically significant differences pooled across all cycles, where a much higher share of low-SES students had novice teachers as compared to the higher-SES students. When focusing only on

the more extreme ends of the socioeconomic spectrum (differences between 90/10 SES groups), Jordan, Botswana, Romania, the United States, Israel and Singapore also displayed statistically significant differences. Based on these results, sorting by experience appears to be more common in economically developing countries, with some exceptions. The maximum degree of sorting in percentage points was about 40 in the case of 90/10 gaps and 26 in the case of 33/66 gaps.

By contrast, sorting by specialization appears much less frequently and is practiced to a lesser degree. Just a handful of education systems show statistically significant differences in exposure rates between high- and low-SES students (Chile, Thailand, Australia, Quebec, and Chinese Taipei) even at the most extreme ends of the socioeconomic spectrum. Chile displays the highest level of inequity with a 23 percentage point difference in the pooled 90/10 gaps and 15 percentage point difference in the 33/66 gaps.

Fewer education systems show statistically significant teacher qualification gaps when only the TIMSS 2019 cycle is considered. Just Turkey, Morocco, Tunisia and Indonesia display statistically significant teacher qualification gaps for recently qualified teachers. For sorting by mathematics education, none of the countries in the sample reach statistical significance. Chile, Australia and New Zealand, USA and Canada (Quebec) display gaps of between 5 and 10 percentage points, however.

As these results are one of the few cross-national studies on the magnitude of teacher sorting by these teacher quality indicators using data which links teachers and students, it is difficult to determine how the findings fit with previous research (Luschei and Jeong, 2019). The results are also based on TALIS data which does not incorporate student-level measures. In the single study which uses a similar measure for the math education specialization variable, Akiba et al. (2007) find that Chile, England, Chinese Taipei and New Zealand have the highest gaps from TIMSS 2003 data, indicating some overlap.

In the case of Chile, low-SES students were less likely to have novice teachers but more likely to have unspecialized teachers. This could be due to the lack of job opportunities in rural communities, thereby limiting opportunities for new teachers to enter such schools. It could also be that rural or remote communities are undesirable and that teachers simply do not apply to work in such settings. In such a case, the underrepresentation of novice teachers may be symptomatic of a teacher shortage. Such an interpretation may be further corroborated if an education system has a higher proportion of unspecialized. Indeed, this is the case in Chile. In this way, even though a high proportion of novice teachers is generally

associated with less desirable working conditions in a school, in certain cases, the absence of novice teachers may also reflect a hiring issue.

Taken together, the results show that lower-SES students face a more modest disadvantage in terms of their access to qualified teachers than the literature suggests. There are however education systems which display dramatic levels of inequity in teacher sorting, including Turkey, Morocco, Tunisia, Indonesia, and Chile.

Within-country trends in teacher sorting between 1999-2019

In the next step, Study I sought to determine whether teacher opportunity gaps between high- and low-SES students were widening. To determine this a logistic regression model was conducted in each education system. The teacher qualification status (novice or out-of-subject) was regressed on an interaction between student SES (coded 1 if a student was in the bottom SES third of his or her country-year) and a time factor, with high-SES students as the reference group. The trend lines for both SES groups can first inform us about where the overall proportion of novice or unspecialized teachers is headed. A systematic upward trend in the overall exposure rates to novice mathematics teachers was observed in Australia, Chile, Jordan, and Morocco. A downward trend was observed in Bahrain, Botswana, Chinese Taipei, England, Hong Kong, Malaysia, Canada (Ontario), Sweden, and Tunisia. Such downward trends may be symptomatic of fewer mathematics teachers entering the field as a whole, or of novice teachers exiting the field faster than they can be replaced, or a combination of both. Very few education systems show an increasing proportion of students with mathematics teachers with no mathematics training.

Just three education systems—Chile, Morocco, and Singapore—display widening novice gap magnitudes based on statistical significance at the 95% level. Bahrain, Indonesia, South Korea and Sweden border on statistical significance and reach only the 90% level. Several education systems display a trend towards more equal allocation of novice teachers including Hong Kong, Romania, Slovenia, Thailand, USA, and Quebec. There are also certain education systems where inequity in novice teacher sorting remained large and stable over time, including Iran, Tunisia and Turkey. Only Chile and Thailand show widening inequity in teacher sorting by specialization at the 95% level of statistical significance, and the USA and Chinese Taipei reach statistical significance at the 90% level. By contrast,

no education systems display an increasingly more equal allocation of teachers by specialization.

The results from these descriptive explorations are much more positive than the picture portrayed by the OECD (2018b), in which a majority of countries report that teachers without proper qualifications are hindering student results. Moreover, it appears that based on the indicators included, inequity in teacher sorting is not increasing over the past two decades. There are a few exceptions, however, where teacher sorting is becoming more prevalent, or where it has remained inequitable and largely unchanged over the past two decades. In addition, we know little from the cross-national literature about which institutional features are associated with teacher sorting. Trends in teacher sorting are therefore not easy to generalize to certain clusters of countries; it appears that countries from a range of economic development levels and with a variety of institutional frameworks are moving in similar and opposite directions.

Teacher sorting and mathematics achievement inequity (Study II)

The findings from Study I show sufficient levels of within-country variability in teacher qualification sorting over time. Study II therefore exploits this country-level variability and investigates whether changes in teacher sorting result in changes in mathematics achievement inequity. The study uses a panel regression technique with country fixed effects and data from six waves of TIMSS from 1999 until 2019. In a first step, the teacher qualification gap magnitudes from Study I were saved as country-level predictors for each country and TIMSS cycle. To control for factors at the school and student levels, the analysis included all students in the participating countries and cycles. Mathematics achievement was then regressed on an interaction term between student SES and country-level teacher qualification gaps. Subsequently, the data were subset by SES (low- and high-SES thirds) and relative mathematics performance was regressed on socioeconomic teacher sorting in both cases. The country fixed effects allow for time-invariant sources of bias to be controlled for, as the variation is limited to within education systems over time. As social segregation has been persistent or widening around the world (Gutiérrez et al., 2020; Jerrim, Volante, Klinger, & Schnepf, 2019; Marcinzcak et al., 2016) and is likely to influence student achievement inequity (Burger, 2019), a measure of social segregation is included as a time-varying control.

First, student mathematics test scores were regressed on the overall share of underqualified mathematics teachers in a country. Contrary to the wide body of literature confirming the lower-competence levels of newly qualified teachers, the panel regressions show that an increasing overall share of novice teachers is not significantly related to student test scores. However, a higher share of out-of-subject mathematics teachers in the workforce did result in lower average mathematics achievement with a point estimate reaching marginal significance at the 10% level. In a next step, the interaction term between student socioeconomic status and inequity in teacher sorting was added to the panel regressions. In line with the earlier model, results show differential effects depending on the indicator of teacher competence. There was a continued null pattern of results for the novice qualification gap and a statistically significant effect of socioeconomic sorting by specialization at the 10% level after controlling for socioeconomic school segregation.

To test the robustness of the interaction effect, we conducted another set of panel regressions with the data subset by SES. In this model, the performance of low- and high-SES students (compared to the mean of all students in a country-year) is the dependent variable. These results show that novice teacher qualification gaps are once again insignificantly related to student test scores. Specialization qualification gaps were related to inequitable outcomes, but only in the case of higher-SES students. These students performed higher on the TIMSS test as a function of wider teacher qualification gaps. The test scores of low-SES students were unaffected by socioeconomic teacher sorting across both qualifications. The conclusion is therefore that socioeconomic teacher sorting by mathematics education specialization exacerbates achievement inequity, but it does so via students in the top SES group. These students not only have the advantage of adequately qualified teachers, but also likely have added educational resources at home.

By and large, the results of the panel regressions support previous cross-national findings which call into question the role of teacher experience in the first few years and support the importance of educational specialization in mathematics (Akiba et al., 2007; Sancassani, 2021). Akiba et al. (2007) report that larger teacher specialization opportunity gaps predicted larger SES-achievement gaps in mathematics. These findings support the theory and empirical findings of Schulman (1986) and Baumert et al. (2010) regarding the importance of content knowledge. As a novice teacher designation says little about the degree of content knowledge that a teacher may have, the null pattern of findings also supports this

conclusion. However, as Jackson et al. (2014), Blömeke and Olsen (2019) and others have mentioned, the importance of qualifications is likely to vary across contexts.

Related to Study I, it is also possible that sorting by mathematics education specialization is more strongly correlated with inequities in working conditions across socioeconomic settings as compared to by experience. While the analysis controlled for social segregation at the system-level, as well as material resource shortage, school location, ability grouping and student/teacher ratio at the school level, it is possible that other unobserved variables correlated with teacher sorting are partially responsible for the significant effect. In certain settings, a low proportion of novice teachers may be just as indicative of undesirable working conditions or hiring difficulties as a high proportion of novice teachers. The novice teacher qualification gaps therefore have a greater vulnerability to effect heterogeneity which may dilute the within-country average estimate. Future research should therefore undertake this question in more depth.

The role of socioeconomic school segregation

An important and central assumption of the fixed effects panel model is the time-invariant nature of potential sources of bias. The time-varying control of social segregation (or socioeconomic school segregation) was included in the analysis due to the strong conceptual link between teacher sorting and social segregation (Jackson, 2009; Sacerdote, 2011), as well as the evidence suggesting a rise in social segregation over the past two decades (Gutiérrez et al., 2020; Marcinzcak et al., 2016). The results of Study II show that many education systems display an upward trend in the measure of between-school social segregation. While the original estimate reflecting the influence of socioeconomic teacher sorting by specialization was large and highly statistically significant, the effect was almost cut in half and reduced in significance after the introduction of the social segregation control. This finding suggests that the relationship between teacher competence and student achievement is moderated by social segregation and the socioeconomic composition of students.

In contrast to socioeconomic teacher sorting, socioeconomic school segregation clearly and persistently exacerbates mathematics achievement inequity. The effect of social segregation is likely to encapsulate peer effects and not other factors correlated with lower-SES schools, such as a shortage of material resources, larger class sizes, or ability grouping, as these are under control in the model at the

school-level. This was found in the point estimates of the original panel regressions as well as in the robustness tests when the data was split into SES subgroups. Conclusions from national and international studies underscore the fact that more social segregation disadvantages lower-SES students (Burger, 2019; Strello et al., 2021; Sacerdote, 2011), and the analysis provides even more evidence that the connection between social background and student performance is strengthened when schools are segregated along socioeconomic lines. As societal inequality is becoming more widespread across Europe and elsewhere (Marcinczak et al., 2016; Yang Hansen & Gustafsson, 2016), the findings are relevant for the majority of countries in the TIMSS sample. If this trend continues, the meritocratic ideals which drive most economically developed nations (Ferreira & Gignoux, 2014) may come under threat.

Taken together, the findings have important implications for the distribution and of educational opportunities. Study I demonstrated that virtually all countries in the study sample fail to allocate teachers specialized in mathematics in favour of socioeconomically disadvantaged students as compared to their high-SES peers (i.e., the 'prioritarian' approach as defined by Brighouse & Swift, 2006). The findings also show that an unequal allocation of unspecialized teachers does not appear to drag performance levels down of low-SES students. However, it does show that when higher-SES students have more qualified mathematics teachers they tend to get further ahead. These results contest the claims of those who advocate against the 'sufficiency' approach to the allocation of educational resources. However, if the goal of policymakers is to close the achievement gap via improvements in the lowest performing students, focusing on ensuring adequate levels of teacher specialization may not be enough, supporting the 'priority' approach.

Institutional correlates of socioeconomic teacher sorting (Study III)

Study I sought to establish the magnitude and development of socioeconomic teacher sorting cross-nationally and within countries over time, and Study II extended this line of inquiry to examine whether within-country changes in sorting could be linked to within-country changes in mathematics achievement inequity. The main conclusion from these studies is first that teacher sorting is less of a 'global phenomenon' (Luschei & Jeong, 2019) than previously suggested using data which does not link teachers to students. Nevertheless, the findings from

Study II show that sorting by specialization may in part exacerbate achievement inequity in mathematics. School leaders and policymakers should therefore question whether amenable characteristics of educational systems mitigate or exacerbate sorting patterns, especially by mathematics education specialization.

Using TALIS 2018 data, Study III employs a three-level random slope model to determine whether the relationship between classroom socioeconomic composition and teacher qualifications varies across institution-level features of 46 education systems. The sample of countries ranges from nations with very high educational, health and economic prosperity such as Norway and Finland to nations at the lower end of the HDI scale, including Brazil, Mexico, South Africa, and Vietnam.

First, the odds of having a novice or unspecialized teacher designation as a function of the teacher's perception of the classroom socioeconomic composition were estimated with logistic regressions for each participating education system. A majority of education systems displayed statistically significant cross-classroom socioeconomic sorting of teachers by specialization, though many of the estimates displayed large confidence intervals and a high degree of uncertainty. Fewer education systems reached statistical significance in the sorting of novice teachers, however. Large inequalities were displayed by Finland, UAE, and Austria for sorting by specialization and Turkey, Latvia, and Slovenia for sorting by novice status, for example. Based on these results, cross-classroom sorting of teacher qualifications appears to be more widespread when taking into account the teacher's perception of the classroom socioeconomic composition as compared to when student and teacher data are linked.

Next, the three-level hierarchical regressions were conducted and showed a mixed pattern of results similar to those found in other cross-national investigations of socioeconomic teacher sorting (Luschei & Jeong, 2019). These results underscore the fact that sorting patterns across countries are not widely generalizable and highly dependent on the measurement of teacher quality. School accountability and between-school tracking were correlated positively with the classroom SES teacher qualification slope in one or two cases. Past work has documented a link between teacher turnover and lower job satisfaction in more stringent accountability environments (Feng et al., 2018), and this analysis provides some evidence that teachers are sorted across classrooms in a more inequitable fashion when their appraisal involves student achievement data. It could be that principals allocate these teachers opportunistically, but it could also be that the selection preferences on the part of teachers are strengthened in such

environments. Similarly, school tracking was positively related to sorting by specialization only, with no evidence that novice teachers are more likely to teacher in low-SES classrooms or schools in systems with early tracking.

The classroom SES-teacher qualification slope was positively related to school competition in several of the models. This was the only indicator which displayed similar effects in direction and significance across both teacher qualifications and cross-classroom and cross-school sorting. There is a mixed consensus regarding the relationship between school competition and educational inequity (Strietholt et al., 2019). Some argue that public schools react positively to competition (Hanushek & Rivkin, 2003) introduced by privatization and increased choice via the pressure put on schools to ameliorate school inputs (including teacher quality) to attract students, thereby positively influencing the working conditions of the school. The original hypothesis of the thesis, however, posited that school competition may exacerbate social divides across schools and that teacher qualifications would then mirror these demographic changes as found in other work (Jackson, 2009). Although partially confirmed, the mixed findings require more investigation across other contexts and should be interpreted with a degree of caution.

School autonomy over hiring and firing was related to the classroom SES teacher qualification slope in an anti-compensatory fashion in one case. This pattern was found for novice teacher sorting only. The estimate almost reached significance via the cross-school sorting estimates and had a similar direction. Many have advocated for centralized teacher allocation (Han, 2018; Kang & Hong, 2008; Luschei et al., 2013; Luschei and Jeong, 2019). The rationale is that decentralized systems are less able to mitigate the opportunistic sorting that happens as a function of teacher preferences and market conditions, and school leaders will exhibit such opportunistic behaviour themselves. Luschei and Jeong (2019) write 'nations can, with some concerted effort, keep cross-school teacher sorting in check through policies like centralized assignment of teachers to schools' (p. 571). The findings from Study III, however, suggest that centralized allocation may not remedy all types of socioeconomic sorting. Centralized staffing may be a better tool in developing educational systems, as shown in past work (Hanushek et al., 2013) and the fact that sorting by specialization was most prevalent in higher HDI countries may explain the inconsistent findings. However, this may be a fertile line of inquiry for future research.

The most consistent finding throughout the study was the statistically significant and positive relationship between the slope and HDI. This was true

regardless of the subset of teachers included in the model and was statistically significant at or above the 95% level in both cases. One possible interpretation of this finding is that higher income countries have more job opportunities for individuals skilled in mathematics and science, and these individuals are therefore less likely to tolerate more challenging working conditions in socioeconomically segregated settings. It could also be that teacher perception of socioeconomic status varies across educational systems. At present these are only speculative interpretations, and future research will also need to investigate this issue more closely.

Taken together, the results from Study III paint a complex picture regarding the correlates of socioeconomic inequalities in teacher sorting. While slope variances for cross-classroom socioeconomic teacher sorting varied significantly across the countries, the institutional features of tracking and ability grouping, school autonomy, school accountability, and school competition showed differential effects in almost every case.

Student performance-based teacher appraisal and teacher turnover intentions (Study IV)

While the first three studies in the thesis examine teacher sorting via the distribution of qualifications, the last study shifts the focus to examine teacher sorting via the attrition of teachers in certain schools. The study uses TALIS 2018 data across 46 education systems and investigates whether the usage of student performance data for formal teacher appraisal moderates the relationship between teacher turnover intentions and classroom socioeconomic composition. The study employs three-level random slope models, where the relationship between turnover intentions and classroom socioeconomic composition (the random slope) is regressed on school system-level accountability. To test the robustness of the findings, a series of within-country models are run. The thesis expands on the analysis and includes two-level hierarchical models with schools nested in countries (Appendix D).

In a first step, the odds of wishing to change schools as a function of classroom socioeconomic composition are estimated using within-country logistic regressions with standard errors clustered by school. Most countries display a statistically significant relationship between teacher turnover intentions and classroom socioeconomic composition. Education systems with the strongest relationships include for instance Belgium, France, UAE, Hungary, Turkey, and

China (Shanghai). There were some exceptions and education systems with no statistically significant relationship (see Georgia, Croatia, Chile and Denmark, for example).

Using hierarchical generalized models, the classroom SES – teacher turnover intention slope was then regressed on school system accountability, distinguished by who is responsible for conducting the teacher appraisals. No statistically significant effects were found for increased results-based accountability when the school management team conducts the appraisals. However, when external authorities or bodies (i.e., school inspectors, municipal or governmental bodies) conduct the appraisal, an anti-compensatory or positive relationship between school autonomy and the slope was confirmed. In this case, school systems with more results-based accountability display a stronger relationship between classroom socioeconomic composition and a teacher's likelihood to report wishing to change schools. Subsequent analyses (see Appendix D) also confirmed a positive relationship between the school socioeconomic composition – teacher turnover (the proportion of teachers wishing to change schools) slope and external accountability and a null relationship for internal accountability. The findings support past work by Jerrim and Sims (2021), Smith and Holloway (2020), and Valli and Buese (2007), whereby increases in accountability may lead to increased workload, stress, and consequently increases in teacher attrition. This study finds this to be especially pronounced in low-SES settings. The results suggest that appraisal from a school inspector or government official is experienced differently by teachers than appraisal monitoring from the school management team which often includes the principal.

The positive relationship between the slope and accountability persists despite a host of teacher characteristics under control, including job satisfaction, self-efficacy, gender, and others, as well as school characteristics such as location, type, and SES composition. It lends support to the initial hypothesis of the paper—with some cautions in mind. First, the within-country models show that the moderating effect of system-level accountability may be cancelled out due to positive and negative effects across countries. There is a lot of country heterogeneity with regards to how appraisal by the school management team interacts with turnover intentions of teachers. There is also the issue of selection effects at the country-level. Countries with issues in the teacher workforce not captured by the data may rely more heavily on external appraisal. It was also apparent that schools in a majority of countries use performance data to assess teacher performance internally. It may be that teacher appraisal is best conducted by those with

knowledge of the school context and trust between educational actors (Qian and Walker, 2019). The findings corroborate the growing number of studies demonstrating the potential pitfalls of performance-driven accountability in socioeconomically disadvantaged settings (Smith and Holloway, 2020; Valli and Buese, 2007).

The results are then re-run by various subgroups. First, by school consequences of appraisal and then by teacher characteristics. There were no differential effects of school accountability on the slope when it comes to more punitive consequences of teacher appraisal. Teachers were no more likely to report wishing to change schools as a function of accountability if they could face dismissal, receive material sanctions or even if they could receive a salary bonus following their appraisal. This finding was surprising, as it runs contrary to the notion that more punitive working environments and stringent accountability may be associated with higher turnover rates. While more research will have to confirm this interpretation, it may be that schools less likely to fire teachers also struggle more with working conditions, which would explain the findings. By and large, however, no cross-national pattern was found with respect to the stakes or consequences of teacher appraisal. The subgroup results by teacher characteristics yielded more interesting results. A similar anti-compensatory relationship to the slope was found for both male and female teachers but was slightly larger in male teachers. Most importantly, contrary to past research (Fantilli and Macdougall, 2009; Hanushek and Rivkin, 2010), results show that experienced teachers (those with more than five years of experience) were more likely to exhibit SES-based turnover in school systems with increased accountability as opposed to newer teachers. This may have negative implications for students over and above their test scores, as schools with a higher proportion of experienced teachers may have better academic climates (OECD, 2018c).

The findings in Studies III and IV suggest yet another a complex picture which is summarized in Table 4. Plus-signs depict statistically significant positive effects. On one hand, the results of Study IV depict clear and consistent evidence that at least in the case of external accountability teachers are more likely to report wishing to change schools if they teach classrooms or schools with more socioeconomically disadvantaged students. On the other, school accountability was not significantly related to teacher sorting in a majority of the models in Study III. School competition, autonomy and tracking were in certain cases positively related to the teacher qualification-classroom SES slope. Importantly, no institutional variables were negatively associated with either of the random slopes.

The most consistently predictive institution-level variable was HDI. Policymakers should weigh these mixed findings in the context of their national or regional educational systems.

Table 4 Summary of findings across Study III and Study IV

	Cross-class		**Cross-class M/S**		**Cross-school**		**Turn. class**	**Turn. school**
	Spec.	Nov.	Spec.	Nov.	Spec.	Nov.	All	All
HDI	+		+		+	+	+	+
Track	+		+					
Abil.								
Comp.		+			+	+		
Aut.		+						
Accnt.	+						+	+

Note. Cross-class refers to cross-classroom sorting, M/S refers to sample of math and science teachers, cross-school refers to cross-school sorting, and turn. Refers to turnover intentions across classrooms and across schools. + refers to a statistically significant positive effect.

Emergent themes across the empirical studies

The empirical studies investigated aspects related to the phenomenon of teacher sorting from a cross-national perspective. In addition to the substantive results answering the initial research questions, several themes materialized which are applicable to each of the studies.

1. Measurement matters
2. Implications for policy
3. The role of socioeconomic context

Measurement matters

Throughout the planning stages for each individual constituent study, the measurement of teacher quality, competence and sorting required deliberate decisions which were based on challenges related to cross-national comparability, validity, and relevance to the research topic in question, to say nothing of the availability and quality of the data.

Two main indicators of teacher quality were the focus of the thesis: experience and specialization. Specifically, teachers with less than five years of experience and no specialization in mathematics were the focus of Studies I through III. Some clear patterns emerged across the studies which are tied to this consistency in measurement. Sorting by specialization was found to be more prevalent in more economically developed education systems, especially across classrooms, whereas

sorting by experience in most cases either displayed a null relationship with economic development or a negative one. The measurement of teacher quality can also help to interpret similarities and differences in terms of contextualizing the findings within the literature on teacher sorting. For instance, the most recent study investigating teacher sorting using TALIS 2013 data by Luschei and Jeong (2019) positions Sweden as showing a high degree of cross- and within- school sorting. In contrast, most investigations in this thesis show Sweden as having a fairly equitable distribution of teachers. As mentioned throughout the thesis, Luschei and Jeong (2019) create a composite measure of teacher quality comprised of self-efficacy, experience, and educational attainment. This is not to say that the findings from their study should be disregarded. Other studies (Hansson and Gustafsson, 2016) have also noted increasing inequity in sorting in Sweden (although the point estimates are similar to this thesis, about 5 percentage points) with respect to the proportion of certified teachers across schools. The important take-away is that conclusions about the degree of inequality in teacher sorting are highly measurement dependent.

Measurement choices related to socioeconomic status should also be emphasized. TIMSS does not test students on their parent's income or occupation, and therefore the socioeconomic status construct created in Studies I and II reflects a dimension of SES probably more in line with cultural and social capital and to a lesser extent economic capital (Bourdieu, 1986). Slight differences in the results may arise should other studies investigate a similar research question with income and occupation in the mix, but they would not be expected to be too large (Chmielewski, 2019). Differences in the measurement of socioeconomic context across TIMSS and TALIS may also underlie some inconsistencies in the results. For example, Chile displayed the highest level of inequity in teacher sorting by specialization in Studies I and II but wasn't near the top of the list in Study III. Chile also does not display a statistically significant result for the proportion of unspecialized teachers and school SES (Appendix D). It could be that the extent of sorting in Chile extends primarily to mathematics teachers. The sampling of the two studies may also be at play. TIMSS is representative of grade 8 students in school, while TALIS is representative of teachers. More research will need to investigate this inconsistency in more detail.

Studies III and IV also reflect deliberate measurement choices regarding educational policies. Study IV differentiates between the use of student results in teacher appraisal when conducted by external authorities versus the school management team. This distinction matters a great deal, as it influences not only

the level of variation in the accountability measure but also the relationship to the slope. There are also theoretical distinctions between the approaches which have been neglected in past cross-national work. Last, Studies III and IV illuminate differences in sorting patterns when the unit of analysis changes from cross-student group, to cross-classroom, to cross-school. The country rankings in terms of the magnitude of sorting are therefore likely to differ depending on how sorting is conceptualized.

Implications for policy

Given the importance of measurement in the contextualization of the findings, what are the implications of the findings for educational policymaking around the world? First, countries that do not display inequalities in teacher sorting in any of the studies in the thesis may have inequities in other dimensions, and therefore should not take the absence of a finding as synonymous with an equitable distribution of teachers. In relation to this, William (2018) puts forth several guidelines around the evaluation of educational policies that can be summed up in his keynote phrase 'everything works somewhere and nothing works everywhere' (2006; 2018). William (2018) urges educational researchers and policymakers to consider the magnitude of policy effects (i.e., not whether something works but how well it works), as well as the cost-benefit tradeoff associated with educational policies. The following sections will discuss the key policy implications of the thesis with these guidelines in mind.

Teachers in low-SES settings: 'love the one you're with'

Study I in particular suggests that many of the alarm bells rung by the international testing bodies and others (OECD, 2018b) regarding the allocation of poorly qualified teachers to low-SES students require some cautions. In many education systems, even at the most extreme ends of the socioeconomic spectrum, low-SES students are just as likely to have novice or unspecialized teachers as their more socioeconomically advantaged peers when teacher and student data are linked. Study III showed a more concerning picture regarding the extent of sorting by specialization, whereby a majority of education systems displayed a strong relationship between specialization level and classroom SES. The findings suggest better quality data is needed linking student socioeconomic backgrounds to the qualifications of their teachers. While the problem of teacher sorting may not be as widespread as originally postulated, the take-away is not that it doesn't matter.

Democratizing access to teacher quality is likely one of the most important school inputs available to policymakers.

Two recommendations emerge out of this conclusion. First, a 'sufficientarian' approach (where all students have 'enough') is likely not enough to close the SES-achievement gap in terms of teacher qualifications. Study II in particular supports this conclusion and underscores the need for an equity as 'fairness' perspective, where low-SES students have teachers with credentials above and beyond what is considered 'enough.' The most challenging policy implication has to do with how educational systems are to achieve this aim. Study II showed that content knowledge likely matters for teacher competence, supporting past work (Baumert et al., 2010; Schulman, 1986). While augmenting attractiveness of the teaching profession and improving teacher education programs is important, it may be unlikely to change the competence of teachers in low-SES settings for two reasons. First, it further supports the growth perspective (Brighouse and Swift, 2006), where the logic implies that simply adding more teachers to the workforce will democratize teacher distribution. Second, changing the social status of teachers will also take a lot of time, on the magnitude of 30 years (William, 2010). It is therefore also important to consider shorter-term solutions to the problem of sorting. One shorter-term strategy is termed the 'love the one you're with' by Dylan William (2010). Policymakers should identify underqualified mathematics teachers in low-SES settings and incentivize them to participate in additional professional development courses with the explicit goal of improving their content and pedagogical content knowledge. Policymakers should be careful not to over-emphasize results in teacher appraisal, however. Such accountability mechanisms, as evidenced by Study IV, should take into account the different levels of difficulty across teaching contexts.

Implications for opportunity of choice, autonomy, and accountability

One of the main debates in educational policy relates to the decentralization of educational decision-making and the resulting checks and balances in such systems. On the one hand, there is evidence that increased opportunity of choice is tied to inequity (Strietholt et al., 2019), as it gives families and educational actors leeway to act in opportunistic ways, which often reproduces inequities already persistent in society (Bourdieu, 1977). On the other hand, proponents of increased decentralization argue that local actors may be better able to match resources with needs (Hanushek et al., 2013; OECD, 2018b). The former perspective was confirmed in Study III, as increased school competition and autonomy was

associated with higher inequity in teacher qualification sorting. This was especially true for school competition, which was positively associated with the slope in a number of cases and especially for sorting by specialization and experience across schools. This finding supported the initial hypothesis of the thesis where policies which change the composition of students are likely to change the composition of teachers (Jackson, 2009). Educational authorities frequently state that schools can mitigate social segregation by investing in the best teachers (OECD, 2022). The findings call into question this approach and suggest that targeting social segregation across schools directly is likely to be more effective.

Limiting school autonomy over staffing is consistently cited as one way to ensure a more equitable allocation of teachers (Han, 2018; Kang and Hong, 2008; Luschei and Jeong, 2019; Luschei et al., 2013) but the thesis yielded mixed results with regards to this variable. While autonomy was associated with cross-classroom novice teacher sorting, it yielded a null pattern with sorting by specialization. Given that sorting by specialization is more common in more developed countries, and that school autonomy has positive outcomes for student achievement in such countries, it could be that effect heterogeneity of autonomy is important for this finding (Hanushek et al., 2013) but more research will have to confirm this interpretation. One often touted example of successful centralized autonomy is South Korea (Luschei et al., 2013; Kang and Hong, 2008), where teachers are hired by a central authority. Despite the frequent praise for the South Korean system, the country displays a degree of inequity in novice teacher sorting in Studies III and I. This is in line with other findings (Luschei and Jeong, 2019). Meanwhile in Turkey, very high rates of novice teacher sorting occur despite the central allocation of teachers (Özoğlu, 2015) as was seen in Study I and III. A closer look at the Turkish system reveals that despite the centralized structure, such mobility patterns are in fact incentivized. The findings show therefore that centralized staffing is not likely to be successful everywhere. In contrast to the above cited arguments for centralization, the OECD (2018b; 2022) consistently advocate for decentralized hiring, arguing that autonomy is better able to mitigate the seniority-related allocation of teachers.

Given the global push towards educational accountability, Studies III and IV examined whether accountability relates to the sorting of teacher qualifications as well as to teacher turnover intentions in low-SES classrooms. Past work states that accountability incentivizes teachers to be sorted more opportunistically (Boyd et al., 2008), and puts undue stress on teachers (Jerrim & Sims, 2021; Smith and Holloway, 2020). Studies III and IV support this conclusion and provide more

evidence that the push towards performance-based monitoring is harming socioeconomically disadvantaged students and may be disincentivizing teachers (Darling-Hammond, 2004). Teachers in low-SES settings should therefore be incentivized to develop their competences (as previously argued), but this should be weighed carefully against the potential to over-monitor and manage. This is likely to be a fruitful area for future research.

Policymakers can weigh their priorities in a framework outlined by Brighouse et al. (2016) who delineate between types of values relevant for educational decisionmakers. This includes education which constitutes the capacity for economic production or personal autonomy, for example, distributive values such as adequacy or priority, and independent values such as freedom of choice, or parents' interests. For example, centralizing autonomy may prioritize the allocation of teachers at the expense of personal autonomy. This is likely to have different downstream effects depending on the education system and their cultural priorities. Nevertheless, decisions should be made transparent based on these and other values and will most likely look quite different across educational systems.

The role of the socioeconomic context

The theoretical basis of the thesis posits that educational leaders, teachers, and families alike enact inequality of educational opportunity ('the opportunity gap') in teacher competence and quality via the phenomenon of social reproduction (Schmidt et al., 1997; Bourdieu, 1977). Though there is variability across countries and measures of sorting, lower-SES students are in certain cases more likely to have teachers with lower qualification levels, or teachers who wish to change schools. Study II finds persistent evidence that social segregation is strengthening the link between student socioeconomic background and mathematics achievement and that access to teacher quality was less important in comparison. In addition to this finding, Studies III and IV showed that country HDI was a much stronger predictor of both sorting dimensions than the policy-related variables. Educational amendments which do not take into consideration the importance of the socioeconomic context at all levels are therefore likely to be much less effective.

There is a widespread and persistent global discourse which states that strengthening education is one of the most effective ways of fighting inequality (Solga, 2014) and that increased educational investments will tackle inequality (OECD, 2019b). The thesis has confirmed that democratizing access to specialized

mathematics teachers is likely to shrink the performance gap between low- and high-SES students to at least some degree. However, this may not be via improvements in performance of low-performing students, and it pales in comparison to the effects of social segregation. Moreover, ensuring an equitable distribution of teachers is just one piece of the puzzle. Another piece pertains to *how* to influence such distributions, which, as has been discussed, involves augmenting the quality of teachers in low-SES contexts, as well as incentivizing teachers to apply and remain in hard-to-staff settings. In most cases, it also involves mitigating social segregation, as evidenced by Study II and to some extent Study III. It is therefore reasonable to wonder whether educational inequity is a problem which can be solved through educational policy alone. There are critiques of the over-emphasis on education (Brown & Tannock, 2009; Crouch et al., 1999) and empirical findings suggesting that education systems which invest both in educational inputs as well as broader social equality through direct redistribution are the most effective at reducing inequality (Solga, 2014). Others find that widening social inequity offsets the equity-oriented policies (Franck and Nicaise, 2022). The long-term foci of educational policymakers should therefore also include amendments outside the realm of education.

It is however important to not minimize the potential of teachers. While they are not a cure-all, compounded over many years, ameliorations in the distribution and magnitude of teacher competence are likely to accrue tangible benefits for socioeconomically disadvantaged students.

Chapter 7 Concluding remarks

The purpose of this thesis has been to acquire a better empirical understanding of the phenomenon of teacher sorting from a cross-national perspective. The main research questions focused on the magnitude and development of teacher sorting, whether such sorting exacerbates achievement inequity, and policy-related correlates of different dimensions of teacher sorting. The following sections outline the contribution of the thesis to existing knowledge on educational inequity and teacher competence, as well as policy investigations using large-scale assessment data.

Contribution of the thesis

This project contributes to the existing body of literature in several ways. First, each constituent study demonstrates different cross-national patterns of sorting inequity which vary depending on the qualification in focus. Past studies have not considered that the nature of sorting patterns (for example, across specialization and experience) likely reflect specific qualitative differences in working and staffing conditions. This underscores the fact that mechanisms related to sorting based on experience and specialization are not interchangeable. There is a moderately consistent overlap between sorting by specialization and more developed economic systems and sorting by experience in less developed economic systems. Policymakers should therefore consider the differential functions of centralization, autonomy, choice and accountability across different development levels as outlined in previous work by Hanushek et al. (2013). Increased opportunity of choice resulted in more pronounced sorting especially in the case of school competition. More stringent accountability practices, particularly as they relate to monitoring teacher effectiveness using student test scores, have an anti-compensatory relationship to sorting by specialization and turnover intentions. These findings present some of the first international evidence on the institutional determinants of teacher sorting and retention.

The dissertation also adds to the scarce body of cross-national evidence using causal and quasi-experimental approaches to estimate the impact of teacher quality

on student outcomes. Few cross-national studies to date have used such an approach to estimate the relationship between test scores and teacher qualifications (e.g., Sancassani, 2021), and the results of the second study in this dissertation generally lined up with these findings. Specifically, there was a consistent pattern of null findings related to the impact of novice teacher exposure and sorting on student achievement and achievement inequity. Conversely, there was a consistent pattern of significant findings related to the impact of exposure to unspecialized mathematics teachers and sorting by specialization on both student achievement and achievement inequity. These results add to the conflicting body of literature on the importance of observable teacher characteristics and present a novel approach considering the effects of how educational systems are able to allocate teacher in terms of qualifications.

A final contribution of the thesis relates to the extent and development of teacher sorting globally and across time. While democratizing access to specialized teachers is likely to diminish achievement inequity in some capacity, the extent of inequity in sorting cross-nationally varies across educational systems. Moreover, it is only increasing in a very few select instances. As such, other non-educational policy inputs related to income inequality and social segregation should not be underestimated. This was underscored by the panel regression estimates depicting the effect of social segregation on mathematics achievement inequity.

Limitations and methodological challenges

Several limitations and methodological challenges characterized the analyses and intepretations of the thesis. First and foremost, only a narrow dimension of teacher quality was in focus due to endogeneity problems as well as problems related to data availability and cross-national comparison. There is a vast body of work examining teaching practices (Kyriakides et al., 2013) which is of potential relevance to the problem of sorting. If the teaching context influences teaching practices (as it is likely to), there is an unexplored question relating to how well teacher education programs across educational systems prepare teachers to work across different settings. Thus far, many studies on sorting have not been able to overcome the endogeneity problem (Jackson et al., 2014). While the thesis has contributed to the importance of the specialization and experience qualifications, more research should investigate differential levels of classroom management, cognitive activation, and other processes across socioeconomic groups within education systems. The thesis also only examines students in grade 8 and their

teachers. This limits the generalizability of the findings to other populations and replication studies are needed.

Another limitation relates to the nature of international large-scale assessment data and the issues related to measurement and validity. While equipped with a rigorous quality-assurance process, there are unanswered questions related to the nature of included participants over time, especially in the context of teachers. Due to the sampling structure of TIMSS which assures the representative nature of students, systematic unobservable changes in the way teachers are sampled within schools over time may bias the findings. Recent work has called into question the validity of certain ILSA sampling features (Anders et al., 2021). Such validity threats also exist at the institutional level. Particularly relevant is the Turkish example, which scores very low on the staffing autonomy measure but which also prioritizes teacher preferences in their allocation of teachers. Indeed, the degree of centralization likely matters (Seebruck, 2021) and it is difficult to capture these nuances with the data available by the OECD TALIS study. However, the thesis shines a light on areas which are worthy of much more future research and attention with more granular national datasets.

While Study III uses a longitudinal panel approach, Studies III and IV are cross-sectional and the findings represent associations between variables. Related to this is the potential for selection effects at the system-level which are uncorrelated with the control variable HDI. Last, while international large-scale assessments present unique opportunities for cross-national research, certain groups of countries do not tend to participate and are underrepresented (Kamens & McNeely, 2010).

Future research

In light of the findings and limitations, several avenues for future research may be suggested. First, there is a need to replicate the findings across other indicators of teacher quality as well as population subgroups. This may be seen as an extension based on the approach of Study I. These new gap measures could then be linked to student outcomes across the subgroups using the panel data approach. Where repeated country-level observations exist, there is also the potential for longitudinal analysis of the institutional correlates included in the thesis in Studies III and IV, including school autonomy, accountability, competition, and stratification. This may provide more insight into, for example, the autonomy over staffing variable.

There are unanswered questions regarding sorting across national levels of economic development. In most cases when teacher and student data were linked, lower- and middle-income countries displayed large sorting patterns. However, the opposite was found when the data considered the perception of the principal and teacher in terms of the school and class socioeconomic composition. National research which links student and teacher data may provide further insight into this problem. Furthermore, why are some education systems which struggle with inequity in student outcomes able to ensure a highly equitable distribution of teachers, while others struggle despite a higher level of overall equity? This applies to Hungary in the former case, and Chinese Taipei, for example, in the latter.

Last, Studies III and IV highlighted in particular that accountability is likely to have negative implications for socioeconomic teacher sorting as well as teacher turnover intentions. Future research can examine whether appraisal which is conducted by peer teachers or on other aspects (not student achievement data) in 'softer' accountability systems (Maroy & Voisin, 2017) has a compensatory relationship to teacher sorting and turnover intentions.

Swedish summary

Det socioekonomiska prestationsgapet inom utbildningsområdet har bestått globalt trots årtionden av forskning och politiska åtgärder (Broer et al., 2019; Chmielewski, 2019). Detta har negativa konsekvenser för såväl individer som nationer, eftersom framgångsrik skolgång är förknippad med social rörlighet och ekonomisk tillväxt (Hanushek & Woessman, 2015). Komparativ forskning har visat att en stor del av variationen i elevernas akademiska prestationer kan förklaras av deras socioekonomiska bakgrund (Sirin, 2005; Harwell et al., 2017). En förklaring till varför prestationsgapet har bestått är att forskare har misslyckats med att rikta in sig på dess viktigaste orsaker (Sims & Allen, 2018; OECD, 2018b); nämligen lärarkompetens och undervisningskvalitet. Tidigare forskning har visat att lärare är den viktigaste insatsen på skolnivå (Goe, 2007; Nye et al., 2004; Rivkin et al., 2005; Wayne & Youngs, 2003) och det hävdas allmänt att den ojämlika fördelningen av lärarkompetens kan bidra till orättvisa utbildningsmöjligheter och till prestationsgapet mellan elever i olika socioekonomiska grupper. Den tvärnationella forskningen om detta ämne är dock knapphändig. Avhandlingen tar upp denna fråga och undersöker tre aspekter relaterade till lärarfördelning mellan elever med olika socioekonomisk bakgrund. För det första dess omfattning och utveckling över länder och tid, för det andra dess effekt på ojämlika studieresultat och för det tredje institutionella och policy faktorer.

Bakgrund

Sedan åtminstone 1948 har internationella organ förklarat utbildning som en "mänsklig rättighet" (FN, 1948). Denna rörelse mot demokratisering av utbildning skedde för mer liberala välfärdsstatliga system såväl som för mer traditionellt socialdemokratiska system, som de nordiska länderna (Arnesen & Lundahl, 2006). Den är mindre uppenbar i system som är explicit meritokratiska (Chmielewski, 2019). Ändå verkar det finnas en rörelse mot att fokusera mer på rättvisa i utbildning globalt (OECD, 2018b). Samtidigt har kritik om skolans roll för att främja social jämlikhet ackumulerats (Coleman, 1966; Broer et al., 2016; Strello et al., 2021). Coleman-rapporten (1966) visade att efter att hänsyn tagits till hemmets

resurser och elevernas socioekonomiska bakgrund stod faktorer på skolnivå för en låg andel av variationen i elevernas prestationer. Sedan dess har försök gjorts att förstärka utbildnings- och skoleffektiviteten, på grundval av metodologisk och teoretisk utveckling som skett under de senaste decennierna (Creemers, 2005; Kyriakides et al., 2018; Scheerens & Bosker, 1997). Vissa har kommit till slutsatsen att skolor och skolpolitik inte är förkämpar för jämlikhet, utan snarare vidmakthåller samhälleliga ojämlikheter (Bourdieu, 1977; Darling-Hammond, 2004). Forskning om utbildningseffektivitet har utvecklats till att inkludera frågor om jämlikhet, som anger sambandet mellan socioekonomisk bakgrund och elevernas akademiska prestationer (Kyriakides et al., 2018). Internationella storskaliga studier som TIMSS, PISA och TALIS har utvecklats för att återspegla variationen i resultat.

En pedagogisk resurs som ofta sägs vara systematiskt ojämnt fördelad mellan eleverna är lärarnas kompetens och kvalifikationer. OECD (2018b) rapporterar att brist på kvalificerade lärare i de flesta länder hindrar elevernas läranderesultat. Akiba et al. (2007) visade att i många länder har elever med låg SES systematiskt mindre tillgång till kvalificerade lärare. En förklaring till den ojämna fördelningen av kvalificerade lärare över skolor har varit preferenser för bättre arbetsvillkor framför lön i kombination med rekryteringssvårigheter till socioekonomiskt missgynnade skolor (Bacolod, 2007; Jackson, 2009; Sims & Allen, 2018). Att undervisa elever från lägre socioekonomisk bakgrund kan vara mer utmanande på grund av lägre nivåer av stöd hemma, utmanande elever, förändringar i skol- och klassrumsklimatet, dåligt arbetsklimat, dåliga skolresurser och ökad övervakning (Allensworth, Ponisciak & Mazzea, 2009; Lazear, 2001; Sacerdote, 2011). Även om man kan dra slutsatsen att lärare motiveras av lägre utbildningsresurser snarare än av elevernas socioekonomiska sammansättning, utnyttjar Jackson (2009) i en undersökning avskaffandet av s.k. ”bussning” i ett skoldistrikt i North Carolina och finner att lärarna återspeglar elevernas demografi. På så sätt kommer policyer som förändrar elevgruppens sammansättning sannolikt också att förändra lärargruppens sammansättning. Allen et al. (2018) finner att socioekonomiskt missgynnade skolor också har systematiskt fler lediga jobb. Lärare tenderar att själva välja vissa arbetsförhållanden när de har självständighet och kontroll över sin arbetssituation. I andra utbildningssystem fördelas lärarna centralt och det är mindre enkelt att redovisa ojämlikheter i lärarfördelningen. Skolledningar kan dock ge lärare i uppdrag att öka elevernas prestationer och förstärka en skolas marknadsposition i vissa fall, beroende på om det finns ansvarsutkrävande

mekanismer eller inte (Boyd et al., 2008). Det finns ett behov av att fastställa vilka andra policyer som samverkar med socioekonomisk lärarsortering.

Stödet för observerbara indikatorer på lärarkvalitet är blandat (Goe, 2007; Hanushek & Rivkin, 2012), men det finns viss konsensus kring särskilt några kvalifikationer. Skolledare tenderar också att anställa lärare utifrån deras kvalifikationer (Engel och Finch, 2014). I synnerhet finns det stöd för att lärares erfarenhet spelar roll under de första åren (Goe, 2007; Rice, 2003; Rockoff, 2004). Det finns också stöd för ämnesspecialisering, särskilt inom matematik (Hill et al., 2005; Sancassani, 2021). Även om undervisningsmetoder är viktiga (Kyriakides et al., 2013), är det svårt att jämföra dessa mellan grupper av elever eftersom undervisningen inte är konstant över skolor och klassrum (Thrupp, 1999; Lazear, 2001). Av dessa skäl fokuserar avhandlingen på socioekonomisk lärarsortering utifrån erfarenhet och ämnesspecialisering.

Syfte

Syftet med avhandlingen är att undersöka problemet med socioekonomiskt relaterad lärarsortering ur ett internationellt perspektiv. Avhandlingen fokuserar på problemet med lärarfördelning, liksom studieresultat och institutionella drag och policyer. I artiklarna granskas olika aspekter av lärarfördelning. I Studie I undersöks mönster i lärarkvalifikationsklyftor inom 32 utbildningssystem, liksom deras förändringar under 20 år. Studie II tar utgångspunkt i måtten på lärarkvalifikationsgapet från Studie I och kopplar dem till förändringar i socioekonomiska skillnader i matematikprestationer över 20 år. Studie III undersöker om sambandet mellan klassrummets socioekonomiska sammansättning och lärarkvalifikationerna varierar som en funktion av skolans stratifiering, skolans autonomi, skolkonkurrens och skolansvar på systemnivå. Studie IV undersöker om lärare i klasser med lägre socioekonomisk nivå rapporterar att de vill byta skola i större utsträckning än i skolsystem med högre prestationsnivåer. Studierna visar att ojämlikheten i lärarfördelningen varierar beroende på hur lärarkvalitet och elevgrupper mäts. Det fanns dock stöd för att ojämlik fördelning av lärare efter ämnesspecialisering resulterar i ojämlika matematikprestationer. Även om det fanns visst stöd för att policy på systemnivå hängde samman med ojämlik lärarfördelning, var socioekonomiska sammanhang en mycket mer konsekvent prediktor för både ojämlikhet i prestationer och lärarsortering.

Teoretiskt Ramverk

Eftersom avhandlingen behandlar lärarsorteringens mångfacetterade karaktär, inkorporeras flera teorier och modeller. En utgångspunkt tas i modellen för utbildningserfarenheter och modellen "möjlighet att lära" (Schmidt et al., 1997) som ett övergripande ramverk för jämförelse av ojämlikhet mellan utbildningssystem, inklusive konceptualisering och mätning av socioekonomisk status. Därefter är Bourdieus teori om social reproduktion och habitus särskilt relevant för hur ojämlikheter produceras av pedagogiska aktörer. Sist övervägs olika teorier om fördelningen av utbildningsresurser.

TIMSS-modellen för potentiella utbildningserfarenheter utökas från tidigare modeller för "möjlighet att lära", där tid var huvuddimensionen av utbildningsmöjligheter (Carroll, 1963; Schmidt et al., 1997). Modellen skiljer mellan de avsedda, implementerade och uppnådda nivåerna. Modellen för potentiella erfarenheter introducerar en mängd faktorer på både avsedd och implementerad nivå som begränsar elevernas utbildningserfarenheter på olika sätt, såsom lärare och deras kvalifikationer. Avhandlingen behandlar hur faktorer på olika nivåer fördelas mellan studenter med olika socioekonomisk bakgrund ('möjlighetsgapet'). Modellen visar också att det kan finnas en bristande överensstämmelse mellan de avsedda nivåerna i läroplanen och den uppnådda läroplanen.

Å ena sidan kan denna bristande överensstämmelse uppstå som ett resultat av att policyer försummar rättvisedimensionen (Lee & Wong, 2003), men det kan också uppstå som ett resultat av medvetna val från utbildningsaktörernas sida. Det senare kan sammanfattas av Bourdieus (1977) teori om social reproduktion. Via begreppet habitus beskriver Bourdieu hur institutionella strukturer som klassrelationer kan ha fysiska förkroppsliganden. Habitus fungerar därför som en länk mellan ’social orsak och social effekt’ (Nash, 1990, s. 434). Bourdieu placerar skolan som den primära plats där habitus genereras. Huvudargumentet är att skolgång gynnar ett litet antal individer som är redo för lärande och skolgång och försummar de återstående eleverna. Elever som uppvisar denna beredskap har ofta medel- eller överklassbakgrund. De former av kapital som familjerna ger sådana elever kan vara ekonomiska, i form av monetära investeringar i deras utbildning, såsom extra lektioner, privat skolgång eller skolgång i socialt segregerade överklassområden, kulturella i form av exponering för vissa typer av kunskap (d.v.s. konst, tal och manér), och sociala i form av kontakter och nätverk.

Bourdieu (1977) hävdar att skolsystemet styrs av de dominerande kulturklasserna och att andra elevers habitus ses som ett underskott. Det finns därför en distinktion att göra med avseende på elevresultat som genereras av skillnader i klass och kulturell habitus och de som genereras av nackdelar för vissa elever. Detta är ingen lätt uppgift, eftersom det upprepade gånger har visat sig att elever från arbetarklassens socioekonomiska bakgrund har olika preferenser för utbildningsresultat och mål (Willis, 1977). I allmänhet återspeglas dock habitus genom beslutsfattandet som är involverat i politik och styrning av mänskliga resurser. Varför är vissa skolor "önskvärda" som arbetsplatser och andra inte? Vilka elever bedöms som ”mer sannolikt” framgångsrika från skolledarnas sida, och hur påverkar detta hur skolresurserna fördelas? Hur samverkar olika utbildningsreformer (som ökad lärarbevakning) med dessa processer? Slutligen återspeglar utformningen av utbildningssystem också habitus, eftersom de är individuella åsikter om social välfärd och individualism som gjorts kollektiva (Lavrijsen & Nicaise, 2016). Detta dialektiska förhållande mellan strukturer och individer är ett centralt tema i Bourdieus teori om social reproduktion.

En sista men lika viktig teoretisk grund för avhandlingen kommer från diskussioner om hur man effektivt och etiskt fördelar utbildningsresurser (Brighouse & Swift, 2006). Enkelt uttryckt behandlaravhandlingen tre huvudsakliga tillvägagångssätt: det "prioriterade" tillvägagångssättet, där resurser fördelas ojämnt till förmån för elever med lägre socioekonomisk status; tillvägagångssättet med "tillräcklighet", där en lägsta tröskel krävs och resurser fördelas över alla grupper; samt en strikt jämlikhetsstrategi, där strävan är att alla resurser ska fördelas lika.

Metod

Analyserna som utförs i avhandlingen baseras på data från två internationella storskaliga studier. Den första är ”Trends in International Mathematics and Science Study” (TIMSS) som genomförs av IEA, och som omfattar sex undersökningsomgångar över 20 år, från 1999 fram till, i skrivande stund, år 2019. För syftet med denna avhandling används endast data från årskurs åtta, med fokus på elever och matematiklärare. Den andra datakällan är den senaste (i skrivande stund) omgången av ”Teaching and Learning International Survey” (TALIS) som genomfördes av OECD 2018 med fokus på lärare för "population 2". TIMSS genomförs vart fjärde år, medan TALIS genomförs vart femte år. Var och en av datamängderna har fördelar relaterade till deras specifika syften. Ett centralt inslag

i de internationella studierna är bakgrundsenkäten som kartlägger relevant lärar- och skolinformation, samt information om elevers socioekonomiska bakgrund och egenskaper. Ett växande antal länder deltar i de internationella studierna, med det högsta antalet i Nordamerika och Europa och det minsta antalet deltagande länder i Afrika (Kamens & McNeely, 2010). På grund av urvalsstrukturerna för både TIMSS och TALIS är uppgifterna representativa på landnivå, när adekvata andelar har uppnåtts. Både TIMSS och TALIS har en nästlad struktur, med elever/lärare på nivå 1, nästlade inom skolor (nivå 2), som är nästlade inom utbildningssystem (nivå 3).

TIMSS använder en tvåstegs stratifierad urvalsdesign, som samplar skolor enligt tidigare fastställda strata proportionellt mot deras storlek samt hela klasser inom skolorna för att täcka en rad nationellt representativa utbildningssammanhang. TIMSS har minimikrav för deltagande för att ett land ska inkluderas. Dessa kräver att minst 150 skolor ska delta per årskurs och minst 4000 elever (Mullis et al., 2016). Det kan också finnas mer än en lärare per elev, vilket leder till att man måste använda lämpliga lärarvikter. TIMSS tillhandahåller enkätvikter för elever, lärare, skolor och länder, som står för studiens urvalsstruktur (Mullis et al., 2016). Enkätvikter för elever och skolor är omvänt proportionella mot deras sannolikhet att bli utvalda, medan vikter för länder står för de olika urvalsstorlekarna mellan deltagande utbildningssystem.

OECD TALIS-studien fokuserar på lärarresultat. TALIS samlar in kontextuella data från lärare och rektorer, relaterade till egenskaper hos deras klassrum och skolor, arbetsmiljön, arbetstillfredsställelse, personliga egenskaper, kvalifikationer och undervisningsprocesser. För denna avhandling ingår endast lärare från ISCED nivå 2 i analyserna. TALIS använder en liknande tvåstegs stratifierad urvalsprocedur som TIMSS. Inom varje utbildningssystem dras 200 skolor med en sannolikhet proportionell mot deras storlek. Inom var och en av dessa skolor väljs 20 lärare slumpmässigt ut.

Variabler

Studie I använde variabler från 6 TIMSS-cykler (1999–2019). Dessa var i synnerhet den högsta nivån av föräldrautbildning från elevernas mor och far, och antalet böcker i deras hem, i ett sammansatt mått på socioekonomisk status. Även lärarvariabler som antal undervisningsår och ämnesfördjupning användes. Kontrollvariabler inkluderade elevegenskaper som kön, utrikesfödd, språk och skolresurser, plats (land/stad) och typ (privat/offentlig). Studie II använde de mått

som konstruerats för Studie I och utökade analysen med testresultat från 1999 - 2019, där s.k. plausibla värden inkluderades som beroendevariabler i analysen. Studie III använde TALIS 2018-data, inklusive lärarens uppfattning om socioekonomisk sammansättning, lärarerfarenhet och ämnesspecialisering, såväl som variabler på systemnivå, som skolans autonomi, ansvarighet, konkurrens och stratifiering. Studie IV använde variabler för intention att flytta till en annan skola från TALIS 2018 tillsammans med skolansvar och klassrummets socioekonomiska sammansättning. Studie II kontrollerar social segregation mellan skolor, och Studie III och Studie IV använder HDI för att kontrollera nationella nivåer av ekonomisk utveckling.

Analysmetod

Avhandlingen bygger på olika statistiska metoder beroende på forskningsfråga och syfte. Studie I använder sig av konfirmatorisk faktoranalys för att fastställa faktorpoäng avseende socioekonomi inom landet och inom cykeln. Studie I använder också deskriptiva statistiska metoder för att undersöka gapets storlek och trender.

Frustrationer kring begränsningarna av korrelationsanalyser för policyrelaterad forskning har lett till en ökning av antalet kvasi-experimentella och kausala angreppssätt inom pedagogisk effektivitetsforskning (Strietholt et al., 2014). Ett sådant tillvägagångssätt är den så kallade ”fixed effects”-metoden med paneldata. Den grundläggande idén med denna metod är att använda de ingående enheterna (t.ex. länderna) som sina egna kontroller och fokusera analysen på förändringar inom enheterna i både de oberoende och beroende variablerna. Detta kan uppnås med hjälp av den longitudinella karaktären hos storskaliga data på landnivå, då dessa ger tillgång till data från upprepade mätningar av samma enheter (s.k., fixa enhetseffekter) (Gustafsson, 2013). Genom att begränsa analysen till förändringar inom enheterna kontrollerar analysen för alla tidsinvarianta faktorer. Studie II använder sig av detta tillvägagångssätt och undersöker effekten av socioekonomisk lärarsortering inom landet över tid på ojämlikhet i elevernas prestationer.

Flernivåmodeller tillåter forskare att hantera den nästlade strukturen av utbildningsdata. Flernivåmodeller används för att undersöka sambandet mellan faktorer på lands-/systemnivå och lärare/klassrumsnivå med hjälp av en kumulativ modellbyggnadsprocess enligt rekommendationer från Bryk och Raudenbush (2002) och Sommet och Morselli (2017). Alla flernivåmodeller i avhandlingen använder s.k. slumpmässiga interceptmodeller, där interceptet för utfallsvariabeln

varierar mellan skolor och länder. Flernivåmodeller tillåter också studier av interaktioner mellan nivåer och slumpmässiga lutningar, d.v.s. antagandet om heterogenitet hos nivå 1-effekten som en funktion av nivå 2- eller 3-effekten kan således testas. Studierna III och IV använder hierarkiska generaliserade linjära modeller med slumpmässig lutning som relateras till systemnivåegenskaper. I studie III fångar den slumpmässiga lutningen sambandet mellan klassrummets socioekonomiska sammansättning och lärarkvalifikationerna, och i Studie IV relationen mellan klassrummets socioekonomiska sammansättning och lärarnas intention att flytta.

Flernivåmodellering kan också undersöka uppdelningen av varians i en variabel över nivåer och är ett användbart sätt att konstruera mått på skillnader mellan och inom enheter. Intraklasskorrelationskoefficienten (ICC) kan användas för att undersöka graden av varians i socioekonomisk status för elever som kan förklaras av skolnivån, som Studie II. Ett högt värde (närmare 1) skulle indikera en hög grad av likhet mellan elever inom skolor i motsats till mellan skolor. Korrelationskoefficienten inom klassen i den beroende variabeln av intresse kan också inkluderas för att avgöra om ett ramverk på flera nivåer behövs (LeBreton & Senter, 2008). En låg ICC för andelen varians på landnivå som förklaras i den beroende variabeln skulle indikera att variabeln av intresse varierar mer inom länder än mellan länder.

Resultat

Studie I

Glassow, L.N., and Jerrim, J. (2022). Is inequitable teacher sorting on the rise? Cross-national evidence from 20 years of TIMSS. *Large-scale Assessments in Education, 10,* doi: 10.1186/s40536-022-00125-9

Ojämlik tillgång till kvalificerade lärare för barn med olika socioekonomisk status—även känd som ojämlik lärarsortering – har framställts som en potentiell faktor som bidrar till det socioekonomiska prestationsgapet. Trots detta har få studier undersökt gränsöverskridande skillnader i lärarsortering, och ingen har undersökt det inom länder över tid. Internationella storskaliga studier inom utbildningsområdet är unikt positionerade för att svara på sådana frågor på grund av deras longitudinella karaktär på systemnivå. Denna studie använder sex undersökningsomgångar av data från Trends in International Mathematics and Science Study (TIMSS) från 1999 till 2019 för 32 utbildningssystem. Vi jämför

matematiklärares kvalifikationer i årskurs 8 för varje land vid varje tidpunkt, mellan topp- och bottengrupper avseende elevernas socioekonomiska bakgrund. Resultaten visar på det hela taget att många länder uppvisar försumbara klyftor i tillgång till lärarkvalitet, med några viktiga undantag. Vad gäller ojämlikhet hos nybörjare är problemet mest utbrett inom utbildningssystem för låg- och medelinkomstländer (dvs i Turkiet, Marocko, Tunisien och Indonesien). Ojämlikhet i sortering baserad på matematikutbildning är mindre vanlig, utan något tydligt mönster när det gäller nivå av ekonomisk utveckling (dvs i Chile, Australien, Nya Zeeland och kinesiska Taipei). Den socioekonomiska ojämlikheten i lärarsortering har också varit i stort sett stabil över tid. På grundval av erfarenhet och matematikutbildning visar mindre än en handfull system systematiska uppåtgående trender i ojämlikhet avseende lärarsortering (dvs i Chile, Marocko, Singapore, och Nya Zeeland). Med tanke på ett ökande fokus på ojämlikhet i tillgången till lärare har dessa resultat ekonomiska och policymässiga implikationer för att ta itu med problemet med det socioekonomiska prestationsgapet.

Studie II

Glassow, L.N., Yang Hansen, K., and Gustafsson, J.E. (under revision). Does socioeconomic sorting of teacher qualifications exacerbate mathematics achievement inequity? Panel data estimates from 20 years of TIMSS.

Nya och äldre studier har rapporterat antingen ett kvarstående eller ett vidgat socioekonomiskt prestationsgap, dvs. skillnaden i prestation mellan elever i de högsta och lägsta socioekonomiska grupperna. Med hjälp av paneldatateknik med fixa landeffekter för 32 utbildningssystem och sex omgångar av data från Trends in International Mathematics and Science Study (TIMSS), undersöker vi om sorteringen av lärare efter specialiseringsnivå i matematikundervisning och nybörjarstatus mellan elever med olika socioekonomisk bakgrund förvärrar ojämlikheten i matematikprestationer trots kontroll för socioekonomisk skolsegregation. Vi finner svagt stöd för att sortering efter matematikutbildning är förknippad med prestationsojämlikhet, och inget stöd för vikten av sortering baserad på lärares erfarenhet. Socioekonomisk skolsegregation förvärrar dock tydligt och effektivt prestationsojämlikheten. Resultaten har policy konsekvenser för en effektiv fördelning av utbildningsresurser.

Studie III
Glassow, L.N., Franck, E., and Yang Hansen, K. (under revision). Institutional characteristics moderating the relationship between classroom socioeconomic composition and teacher qualifications: Evidence from 46 education systems in TALIS 2018

Denna studie undersöker omfattningen av ojämlik sortering av lärare över utbildningssystem liksom relevanta institutionella korrelat. Vi använder data från OECD TALIS-studien från 2018 med totalt deltagande av 144 316 lärare och 9 063 skolor i 46 utbildningssystem. Vi skattar först sambandet mellan klassrummets socioekonomiska sammansättning och lärarkvalifikationer inom varje deltagande utbildningssystem. Därefter undersöker vi med hjälp av en hierarkisk generaliserad linjär modell på tre nivåer om förhållandet mellan klassrummets socioekonomiska sammansättning och lärarkvalifikationer varierar som en funktion av institutionella särdrag via en slumpvarierad lutning, med regression på stratifiering på skolsystemnivå, ansvarighet, autonomi och konkurrens, med kontroller för nationella nivåer av ekonomisk utveckling och lärarbrist. Resultaten visar att sortering över klassrum efter specialisering är mer framträdande i ekonomiskt mer utvecklade system. Nivågruppering mellan skolor, prestationsdatabaserat skolansvar samt högre nivåer av skolautonomi över personalstyrka och skolkonkurrens var förknippade med mer uttalad socioekonomisk lärarsortering. Resultaten visar dock att institutionella bestämningsfaktorer inte lätt generaliserar över olika mätningar av lärares kompetens eller sortering mellan skolor.

Studie IV
Glassow, L.N. (under revision). Inequitable teacher turnover and performance-based appraisal: A global trend?

Elevprestationsdata används i allt högre grad för att övervaka och utvärdera lärare. Denna studie undersöker huruvida intentioner att byta skola bland lärare i socioekonomiskt missgynnade klassrum modereras av lärarbedömningsmetoder baserade på elevers akademiska prestationsdata på skolsystemsnivå. Tre-nivå hierarkisk modellering i 46 utbildningssystem genomförs utifrån Teaching and Learning International Survey (TALIS) från 2018. Resultaten visar att effekten av klassrummets socioekonomiska sammansättning på lärarnas intentioner ökar som en funktion av prestationsdatabaserad lärarbedömning. Detta gäller dock endast när bedömningen görs av externa myndigheter och inte när den görs av skolans

ledningsgrupp. Modellerna körs sedan om efter skolkonsekvenser av bedömning såsom uppsägningar, ekonomiska bonusar eller sanktioner, och sedan av lärarens egenskaper, inklusive kön, erfarenhet och undervisningsämne. Erfarna lärare i socioekonomiskt missgynnade klassrum är mer benägna att byta skola i skolsystem med mer prestationsbaserad lärarbedömning. Dessa resultat understryker de potentiella fallgroparna med prestationsdatabaserade ansvarssystem för elever i socioekonomiskt missgynnade utbildningsmiljöer.

Diskussion och slutsatser

Resultaten pekar på flera viktiga slutsatser. För det första finns det belägg för ojämlikhet i lärarsorteringen över många utbildningssystem, i varierande grad. Mönstren varierade beroende på hur lärarkvalifikationer och socioekonomisk status mättes samt hur eleverna grupperades. Därefter gav studierna blandade resultat när det gäller skolautonomi, ansvarighet, konkurrens och stratifiering, vilket tyder på att bestämningsfaktorerna för socioekonomisk lärarsortering inte lätt generaliseras enligt gränsöverskridande mönster. Trots detta var prestationsdatabaserad ansvarsskyldighet (lärarbedömning) konsekvent förknippad med högre omsättningsgrad i låga SES-miljöer. Bedömning av lärarprestationer för dem som arbetar i klassrum med lägre SES bör förlita sig på andra mätvärden än prestationsdata och bedömningarna bör utföras av dem med adekvat kunskap om skolans sammanhang. När det gäller ojämlikhet i elevresultat visade sig socioekonomisk lärarsortering enligt specialisering ha en blygsam effekt. I de flesta fall bör man prioritera att demokratisera tillgången till lärare med lämplig kunskap om pedagogiskt innehåll, men utbildningssystemen måste gå längre än att ge socioekonomiskt missgynnade elever lärare med grundläggande kvalifikationsnivåer. Även om det är en ständig utmaning för många utbildningssystem att uppmuntra de mest kompetenta lärarna att arbeta i socioekonomiskt missgynnade miljöer, är det en givande ansträngning att bygga vidare på innehållskännedomen hos underkvalificerade matematiklärare som för närvarande arbetar i miljöer med rekryteringsproblem. Slutligen, även om en minskning av lärarsortering efter specialisering sannolikt delvis kommer att lindra ojämlikheten i utbildningsresultat, är det inte ett universalmedel i det större sammanhanget med ökande inkomstskillnader och social segregation i många utbildningssystem.

References

Abadie, A., Athey, S., Imbens, G.W., & Wooldridge, J. (2017). When should you adjust standard errors for clustering? NBER Working Paper 24003.

Akiba, M., LeTendre, G.K., & Scribner, J.P. (2007). Teacher quality, opportunity gap, and national achievement in 46 countries. *Educational Researcher, 36*, 369–387.

Alatalo, T., Hansson, Å., & Johansson, S. (2021). Teachers' academic achievement: evidence from longitudinal register data. *European Journal of Teacher Education*, DOI: 10.1080/02619768.2021.1962281

Allen, R., Burgess, S., & Mayo, J. (2018). The teacher labour market, teacher turnover and disadvantaged schools: new evidence for England. *Education Economics, 26*, 4-23.

Allensworth, E., Ponisciak, S., & Mazzeo, C. (2009). The schools teachers leave: teacher mobility in Chicago public schools. Consortium on Chicago School Research. 1-52.

Anders, J., Has, S., Jerrim, J., & Zieger, L. (2021). Is Canada really an education superpower? The impact of non-participation on results from PISA 2015. *Educational Assessment, Evaluation and Accountability, 33*, 229-249.

Anderson, E. (2007). Fair opportunity in education: a democratic equality perspective. *Ethics, 117*, 595-622.

Arnesen, A. L., & Lundahl, L. (2006). Still social and democratic? Inclusive education policies in the Nordic welfare states. *Scandinavian Journal of Educational Research, 50,* 285-300.

Asparouhov, T., & Muthèn, B. (2021). Advances in Bayesian model fit evaluation for structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal, 28,* 1-14.

Bacolod, M. (2007). Who teaches and where they choose to teach: college graduates of the 1990s. *Educational Evaluation and Policy Analysis, 29*, 155-168.

Ball, S. J. (2003). The teacher's soul and the terrors of performativity. *Journal of Education Policy, 18*, 215-228.

Baumert, J. et al. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal, 47,* 33-180.

Betts, J.R., Reuben, K.S., & Danenberg, A. (2000). Equal resources, equal outcomes? The distribution of school resources and student achievement in California. Public Policy Institute of California: San Francisco.

Biesta, G. (2007). Education, accountability, and the ethical demand: can the democratic potential of accountability be regained? *Educational Theory; Urbana, 54*, 233-250.

Biesta, G. (2009). Good education in an age of measurement: on the need to reconnect with the question of purpose in education. *Educational Assessment, Evaluation and Accountability, 21,* 33-46.

Blömeke, S., Olsen, R. V., & Suhl, U. (2016). Relation of student achievement to the quality of their teachers and instructional quality. In T. Nilsen & J.- E. Gustafsson (Eds.), Teacher

quality, instructional quality and student outcomes (pp. 21–50). IEA Research for Education, vol 2. Springer, Cham.

Blömeke, S., & Olsen, R. V. (2019). Consistency of results regarding teacher effects across subjects, school levels, outcomes and countries. *Teaching and Teacher Education, 77*, 170-182.

Blömeke, S., & Delaney, S. (2012). Assessment of teacher knowledge across countries: a review of the state of research. ZDM: *The International Journal on Mathematics Education, 44*.

Baird, J.A., et al. (2016). On the supranational spell of PISA in policy. *Educational Research, 58,* 121-138.

Belsky, K., et al. (2019). Genetics & the geography of health, behavior, and attainment. *Nature Human Behaviour, 3,* 576-586.

Bitler, M., Corcoran, S., Domina, T., & Penner, E.K. (2019). Teacher effects on student achievement and height: a cautionary tale. in NBER Working Paper Series.

Bliese, P. D. (*1998*). Group size, ICC values, and group-level correlations: A simulation. *Organizational Research Methods, 1,* 355-373.

Bonesrønning, H., Falch, T., & Strøm, B. (2005). Teacher sorting, teacher quality, and student composition. *European Economic Review, 49,* 457-483.

Bourdieu, P. (1977*). Cultural Reproduction and Social Reproduction*. In J. Karabel, & A. H. Halsey (Eds.), Power and Ideology in Education (pp. 487-511).

Bourdieu, P. (1986). The Forms of Capital. In J. G. Richardson (Ed.), *Handbook of Theory and Research for the Sociology of Education* (pp. 241-258).

Bos, W., & Schwippert, K. (2003). The use and abuse of international comparative research on student achievement. *European Educational Research Journal, 2,* 559–573.

Boyd, D., Lankford, H., Loeb, S., Rockoff, J., & Wyckoff, J. (2008). The narrowing gap in New York City teacher qualifications and its implications for student achievement in high poverty schools. NBER Working Paper Series.

Boyd, D., et al. (2011). The influence of school administrators on teacher retention decisions. *American Educational Research Journal, 48*, 1-24.

Brese, F., & Mirazchiyski, P. (2013). Issues and methodologies in large-scale assessments. Special Issue 2: Measuring Students' Family Background in Large-scale International Education Studies. IERIE Monograph Series.

Brighouse, H., & Swift, A. (2006). Equality, priority, and positional goods. *Ethics, 116,* 471-497.

Brighouse, H., Ladd, H.F., Loeb, S., & Swift, A. (2016). Educational goods and values: A framework for decision makers. *Theory and Research in Education, 14,* 3-25.

Brill, F., Grayson, H., Kuhn, L., & O'Donnell, S. (2018). What impact does accountability have on curriculum, standards and engagement in education? A literature review. National Foundation for Educational Research.

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

Brown, T. A. (2015). Confirmatory factor analysis for applied research (2nd ed.). New York; London: The Guilford Press.

Brown, P., & Tannock, S. (2009). Education, meritocracy and the global war for talent. *Journal of Education Policy, 24,* 377-392.

Bryan, M.L., & Jenkins, S.P. (2016). Multilevel modelling of country effects: a cautionary tale. *European Sociological Review, 32,* 3-22.

Burger, K. (2016). Intergenerational transmission of education in Europe: Do more comprehensive education systems reduce social gradients in student achievement? *Research in Social Stratification and Mobility, 44,* 54–67.

Burger, K. (2019). The socio-spatial dimension of educational inequality: a comparative European analysis. *Studies in Educational Evaluation, 62,* 171-186.

Caro, D., & Biecek, P. (2018). Intsvy: an R package for analyzing international large-scale assessment data. *Journal of Statistical Software, 81*.

Carroll, J. B. (1963). A model of school learning. *Teachers College Record. 64*, 723-733.

Carver-Thomas, D., & Darling-Hammond, L. (2017). Teacher turnover: why it matters and what we can do about it. Learning Policy Institute.

Chetty, R., Friedman, J. N., & Rockoff, J.E. (2014). Measuring the impacts of teachers II: teacher value added and student outcomes in adulthood. *The American Economic Review,* 104, 2633–2679.

Chmielewski, A. (2019). The global increase in the socioeconomic achievement gap, 1964 2015. *American Sociological Review, 84,* 517-544.

Chiu, M. M. (2015). Family inequality, school inequalities, and mathematics achievement in 65 countries: Microeconomic mechanisms of rent seeking and diminishing marginal returns. *Teachers College Record, 117*, 1–32.

Chudgar, A, & Luschei, T.F. (2011). Teachers, student achievement and national income: A cross-national examination of relationships and interactions. *Prospects, 41,* 507-533.

Clotfelter, C., Ladd, H., & Vigdor, J.L. (2005). Who teaches whom? Race and the distribution of novice teachers. *Economics of Education Review, 24,* 377-392.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J.L. (2007). Teacher credentials and student achievement: Longitudinal analysis with student fixed effects. *Economics of Education Review, 26*, 673–682.

Clotfelter, C., Glennie, E., Ladd, H., & Vigdor, J.L. (2008). Would higher salaries keep teachers in high-poverty schools? Evidence from a policy intervention in North Carolina. *Journal of Public Economics, 92,* 1352–1370.

Clotfelter, C., Ladd, H., & Vigdor, J.L. (2006). Teacher-student matching and the assessment of teacher effectiveness. NBER Working Paper.

Clotfelter, C., Ladd, H., & Vigdor, J.L. (2011). Teacher mobility, school segregation, and pay-based policies to level the playing field. *Education Finance and Policy, 6,* 399-438.

Coleman, J. (1966). *Equality of educational opportunity*. United States and National Center for Education Statistics. Washington: U.S. Dept.

Cordero, J.M., Cristobal, V., & Santin, D. (2017). Causal inference on education policies: a survey of empirical studies using PISA, TIMSS, and PIRLS. *Journal of Economic Surveys*, *32*, 878-915.

Craig, C. (2017). International teacher attrition: multiperspective views. *Teachers and Teaching, 23*, 859-862.

Creemers, B. (2005). Educational effectiveness: the development of the field. Keynote address presented at the first International Conference on School Effectiveness and School Improvement in China, September.

Creemers, B., & Kyriakides, L. (2007). *The dynamics of educational effectiveness.* Routledge.

Creemers, B. P. M., & Schaveling, J. (1985). *Verhoging van onderwijseffectiviteit [Improving educational effectiveness]*. The Hague: WRR.

Cronbach, L.J., & Meehl, P.E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52,* 281-302.

Crouch, C., Finegold, D., & Sako, M. (1999). Are skills the answer? New York: Oxford University Press.

Darling-Hammond, L. (2000). Teacher quality and student achievement: a review of state policy evidence. *Educational Policy Analysis Archives, 8*, 1-44.

Darling-Hammond, L. (2004a). Access to qualified teachers in California's Public Schools. *Teachers College Record, 106*, 1936-1966.

Darling-Hammond, L. (2004b). Standards, accountability, and school reform. *Teachers College Record, 106,* 1047-1085.

Darling-Hammond, L. (2006). Constructing 21st century teacher education. *Journal of Teacher Education, 57,* 300-314.

Darling-Hammond, L. (2012). *Powerful teacher education: lessons from exemplary programs.* John Wiley and Sons.

Darling-Hammond, L. (2017). Teacher education around the world: What can we learn from international practice? *European Journal of Teacher Education, 40,* 291-309.

DeAngelis, K.J., White, B., & Presley, J.B. (2010). The changing distribution of teacher qualifications across schools: a statewide perspective post NCLB. *Education Policy Analysis Archives, 18,* 1-34.

Dieterle, S., Guarino, C.M., Reckase, M.R., & Wooldridge, J.F. (2014). How do principals assign students to teachers? Finding evidence in administrative data and implications for value-added. *Journal of Policy Analysis and Management, 34,* 32-58.

Donaldson M. L. (2009). *So long, Lake Wobegon? Using teacher evaluation to raise teacher quality.* Washington, DC: Center for American Progress. Retrieved from https://cdn.americanprogress.org/wp-content/uploads/issues/2009/06/pdf/teacher_evaluation.pdf

Downey, D.B., & Condron, D.J. (2016). Fifty years since the Coleman Report. *Sociology of Education, 89,* 207-220.

Dupriez, V., & Dumay, X. (2006). Inequalities in school systems: effect of school structure or of society structure? *Comparative Education, 42,* 243-260.

Ecalle, J., Magnan, A., & Gilbert, F. (2006). Class size effects on literacy skills and literacy interest in first grade: A large-scale investigation. *Journal of School Psychology, 44,* 191-209.

Elacqua, G., Hincapie, D., & Martinez, M. (2019). The effects of the Chilean school accountability system on teacher turnover. IDB Working Paper Series Number 1075.

Engel, M., & Finch, M. (2014). How urban principals find teachers and make hiring decisions. *Leadership and Policy in Schools, 14,* 12-41.

Engel, M., Jacob, B.A., & Curran, C. (2014). New evidence on teacher labor supply. *American Educational Research Journal, 51,* 36-72.

Engzell, P. (2019). What do books in the home proxy for? A cautionary tale. *Sociological Methods and Research*, 1-28.

Fantilli, R.D., & Macdougall, D.E. (2009). A study of novice teachers: challenges and supports in the first years. *Teaching and Teacher Education, 25*, 814-825.

Finn, J.D., Pannozzo, G.M., & Achilles, C.M. (2003). The "why's" of class size: student behavior in small class sizes. *Review of Educational Research, 73,* 321-368.

Feng, L. (2010). Hire today, gone tomorrow: New teacher classroom assignments and teacher mobility. *Education Finance and Policy, 5*, 278–316.

Feng, L., Figlio, D.N., & Sass, T. (2018). School accountability and teacher mobility. *Journal of Urban Economics, 103*, 1-41.

Ferreira, F., & Gignoux, J. (2014). The measurement of educational inequality: achievement and opportunity. *The World Bank Economic Review, 28,* 210-246.

Franck, E., & Nicaise, I. (2022). The effectiveness of equity funding policies in schools in Europe and North America: A systematic literature review. *Issues in Educational Research, 32,* 494-512.

Freeman, R.B., Machin, S.J., & Viarengo, M. (2011). Inequality of educational outcomes: international evidence from PISA. Regional and Sectoral Economic Studies, Euro-American Association of Economic Development, 11.

Goe, L. (2007). The Link Between Teacher Quality and Student Outcomes: A Research Synthesis. National Comprehensive Center for Teacher Quality, Washington, DC, USA. Retrieved from: http://www.gtlcenter.org/sites/default/files/docs/LinkBetweenTQandStudentOutcomes.pdf

Goldhaber, D. D., & Brewer, D. J. (2000). Does teacher certification matter? High school teacher certification status and student achievement. *Educational Evaluation and Policy Analysis, 22*, 129-145.

Goldhaber, D., Lavery, L., & Theobald, R. (2015). Uneven playing field? Assessing the teacher quality gap between advantaged and disadvantaged students. *Educational Researcher, 44,* 293-307.

Goldhaber, D., Quince, V., & Theobald, R. (2018). Has it always been this way? Tracing the evolution of teacher quality gaps in U.S. public schools. *American Educational Research Journal, 55*, 171-201.

Gonzalez, E. (2014). Calculating standard errors of sample statistics when using international large scale assessment data. *Educational Policy Evaluation Through International Comparative Assessments, 7*, 59-73.

Grissom, J.A., Kalogrides, D., & Loeb, S. (2015). The micropolitics of educational inequality: the case of teacher-student assignments. *Peabody Journal of Education, 90*, 601-614.

Gustafsson, J.E. (2008). Effects of International Comparative Studies on Educational Quality on the Quality of Educational Research. *European Educational Research Journal, 7,* 1–17.

Gustafsson, J.E. (2013). Causal inference in educational effectiveness research. *School Effectiveness and School Improvement, 24,* 275-295.

Gustafsson, J. E. (2016). Lasting effects of quality of schooling: Evidence from PISA and PIAAC. *Intelligence, 57*, 66–72.

Gustafsson, J.E., & Rosèn, M. (2014). Quality and credibility of international studies. in Strietholt, Rolf, Wilfried Bos, Jan-Eric Gustafsson, and Monica Rosén. 2014. Educational Policy Evaluation through International Comparative Assessments. Münster/New York: Waxmann.

Gustafsson, J.E., Nilsen, T., & Yang Hansen, K. (2018). School characteristics moderating the relation between student socioeconomic status and mathematics achievement in grade 8. Evidence from 50 countries in TIMSS 2011. *Studies in Educational Evaluation, 57*, 16-30.

Gutiérrez, G., Jerrim, J., & Torres, R. (2019). School segregation across the world: has any progress been made in reducing the separation of the rich from the poor? *The Journal of Economic Inequality, 18, 157*-179.

Han, S.W. (2018). School-based teacher hiring and achievement inequality: a comparative Perspective. *International Journal of Educational Development, 61*, 82-91.

Hansson, Å., & Gustafsson, J.E. (2016). Pedagogisk Segregation: Lärarkompetens i en Svenska Grundskolan ur ett Likvärdighetsperspektiv [Pedagogical Segregation: Teacher Competence in Swedish Elementary School from an Equality Perspective, in Swedish]. *Pedagogisk Forskning i Sverige, 21*, 56-78.

Hanushek, E. A. (1971). Teacher Characteristics and Gains in Student Achievement: Estimation Using Micro Data. *American Economic Review, 61,* 280–88.

Hanushek, E.A., & Kain, J.F. (1972). On the value of `equality of educational opportunity' as a guide to public policy. In *On equality of educational opportunity.* (Eds). Mosteller, F., & Moynihan, D-P. New York: Random House: 116-145.

Hanushek, E.A., Kain, J.F., O'Brien, D.M., & S.G. Rivkin. (2005). The market for teacher quality. NBER Working Paper Series.

Hanushek, E.A., Piopiunik, M., & Wiederhold, S. (2014). The value of smarter teachers: international evidence on teacher cognitive skills and student performance. NBER, 1-55.

Hanushek, E.A. & Woessman, L. (2011). *The Economics of International Differences in Educational Achievement.* Pp. 89-200 in Handbook of the Economics of Education, Vol. 3, edited by E. A. Hanushek, S. Machin, and L. Woessmann . Amsterdam: the Netherlands.

Hanushek, EA., & Rivkin, S. (2003). Does public school competition affect teacher quality? NBER Working Paper.

Hanushek, E.A., & Rivkin, S.G. (2010a). The quality and distribution of teachers under the No Child Left Behind Act, *The Journal of Economic Perspectives, 24,* 133-150.

Hanushek, E.A., & Rivkin, S.G. (2010b). Constrained job matching: does teacher job search harm disadvantaged urban schools? NBER Working Paper w15816.

Hanushek, E.A., & Rivkin, S.G. (2012). The distribution of teacher quality and implications for policy. *Annual Review of Economics, 4*, 131–157.

Hanushek, E. A., & Woessmann, L. (2015). *The knowledge capital of nations: education and the economics of growth.* Cambridge: MIT Press.

Hanushek, E., & Woessmann, L. (2006). Does educational tracking affect performance and inequality? Differences-in-differences evidence across countries. *The Economic Journal, 116,* 63-76.

Hanushek, E., Link, S., & Woessmann, L. (2013). Does school autonomy make sense everywhere? Panel estimates from PISA, *Journal of Development Economics, 104,* 212-232.

Hanushek, E.A., & Woessmann, L. (2014). Institutional structures of the education system and student achievement: a review of cross-country economic research. Pp. 129-144 in Rolf Strietholt, Jan-Eric Gustafsson, Monica Rosen and Wilfried Bos (eds) *Educational policy evaluation through international comparative assessments*. New York: Waxmann.

Hanushek, E.A., et al. (2016). Dynamic effects of teacher turnover on the quality of instruction. *Economics of Education Review, 55,* 132-148.

Harris, D. N., & Sass, .T.R. (2011). Teacher Training, Teacher Quality and Student Achievement. *Journal of Public Economics, 95,* 798–812.

Harwell, M., Maeda, Y., Bishop, K., & Xie, A. (2017). The surprisingly modest relationship between SES and educational achievement. *The Journal of Experimental Education, 85,* 197-214.

Hattie, J. (2003). Teachers make a difference: what is the research evidence? Paper presented at the Australian Council for Educational Research Annual Conference on Building Teacher Quality, Melbourne. https://research.acer.edu.au/research_conference_2003/4

Hattie, J. A. (2009). Visible learning: A synthesis of 800+ meta-analyses on achievement. Abingdon: Routledge.

Hill, H.C., Rowan, B., & Ball, D.L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal, 42,* 449-466.

Hill, H. C., Ball, D. L., & Schilling, S. G. (2008). Unpacking pedagogical content knowledge: Conceptualizing and measuring teachers' topic-specific knowledge of students. Journal for research in mathematics education, 372-400.

Horng, E.L. (2005). Poor working conditions make urban schools hard-to-staff. UCACCORD Public Policy Series, PB-010-0305.

Husén, T. (1979). An International Research Venture in Retrospect: The IEA Surveys. *Comparative Education Review, 23,* 371–385.

IEA. (2020). TIMSS Technical Report. https://timssandpirls.bc.edu/timss2019/methods/

Ingersoll, R.M. (2001). Teacher turnover and teacher shortages: an organizational analysis. *American Educational Research Journal, 38*, DOI: 10.3102/00028312038003499

Ingersoll, R.M. (2002). Who controls teachers' work. Cambridge, MA: Harvard University Press.

Ingersoll, R., Merrill, L., & May, H. (2016). Do accountability policies push teachers out? *Educational Leadership, 73,* 44-49.

Ingersoll, R. M. (2017). Misdiagnosing America's teacher quality problem. In G. K. LeTendre & M. Akiba (Eds.), International handbook of teacher quality and policy (pp. 79–96). London: Routledge.

Isenberg, E., Max, J., Gleason, P., & Deutsch, J. (2021). Do low-income students have equal access to effective teachers? *Educational Evaluation and Policy Analysis,* https://doi.org/10.3102%2F01623737211040511

Jackson, C.K. (2009). Student demographics, teacher sorting, and teacher quality: Evidence from the end of school desegregation. *Journal of Labour Economics, 27,* 213-256.

Jackson, C.K. (2018). What do test scores miss? The importance of teacher effects on non–test score outcomes. *Journal of Political Economy,* 126, 2072-2107.

Jackson, C.K., Rockoff, J.E., & Staiger, D.O. (2014). Teacher effects and teacher-related policies. *Annual Review of Economics, 6,* 801-825.

Jerrim, J., & Micklewright, J. (2014). Socioeconomic gradients in children's cognitive skills: are cross-country comparisons robust to who report family background? *European Sociological Review, 30*, 766-781.

Jencks, C., et al. (1972). *Inequality: A reassessment of the effects of family and schooling in America.* New York, NY: Basic Books.

Jenkins, R. (1982). Pierre Bourdieu and the reproduction of determinism. *Sociology, 16,* 270-281.

Jerrim, J. (2012). The socio-economic gradient in teenagers' reading skills: How does England compare with other countries? *Fiscal Studies, 33*, 159–184.

Jerrim, J., Volante, L., Klinger, D. A., & Schnepf, S. V. (2019). Socioeconomic inequality and student outcomes across education systems. In L. Volante, S. V. Schnepf, J. Jerrim, & D. A. Klinger (Eds.), Socioeconomic inequality and student outcomes: Cross-national trends, policies, and practices (pp. 3-16). Singapore: Springer.

Jerrim, J, Lopez-Agudo, L.A., Marcenaro-Gutierrez, & Shure, N. (2017). What Happens When Econometrics and Psychometrics Collide? An Example Using the PISA Data. *Economics of Education Review, 61,* 51–58.

Jerrim, J., & Sims, S. (2021). School accountability and teacher stress: international evidence from the OECD TALIS study. *Educational Assessment, Evaluation and Accountability, 34,* 5-32.

Jerrim, J., Sims, S., Taylor, H., & Allen, S. (2021). Has the mental health of teachers in England changed over time? New evidence from three datasets. *Oxford Review of Education, 47,* 805-825.

Johansson, S. (2016). International large-scale assessments: what uses, what consequences? *Educational Research, 58*, 139-148.

Johansson, S., & Myrberg, E. (2019). Teacher specialization and student perceived instructional quality: what are the relationships to student reading achievement? *Educational Assessment, Evaluation and Accountability*, *31*, 1-24.

Kahneman, D., & Tversky, A. (1979). Prospect theory: an analysis of decision under risk. *Econometrica, 47,* 263-292.

Kang, N. H., & Hong, M. (2008). Achieving excellence in teacher workforce and equity in learning opportunities in South Korea. *Educational Researcher, 37*, 200–207.

Kalogrides, D., Loeb, S., & Beteille, T. (2013). Systematic sorting of teacher characteristics and class Assignments. *Sociology of Education, 86*, 103-123.

Kamens, D. H., & McNeely, C. L. (2010). Globalization and the growth of international educational testing and national assessment. *Comparative Education Review, 54*, 5–25.

Karsentil, T., & Collin, S. (2013). Why are new teachers leaving the profession? Results of a Canada-wide survey. *Education, 3,* 141-149.

Kunter, M., Klusmann, U., Baumert, J., Richter, D., Voss, T., & Hachfeld, A. (2013b). Professional competence of teachers: Effects on instructional quality and student development. *Journal of Educational Psychology, 105,* 805-820.

Komatsu, H., & Rappleye, J. (2017). A new global policy regime founded on invalid statistics? Hanushek, Woessmann, PISA, and economic growth. *Comparative Education, 53*, 166-191.

Kreft, I. & De Leeuw, J. (1998) *Introduction to Multilevel Modelling.* Sage Publications, London.

Kyriakides, L. (2007). Generic and differentiated models of educational effectiveness: implications for the improvement of educational practice. In T., Townsend. (Eds). *International Handbook of School Effectiveness and Improvement.* Springer. 41-56.

Kyriakides, L., Christoforou, C., & Charalambous, C.Y. (2013). What matters for student learning outcomes: a meta-analysis of studies exploring factors of effective teaching. *Teaching and Teacher Education, 35*, 143-152.

Kyriakides, L., Creemers, B., Antoniou, P., & Demetriou, D. (2010). A synthesis of studies searching for school factors: implications for theory and research. *British Educational Research Journal, 36,* 807-830.

Kyriakides, L., Creemers, B., & Charambolous, E. (2018). Equity and quality dimensions in educational effectiveness.

Lankford, H., Loeb, S., & Wyckoff, J. (2002). Teacher sorting and the plight of urban schools: a descriptive analysis. *Educational Evaluation and Policy Analysis, 24*, 37-62.

Lavrijsen, J., & Nicaise, I. (2016). Ascription, achievement and perceived equity of educational regimes: an empirical investigation. *Social Sciences, 5,* 1-18.

Lazear, E.P. (2001). Educational production. *The Quarterly Journal of Economics, 116,* 777-803.

LeBreton, J.M., & Senter, J.L. (2008). Answers to 20 questions about interrater reliability and interrater agreement. *Organizational Research Methods, 11,* 815-852.

Lee, J., & Wong, K.K. (2003). The impact of accountability on racial and socioeconomic equity: considering both school resources and achievement outcomes. *American Educational Research Journal, 41,* 797-832.

Leinket, J., & Caro, D. (2014). Performance status and change – measuring education system effectiveness with data from PISA 2000-2009. *Educational Research and Evaluation, 20,* 146-174.

Lindqvist, P., Nordänger, U. K., & Carlsson, R. (2014). Teacher attrition the first five years – A multifaceted image. *Teaching and Teacher Education, 40*, 94-103.

Lundgren, U.P. (1972). Frame factors and the teaching process: a contribution to curriculum theory and theory on teaching. PhD Diss.

Luschei, T.F. (2012). In search of good teachers: Patterns of teacher quality in two Mexican states. *Comparative Education Review, 56,* 69-97.

Luschei, T.F., & Carnoy, M. (2010). Educational production and the distribution of teachers in Uruguay. *International Journal of Educational Development, 30,* 169-181.

Luschei, T.F., Chudgar, A., & Rew, J. (2013). Exploring differences in the distribution of teacher qualifications across Mexico and South Korea: Evidence from the teaching and learning international survey. *Teachers College Record, 115,* 1–38.

Luschei, T.F., & Jeong, D.W. (2019). Is teacher sorting a global phenomenon? Cross-national evidence on the nature and correlates of teacher quality opportunity gaps. *Educational Researcher, 47,* 556-576.

Marks, G.N., & O'Connell, M. (2021). Inadequacies in the SES-achievement model: evidence from PISA and other studies. *Review of Education, 9,* https://doi.org/10.1002/rev3.3293

Maroy, C., & Voisin, A. (2017). THINK PIECE ON ACCOUNTABILITY: Think piece on accountability background paper prepared for the 2017/8 Global Education Monitoring Report. [Research Report] UNESCO. 2017. halshs-01705982

Marsh, H.W., et al. (2009). Doubly-latent models of school contextual effects: integrating multilevel and structural equation approaches to control measurement and sampling error. *Multivariate Behavioral Research, 44*, 764-802.

Marcinczak, S., Musterd, S., van Ham, M., & Tammaru, T. (2016). Inequality and rising levels of socioeconomic segregation: lessons from a pan-European comparative study. In T. Tammaru, S. Marcinzcak, M. van Ham, and S. Musterd (Eds). *East meets West: New perspectives on socio-economic segregation in European capital cities (p. 358-382).* London: Routledge.

McDonnell, L. M. (1995). Opportunity to learn as a research concept and a policy instrument. *Educational Evaluation and Policy Analysis, 17*, 305-322.

Meckes, L., & Bascopé, M. (2012). Uneven distribution of novice teachers in the Chilean primary school system. *Educational Policy Analysis Archives, 20,* 1-27.

Meyer, H.D., & Benavot, A. (2013). *PISA, power and policy: the emergence of global education governance.* Oxford Studies in Comparative Education.

Mons, N. (2007). *Les nouvelles politiques educatives: La France fait-elle les bons choix?* Paris: Presses Universitaires de France.

Mullis, I., Michael, V.S., Martin., O., & Loveless, T. (2016). 20 Years of TIMSS: International Trends in Mathematics and Science Achievement, Curriculum, and Instruction: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College and International Association for the Evaluation of Educational Achievement (IEA).

Mummolo, J., & Peterson, E. (2018). Improving the interpretation of fixed-effects regression results. *Political Science Research and Methods, 6,* 829-835.

Murnane, R. (1975). *The impact of school resources on the learning of inner-city children.* Cambridge, MA: Ballinger.

Muthén, L. K., & Muthén, B. O. (1998-2018). Mplus User's Guide. Sixth Edition. Los Angeles, CA: Muthén & Muthén.

Muthén, B. (2010). *Bayesian analysis in Mplus: A brief introduction.*

Nash, R. (1990). Bourdieu on education and social and cultural reproduction. *British Journal of Sociology of Education, 11,* 431-447.

Nilsen, T., & Gustafsson, J.E. (2016). *Teacher Quality, Instructional Quality, and Student Outcomes: Evidence across Countries, Cohorts, and Time.* Springer.

Nilsen, T., Scherer, R., & Blömeke, S. (2018). The relation of science teachers' quality and instruction to student motivation and achievement in 4th and 8th Grade: a Nordic perspective. Nordic Council of Ministers: *Northern lights on TIMSS and PISA.*

Nóvoa, A., & Yariv-Mashal, T. (2003). Comparative research in education: A mode of governance or a historical journey? *Comparative Education, 39*, 423–438.

Nye, B., Konstantopoulos, S., & Hedges, L.V. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis, 26,* 237-57.

OECD. (2005). *Teachers matter: attracting, developing and retaining effective teachers.* Paris: OECD Publishing.

OECD. (2018a). *Equity in education: breaking down barriers to social mobility*. OECD publishing: Paris.

OECD. (2018b). *Effective Teacher Policies. Insights from PISA.* Paris: OECD Publishing.

OECD. (2019a). *TALIS 2018 technical report*. Paris: OECD Publishing.

OECD. (2019b). *Balancing school choice and equity. An international perspective based on PISA.* Paris: OECD Publishing.

OECD. (2019c). *PISA 2018 results (Volume I). Annex B3 PISA 2018 system-level indicators*. Paris: OECD Publishing.

OECD. (2022). *Mending the education divide: getting strong teachers to the schools that need them most.* Paris: TALIS, OECD Publishing.

Özoğlu, M. (2015). Mobility-related teacher turnover and the unequal distribution of experienced teachers in Turkey. *Educational Sciences: Theory and Practice, 15,* 891-909.

Palardy, G.J. (2013). High school socioeconomic segregation and student attainment. *American Educational Research Journal, 50,* 714-754.

Papay, J. P., & Kraft, M.A. (2015). Productivity returns to experience in the teacher labor market: methodological challenges and new evidence on long-term career improvement. *Journal of Public Economics, 130,* 105–119.

Park, J. (2019). Elementary science teacher education in Korea: Past, present, and future. *Asia-Pacific Science Education*, *5*, 1–11.

Perryman, J., Ball, S., Maguire, M., & Braun, A. (2011). Life in the pressure cooker – school league tables and English and Mathematics teachers' responses to accountability in a results driven era. *British Journal of Educational Studies, 59*, 179-195.

Pokropek, A., Borgonovi, F., & McCormick, C. (2017). On the cross-country comparability of indicators of socioeconomic resources in PISA. *Applied Measurement in Education, 30,* 243-258.

Qian, H., & Walker, A. (2019). Reconciling top-down policy intent with internal accountability: the role of Chinese school principals. *Educational Assessment, Evaluation and Accountability, 31, 495*-517.

Raudenbush, S.R., & Bryk, A.S. (2002). *Hierarchical linear models: applications and data analysis methods.* CA: Sage Publications.

Reardon, S. F. (2011). The widening academic achievement gap between the rich and the poor: New evidence and possible explanations. In G. J. Duncan & R. J. Murnane (Eds.), *Whither Opportunity?: Rising Inequality, Schools, and Children's Life Chances* (pp. 91- 116). New York, NY: Russel Sage Foundation.

Rice, J. (2003). *Teacher Quality: Understanding the Effectiveness of Teacher Attributes*. Washington, DC: Economic Policy Institute.

Rivkin, S.G., Hanushek, E.A, & Kain, J.F. (2005). Teachers, schools, and academic achievement. *Econometrica, 73*, 417-458.

Rockoff, J. (2004). The impact of individual teachers on student achievement: evidence from panel data. *American Economic Review, 94*, 247–252.

Rockoff, J. E., Jacob, B. A., Kane, T. J., & Staiger, D. O. (2011). Can you recognize an effective teacher when you recruit one? *Education Finance and Policy, 6,* 43–74.

Roemer, J. (1998). Equality of Opportunity. Cambridge University Press.

Rolfe, V., Strietholt, R., & Yang Hansen, K. (2021). Does inequality in opportunity perpetuate inequality in outcomes? International evidence from four TIMSS cycles. *Studies in Educational Evaluation, 71,* https://doi.org/10.1016/j.stueduc.2021.101086

Rothstein, J. (2012). Teacher quality policy: when supply matters. NBER Working Paper.

Rubin, D.B., (1987). *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley and Sons.

Rutkowski, L., Gonzalez, E., Joncas, M., & von Davier, M. (2010). Methodological challenges with international large-scale assessment data. *Educational Researcher, 39*, 142-151.

Rutkowski, L., & Rutkowski, D. (2009). Trends in TIMSS responses over time: Evidence of global forces in education? *Educational Research and Evaluation, 15*, 137-152.

Rutkowski, D., & Rutkowski, L. (2013). Measuring socioeconomic background in PISA: One size might not fit all. *Research in Comparative and International Education, 8,* 259-278.

Rutkowski, D., & Delandshere, G. (2016). Causal Inferences with Large Scale Assessment Data: Using a Validity Framework. *Large-Scale Assessments in Education 4,* 1-18.

Rutter, M., Maughan, B., Mortimore, P., & Ouston, J. (1979). *Fifteen thousand hours: Secondary schools and their effects on children*. London: Open Books.

Sacerdote, B. (2011). Peer effects in education: how might they work, how big are they and how much do we know thus far. Chapter 04 in *Handbook of the Economics of Education, 3,* 249-277.

Sancassani, P. (2021). The effect of teacher characteristics on students' science achievement. Ifo Working Paper Series 348, ifo Institute, Leibniz Institute for Economic Research at the University of Munich.

Scheerens, J. (2013). The use of theory in school effectiveness research revisited. *School Effectiveness and School Improvement, 24*, 1-38.

Scheerens, J., & Bosker, R. J. (1997). *The foundations of educational effectiveness*. Oxford, UK: Pergamon.

Schmidt, W.H., & McKnight, C.C. (2012). *Inequality for all: The challenge of unequal opportunity in American schools*. New York, NY: Teachers College.

Schmidt, W. H., Zoido, P., & Cogan, L. (2013). Schooling matters: Opportunity to learn in PISA 2012. OECD Education Working Papers (95). doi:10.1787/5k3v0hldmchl-en

Schmidt, W. H., Raizen, S., Britton, E., Bianchi, L., & Wolfe, R. (1997). *Many visions, many aims volume 2: A cross-national investigation of curricular intentions in school science*. Dordrecht: Springer.

Schlotter, M., Schwerdt, G., & Woessmann, L. (2009). Econometric methods for causal evaluation of education policies and practices: a non-technical guide. CESifo Working Paper, No. 2877, Center for Economic Studies and Ifo Institute (CESifo), Munich.

Schwille, M.T. Ingvarsson, L., Tatto, G. Rowley, R. Peck, & S. Senk. (2013). (eds.). An Analysis of
Teacher Education Context, Structure, and Quality Assurance Arrangements in *TEDS-M Countries: Findings from the IEA Teacher Education and Development Study in Mathematics (TEDS-M).* Amsterdam: International Association for the Evaluation of Educational Achievement (IEA).

Seebruck, R. (2021). How teacher rotation in Japanese high schools affects clustering of teacher quality: comparing the distribution of teachers across public and private sectors. *Education Policy Analysis Archives, 29,* https://doi.org/10.14507/epaa.29.5362

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher, 15,* 4-14.

Shulman, L.S. (1987). Knowledge and teaching: foundations of the new reform. *Harvard Educational Review, 57,* 1-23.

Sibieta, L. (2018). The teacher market in England: shortages, subject expertise and incentives. Teaching and Leadership: supply and quality. Report from the Education Policy Institute.

Sims, S., & Allen, R. (2018). Do pupils from low-income families get low-quality teachers? Indirect evidence from English schools. *Oxford Review of Education, 44,* 441-458.

Sirin, S.R. (2005). Socioeconomic status and academic achievement: a meta-analytic review of the research. *Review of Educational Research, 75,* 417-453.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2001). *Experimental and quasi-experimental designs for generalized causal inference.* Boston: Houghton Mifflin.

Slater, H., Davies, N. M., & Burgess, S. (2012). Do teachers matter? Measuring the variation in teacher effectiveness in England. *Oxford Bulletin of Economics and Statistics, 74*, 629–645.

Solga, H. (2014). Education, economic inequality and the promises of the social investment state. *Socio-Economic Review, 12*, 269-297.

Sommet, N., & Morselli, D. (2017). Keep calm and learn multilevel logistic modeling: A simplified three-step procedure using Stata, R, Mplus, and SPSS. *International Review of Social Psychology, 30*, 203-218.

Sorensen, L.C., & Ladd, H.F. (2020). The hidden costs of teacher turnover. *AERA Open, 6,* https://doi.org/10.1177/2332858420905812

Smith, W.C., & Holloway, J. (2020). School testing culture and teacher satisfaction. *Educational Assessment, Evaluation and Accountability, 32,* 461-479.

Strello, A.I., Strietholt, R., Steinmann, I., & Siepmann, C. (2020). Early tracking and different types of inequalities in achievement: difference-in-differences evidence from 20 years of large-scale assessments. *Educational Assessment, Evaluation and Accountability, 33,* 139-167.

Strietholt, R., Bos, W., Gustafsson, J.E., Rosén, M., & Bos, W. (2014). *Educational policy evaluation through international comparative assessments.* Waxmann: New York.

Strietholt, R., Gustafsson, J.E., Hogrebe, N., & Rolfe, V. (2019). The impact of education policies on socioeconomic inequality in student achievement: a review of comparative studies. In: (Eds) Volante, L., Schnepf, S. V., Jerrim, J., Klinger, D. A. *Socioeconomic Inequality and Student Outcomes: Cross-National Trends, Policies, and Practices.* Springer. 17-38.

Teltemann, J., & Schunk, R. (2019). Education systems, school segregation, and second generation immigrants' educational success: Evidence from a country-fixed effects approach using three waves of PISA. *International Journal of Comparative Sociology, 57*, 401-424.

Toropova, A, Johnson, S., & Myrberg, E. (2019). The role of teacher characteristics for student achievement in mathematics and student perceptions of instructional quality. *Education Inquiry, 10*, 274-299.

Traynor, A., & Raykov, T. (2013). Household possessions indices as wealth measures: a validity evaluation. *Comparative Education Review, 57*, 662-688.

United Nations. (1948). *Declaration of human rights. Adopted and proclaimed by General Assembly resolution 217 A (III) of 10 December 1948*.

UNDP. (2020). Human development report (PDF) Technical notes. United National Development Programme, 2-4.

Valli, L., & Buese, D. (2007). The changing roles of teachers in an era of high-stakes accountability. *American Educational Research Journal, 44,* 519-558.

Van Buuren, S., & Groothuis-Oudshoorn, K. (2011). Mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software, 45*, 1-67.

van de Werfhorst, H.G., & Mijs, J.B. (2010). Achievement inequality and the institutional structure of educational systems: a comparative perspective. *Annual Review of Sociology, 36*, 407-428.

Wang, A.H., Coleman, A.B., Coley, R.J., & Phelps, R.P. (2003). *Preparing teachers around the world: Policy information report.* Educational Testing Service: Princeton.

Wayne, A. J., & Youngs, P. (2003). Teacher characteristics and student achievement gains: A review. *Review of Educational Research,* 73, 89–122.

White, K. R. (1982). The relation between socioeconomic status and academic achievement. *Psychological Bulletin, 91,* 461-481

William, D. (2006). Assessment for learning: why, what and how? Paper presented at the Cambridge Assessment Network Conference.

William, D. (2010). Teacher quality: how to get more of it. Paper presented at Spectator Schools Revolution Conference: March.

William, D. (2018). *Creating the schools our children need: why what we're doing now won't help much (and what we can do instead).* Learning Sciences International.

Willis, P. (1977). *Learning to labor: How working class kids get working class jobs.* United Kingdom: Saxon House.

Wiswall, M. (2013). The dynamics of teacher quality. *Journal of Public Economics, 100*, 61–78.

Woessmann, L. (2004). How equal are educational opportunities? Family background and student achievement in Europe and the US. IZA DP No. 1284.

Wu, M. (2005). The role of plausible values in large-scale surveys. *Studies in Educational Evaluation, 31,* 114-128.

Wu, M. (2009). A critical comparison of the contents of PISA and TIMSS mathematics assessments. University of Melbourne. June 2, 2009.

Yang, Y., and Gustafsson, J.E. (2004). Measuring socioeconomic status at individual and collective levels. *Educational Research and Evaluation, 10,* 259-288.

Yang Hansen, K., and Gustafsson, J.E. (2016). Causes of educational segregation in Sweden – school choice or residential segregation. *Educational Research and Evaluation, 22,* 23-44.

Appendix A

Table A Descriptive statistics and missing data (TIMSS 1999-2019)

Variables	N	Min	Max	Mean	SD	% Missing
PV1math	870457	17.99	918.10	491.113	105.880	0
PV2math	870457	5.00	937.73	491.367	106.656	0
PV3math	870457	5.00	937.21	491.004	106.007	0
PV4math	870457	5.00	939.95	491.995	106.889	0
PV5math	870457	5.00	939.03	491.422	106.925	0
tnew	813129	0	1	.2249	.41750	6.4
tnspec	789127	0	1	.1626	.36903	9.9
sfobo	852852	0	1	.07		1.9
smoed	633666	1	7	3.29	1.852	24.2
sfaed	601409	1	7	3.28	1.859	28.4
sbooks	856547	1	5	2.83	1.276	1.3
cloc	828452	0	1	.40		5.3
cdis	673550	0	1	.27		25
stratio	725667	1	200	21.17		17
cabil	799865	0	1	.49		9.2
cshort	743556	1	3	1.29		14
ICCses	870457	.02	.58	.218	.10366	0
expnt	870457	.02	.65	.2306	.13742	0
exput	870457	.00	.84	.186	.17435	0
gapnt	870457	-.10	.35	.045	.07062	0
gaput	870457	-.07	.28	.023	.044	0

Appendix B

Table B Descriptive statistics and missing data (TALIS 2018)

Variables	N	Min	Max	Mean	SD	% Missing
tchange	147706	.00	1.00	.2180		3.9
tnew	152081	.00	1.00	.1829		1.0
tnewMS	145090	.00	1.00	.0953		5.6
tnspecALL	141163	.00	1.00	.2503		8.1
tnspecMS	151246	.00	1.00	.0691		1.6
tfemale	153674	.00	1.00	.6906		.00
T3SELF	145247	.67	19.22	12.6786	2.00042	5.5
T3JSPRO	148329	4.09	15.36	11.4367	1.98740	3.5
Tsesdis	147842	1.00	5.00	2.2904		3.8
Timm	145617	1.00	5.00	1.7166		5.2
Tspene	148094	1.00	5.00	1.9971		3.6
Tlang	148289	1.00	5.00	1.9066		3.5
spriv	137700	.00	1.00	.1828		10.4
srural	146690	.00	1.00	.3445		4.5
TC3G17C	147103	1.00	5.00	2.5338		4.3
sfire	138784	.00	1.00	.2565		9.7
ssalincrease	126129	.00	1.00	.1320		17.9
ssanction	132064	.00	1.00	.2019		14.1
ctrabet	150025	.00	1.00	.7930		0
ctrawith	143673	.10	.98	.4050	.22828	0
cautsfu	153682	.01	1.00	.5637	.31218	0
sautsfu	147070	.00	1.00			4.3
csaccnt	153682	.07	1.00	.4002	.17316	0
saccnt	136333	.00	1.00			11.3
cscomp	153682	.29	.96	.6325	.15990	0
sscomp	137212	.00	1.00			10.7
csshort	153682	.35	.95	.6680	.16096	0
sshort	147272	.00	1.00			4.2
HDI	147471	.71	.95	.8673	.05549	0

Appendix C

Table C Factor loadings by country and cycle for SES variables TIMSS 1999-2019

	1999	2003	2007	2011	2015	2019		1999	2003	2007	2011	2015	2019
AUS							LTU						
Mother ed	0.854	0.823	0.836	0.867	0.863	0.864	Mother ed	0.868	0.838	0.857	0.861	0.855	0.859
Father ed	0.852	0.822	0.84	0.85	0.847	0.848	Father ed	0.859	0.827	0.849	0.852	0.845	0.856
Books	0.594	0.595	0.662	0.683	0.653	0.644	Books	0.678	0.646	0.687	0.706	0.692	0.735
% of Var	60.26	56.89	61.4	64.7	62.98	62.67	% of Var	65.06	60.11	64.27	65.5	64.16	67.04
ARM							MYS						
Mother ed	NA	0.835	0.837	0.83	0.882	NA	Mother ed	0.862	0.875	0.864	0.873	0.885	0.88
Father ed	NA	0.839	0.842	0.838	0.879	NA	Father ed	0.877	0.881	0.872	0.874	0.881	0.881
Books	NA	0.685	0.697	0.687	0.623	NA	Books	0.644	0.667	0.634	0.632	0.654	0.639
% of Var	NA	62.32	63.13	61.98	64.58	NA	% of Var	64.23	66.22	63.57	64.15	66.25	65.83
BHR							MOR						
Mother ed	NA	0.84	0.854	0.853	0.858	0.87	Mother ed	0.824	0.828	0.812	0.852	0.832	0.846
Father ed	NA	0.859	0.866	0.861	0.869	0.865	Father ed	0.852	0.84	0.844	0.867	0.846	0.845
Books	NA	0.579	0.605	0.58	0.597	0.537	Books	0.647	0.662	0.664	0.661	0.652	0.667
% of Var	NA	59.3	61.52	60.19	61.55	59.74	% of Var	60.8	60.95	60.44	63.83	61.1	62.48
BWA							NZE						
Mother ed	NA	0.875	0.845	0.855	0.866	NA	Mother ed	0.854	0.843	NA	0.852	0.85	0.868
Father ed	NA	0.849	0.851	0.843	0.856	NA	Father ed	0.856	0.849	NA	0.855	0.85	0.854
Books	NA	0.684	0.594	0.571	0.58	NA	Books	0.588	0.587	NA	0.592	0.629	0.659
% of Var	NA	65.11	59.76	58.94	60.66	NA	% of Var	59.92	59.17	NA	60.26	61.35	63.9
CHL							NOR						
Mother ed	0.878	0.879	NA	0.877	0.876	0.87	Mother ed	NA	0.859	0.863	0.864	0.856	-856
Father ed	0.882	0.883	NA	0.876	0.878	0.872	Father ed	NA	0.837	0.847	0.838	0.837	0.83
Books	0.714	0.765	NA	0.696	0.657	0.639	Books	NA	0.617	0.628	0.641	0.68	0.671
% of Var	68.63	71.22	NA	67.34	65.63	64.18	% of Var	NA	60.64	61.87	61.96	63.17	62.37
CTPE							ONT						
Mother ed	0.845	0.858	0.865	0.858	0.875	0.87	Mother ed	NA	0.837	0.848	0.842	0.861	0.862

Father ed	0.839	0.867	0.86	0.874	0.859	0.861	Father ed	NA	0.848	0.841	0.842	0.86	0.847
Books	0.696	0.658	0.681	0.699	0.714	0.719	Books	NA	0.608	0.583	0.568	0.592	0.561
% of Var	63.38	64.06	65.06	66.31	66.1	67.16	% of Var	NA	59.62	58.9	58.06	61.06	59.14
ENG							QUE						
Mother ed	0.84	0.825	0.84	0.829	0.837	0.841	Mother ed	NA	0.821	0.823	0.819	0.835	0.847
Father ed	0.821	0.819	0.832	0.822	0.833	0.828	Father ed	NA	0.81	0.832	0.812	0.831	0.838
Books	0.613	0.658	0.681	0.662	0.672	0.648	Books	NA	0.638	0.647	0.666	0.661	0.597
% of Var	58.49	59.5	62.09	60.04	61.55	60.43	% of Var	NA	57.9	59.56	59.08	60.84	59.24
HK							ROM	0.883	0.881	0.882	0.888	NA	0.915
Mother ed	0.839	0.847	0.851	0.86	0.869	0.874	Mother ed	0.892	0.885	0.877	0.882	NA	0.905
Father ed	0.843	0.837	0.86	0.867	0.867	0.876	Father ed	0.739	0.715	0.7	0.741	NA	0.733
Books	0.595	0.65	0.664	0.687	0.672	0.682	Books	70.68	68.97	67.89	70.53	NA	73.09
% of Var	58.98	61.34	63.52	65.43	65.26	66.6	% of Var						
HUN							RUS						
Mother ed	0.869	0.877	0.858	0.873	0.883	0.874	Mother ed	0.823	0.827	0.818	0.83	0.849	0.854
Father ed	0.86	0.857	0.842	0.857	0.87	0.861	Father ed	0.825	0.811	0.817	0.826	0.842	0.85
Books	0.731	0.735	0.746	0.756	0.798	0.768	Books	0.697	0.694	0.653	0.69	0.612	0.653
% of Var	67.58	68.09	66.67	68.97	72.4	69.89	% of Var	61.45	60.74	58.75	61.54	60.17	62.65
IND							SGP						
Mother ed	0.89	0.9	0.89	0.89	NA	NA	Mother ed	0.858	0.852	0.861	0.854	0.855	0.87
Father ed	0.89	0.89	0.88	0.89	NA	NA	Father ed	0.85	0.854	0.866	0.868	0.862	0.872
Books	0.454	0.423	0.552	0.498	NA	NA	Books	0.565	0.594	0.632	0.642	0.617	0.627
% of Var	59.7	60	62.49	61.3	NA	NA	% of Var	59.26	60.27	63.03	63.2	61.8	63.66
IRN							SVN						
Mother ed	0.857	0.866	0.87	0.873	0.87	0.877	Mother ed	0.862	0.865	0.864	0.88	0.871	NA
Father ed	0.881	0.874	0.88	0.888	0.885	0.882	Father ed	0.839	0.835	0.855	0.852	0.848	NA
Books	0.708	0.719	0.773	0.734	0.723	0.678	Books	0.697	0.677	0.651	0.668	0.688	NA
% of Var	67.1	67.7	71.42	69.64	68.75	66.9	% of Var	64.43	63.43	63.37	62.42	65.04	NA
ISR							SWE						
Mother ed	0.873	0.85	0.863	0.868	0.871	0.878	Mother ed	NA	0.839	0.846	0.846	0.845	0.847
Father ed	0.861	0.856	0.863	0.868	0.868	0.875	Father ed	NA	0.817	0.836	0.831	0.83	0.841
Books	0.62	0.604	0.646	0.645	0.651	0.616	Books	NA	0.674	0.644	0.624	0.682	0.67
% of Var	62.9	60.61	63.59	64.09	64.55	63.85	% of Var	NA	60.82	61	59.81	62.23	62.49
ITA							THL						
Mother ed	0.845	0.847	0.851	0.837	0.849	0.868	Mother ed	0.902	NA	0.892	0.896	0.888	NA
Father ed	0.843	0.83	0.838	0.842	0.858	0.853	Father ed	0.907	NA	0.9	0.895	0.891	NA
Books	0.662	0.645	0.709	0.683	0.702	0.687	Books	0.624	NA	0.713	0.705	0.697	NA
% of Var	62.11	60.8	64.29	62.53	64.71	65.06	% of Var	67.54	NA	70.5	70.07	68.97	NA
JPN							TUN						
Mother ed	0.833	0.838	0.822	0.826	0.832	0.833	Mother ed	0.843	0.856	0.861	0.867	NA	NA
Father ed	0.844	0.839	0.846	0.839	0.838	0.829	Father ed	0.852	0.863	0.862	0.863	NA	NA

Books	0.548	0.548	0.582	0.538	0.572	0.553	Books	0.703	0.714	0.747	0.69	NA	NA
% of Var	56.91	56.89	57.66	55.87	57.33	56.2	% of Var	64.33	66.29	68.06	65.76	NA	NA
JOR							TUR						
Mother ed	0.813	0.826	0.806	0.806	0.818	0.816	Mother ed	0.797	NA	0.845	0.832	0.833	0.849
Father ed	0.854	0.849	0.846	0.837	0.838	0.834	Father ed	0.843	NA	0.87	0.852	0.853	0.858
Books	0.626	0.582	0.611	0.593	0.569	0.555	Books	0.658	NA	0.724	0.682	0.719	0.722
% of Var	59.43	58.12	57.97	56.75	56.39	55.67	% of Var	59.3	NA	66.48	62.76	64.61	65.93
KOR							USA						
Mother ed	0.868	0.841	0.851	0.852	0.852	0.86	Mother ed	0.848	0.834	0.849	0.848	0.85	0.866
Father ed	0.884	0.856	0.86	0.859	0.86	0.853	Father ed	0.852	0.853	0.858	0.858	0.852	0.864
Books	0.596	0.615	0.611	0.657	0.611	0.625	Books	0.69	0.678	0.694	0.699	0.687	0.681
% of Var	62.98	60.66	61.24	63.2	61.33	61.91	% of Var	64.08	62.75	64.58	64.87	63.97	65.32

Appendix D

Table D1 Within-country linear regression model estimates of cross-school turnover intentions

	DV: Proportion of teachers wishing to change schools	
	Coeff.	SE
AUS	.051-	.029
AUT	.025**	.005
BEL	.034***	.009
BRA	.017-	.009
BGR	.050***	.017
CHL	.020	.012
CTPE	.065***	.017
COL	-0.017	.026
HRV	.039-	.020
CYP	.026	.033
CZE	.028-	.017
DEN	.048	.050
EST	.052**	.017
FIN	.009	.014
FRA	.064***	.018
GEO	-.007	.009
HUN	.035**	.012
ISR	.040**	.016
ITA	.026-	.015
JPN	.037**	.018
KAZ	.002	.013
KOR	.006	.038
LVA	.024	.028
LTU	.052*	.021
MLT	.013	.020
MEX	.008	.019
NLD	.005	.014
NZE	.043	.073
NOR	.034-	.018

PRT	.008	.016
RUS	-.007	.029
SAU	.053**	.017
SGP	.011	.022
SVK	.030*	.014
VNM	.029	.020
SVN	.005	.020
ZAF	.022	.030
ESP	.017	.013
SWE	.005	.023
UAE	.033***	.009
TUR	.0009	.023
USA	.033	.046
ENG	.025-	.013
CAN	.067-	.040
ROM	.042***	.012
ARG	.032**	.010
CHN	.011	.016

*** = $p < .000$, ** = $p < .01$, * = $p < .05$, - = $p < .10$.

Note. Standard errors clustered by school. Estimates weighted by the school weights provided by TALIS. Analysis controls for school location and type.

Table D2 2-level random slope models

	DV: Proportion of teachers wishing to change schools	
	Coeff.	SE
School location	0.015	0.009
School type	0.023	0.020
HDI	-0.513**	0.084
Accountability (External)	-0.294**	0.019
Accountability (SMT)	-0.032	0.019
Socioeconomic SES composition × Account (Exte)	0.097***	0.023
Socioeconomic SES composition × Account (SMT)	-0.033	0.057
Socioeconomic SES composition × HDI	0.0142**	0.052

N schools	8814
N system	46

*** = $p < .000$, ** = $p < .01$, * = $p < .05$.

Note. Controls include school location and school type. Senate school weights applied.

Studies I – IV

Study I.
Glassow, L.N., & Jerrim, J. (2022). Is inequitable teacher sorting on the rise? Cross-national evidence from 20 years of TIMSS. *Large-scale Assessments in Education, 10,* doi: 10.1186/s40536-022-00125-9

Study II.
Glassow, L.N., Yang Hansen, K., & Gustafsson, J.E. (under revision). Does socioeconomic sorting of teacher qualifications exacerbate mathematics achievement inequity? Panel data estimates from 20 years of TIMSS.

Study III.
Glassow, L.N., Franck, E., & Yang Hansen, K. (under revision). Institutional characteristics moderating the relationship between classroom socioeconomic composition and teacher qualifications: Evidence from 46 education systems in TALIS 2018.

Study IV.
Glassow, L.N. (under revision). Inequitable teacher turnover and performance-based appraisal: A global trend?

The allocation of teacher competence is frequently cited as a determinant of socioeconomic inequity in student outcomes. This thesis explores several aspects related to the phenomenon of 'teacher sorting' from a cross-national perspective. It aims to empirically validate theoretical questions regarding social reproduction and the allocation of educational resources, as well as whether inequitable allocation contributes to socioeconomic gaps in student performance.

The thesis presents four empirical studies with two pairs of related research questions. Study I examined the magnitude of teacher sorting across education systems and over the past 20 years. Study II extends this line of questioning and examines whether the sorting gap magnitudes exacerbate mathematics achievement inequity within countries. Studies III and IV examined whether institution- level policies are associated with several different dimensions of teacher sorting. The findings lend partial support to the role of inequitable teacher competence in perpetuating inequitable student outcomes, but also point to the importance of the socioeconomic context and construct measurement in considering inequity and policy across education systems.

Leah Natasha Glassow holds a M.Sc and a M.Ed in Education. Her research interests include quasi-experimental techniques, educational policy evaluation, and teacher sorting.

ISBN 978-91-7963-127-7
9789179631277