Large-scale Assessments in Education

Is inequitable teacher sorting on the rise? Cross-national evidence from 20 years of TIMSS

Leah Natasha Glassow[1*] and John Jerrim[2]

*Correspondence:
leah.glassow-hill@gu.se

[1] Department of Education and Special Education, University of Gothenburg, Gothenburg, Sweden
[2] Department of Social Science, University College London Institute of Education, London, UK

Abstract

Unequal access to qualified teachers for children of different socioeconomic status—also known as inequitable teacher sorting—has been increasingly put forth as one potential factor contributing to the socioeconomic achievement gap. Despite this, few studies have investigated cross-national differences in teacher sorting, and none have examined it within-countries over time. International large-scale assessments in education are uniquely positioned to answer such questions due to their longitudinal nature at the system level. This study uses six waves of data from the Trends in International Mathematics and Science Study (TIMSS) from 1999 to 2019 for 32 education systems. We compare differences in grade 8 mathematics teacher qualifications for each country at each time point, across top and bottom groups on the student socioeconomic spectrum. Results show that on the whole many countries display negligible gaps in access to teacher quality, with some key exceptions. With respect to inequity in novice teacher sorting, the problem is most prevalent in low- and middle- income education systems (i.e. in Turkey, Morocco, Tunisia and Indonesia). Inequity in sorting based on mathematics education is less common, with no clear pattern in regards to level of economic development (i.e. in Chile, Australia, New Zealand, and Chinese Taipei). Socio-economic inequality in teacher sorting has also remained broadly stable over time. Based on experience and mathematics education, less than a handful of systems show systematic upward trends in teacher sorting inequity (i.e. in Chile, Morocco, Singapore, and New Zealand). Given the increasing focus on inequity in access to teacher competence, these results have economic and policy implications for tackling the socioeconomic achievement gap.

Keywords: TIMSS, Teacher sorting, Educational inequity, Achievement gap

Introduction

Since the publication of the Coleman report (1966), rather than aiming for equality of outcomes in education, policy makers have focused their efforts towards minimizing such inequities through educational opportunities (i.e. inputs) susceptible to policy amendment. However, despite decades of effort devoted to promoting educational equity, the socioeconomic status (SES) related achievement gap persists (Broer et al., 2019). Recent work has also shown that globally, the SES-achievement gap is widening (Chmielewski, 2019). Minimizing the performance gap between high- and low-SES students through improvements from the lowest performing students yields economic

growth and social mobility (Chetty et al., 2014; Hanushek & Woessman, 2015; Solga, 2014). Among the proposed explanations for the persistence of the SES achievement gap is differential access to teacher competence, which some argue has not been adequately investigated (Sims & Allen, 2018). A growing number of internationally-focused studies have raised the alarm about this issue (Akiba et al., 2007; Han, 2018; Luschei & Jeong, 2019; OECD, 2018).

The effects of socioeconomic inequity are apparent at birth- and it is becoming ever clearer that such effects are cumulative (Kratz et al., 2022; Walsh et al., 2019). Instead of the role of 'great equalizer' (Downey & von Hippel, 2004), education systems are at risk of compounding the effect of familial SES on achievement with differential access to educational resources. Determinants of inequity at the system level include within- and across- school ability grouping, school competition, school autonomy, accountability practices, school composition and segregation, public expenditure on education, class size, and the segmentation of teaching and teacher qualifications (Strietholt et al., 2019). Class size reductions are one of the most debated educational policy reforms of the past several decades, and have yielded mixed results as smaller classes do not necessarily guarantee better teaching (Shen & Konstantopoulos, 2019). For instance, Hanushek et al. (2005) show that achievement gains from a 'costly' ten-student reduction in class size were smaller than moving one standard deviation up the teacher quality distribution. Much of the discussion has therefore turned to focus on the role of teacher quality.

In this study, we aim to provide an up-to-date descriptive account of the phenomenon and evolution of inequitable teacher sorting, or the teacher 'opportunity gap' as it has been previously named (Akiba et al., 2007), which we term 'teacher sorting' or the 'teacher qualification gap.' Our aim is to compare gap magnitudes across countries and to determine whether inequities have grown over time. We use evidence from six waves of the Trends in International Mathematics and Science Study (TIMSS) from 1999 to 2019 and examine data from 32 education systems.

Inequitable teacher sorting: evidence from past research

The phenomenon of teacher sorting has generally been attributed to teachers self-selecting into more preferable working conditions coupled with labor market conditions (Engel et al., 2014). Accounting for sorting inequity in education systems where the allocation of teachers is more centralized is less straightforward, but may be due to increasing socioeconomic segregation, accountability measures, or ability grouping within schools (Han, 2013; Luschei & Jeong, 2019). In most cases, however, the first part the problem is *attracting* the best teachers to schools or classrooms with a high proportion of disadvantaged students. This type of sorting has generally been measured at the national level, by examining differences in average qualification levels of teachers between high- and low-SES students, schools or classrooms. In the US, Goldhaber et al. (2015) find evidence of inequities in teacher qualifications across schools in Washington for all input measures. This has been confirmed by studies in New York, North Carolina and California, and across racial disparities as well (Clotfelter et al., 2005, 2007; Goldhaber, 2018; Lankford et al., 2002). In England, Sims and Allen (2018) find evidence of inequities for teacher experience, certification, and subject matter specialization. This may in part be due to teacher motivation (novice teachers tend to be more ideologically

motivated to 'make a difference'), coupled with the fact that there are generally more job opportunities for novice teachers in disadvantaged schools (Sims & Allen, 2018). National studies from Turkey and Chile also show evidence of sorting (Meckes & Bascope, 2015; Özoğlu, 2015).

The next part of the problem relates to *keeping* teachers in disadvantaged settings. Here, the role of salary incentives has mixed evidence (Clotfelter et al., 2008; Sims & Allen, 2018). Teachers may prioritize working conditions over salary (Bacolod, 2007), but this is difficult to generalize, as there is variation in teacher salaries and working conditions worldwide. Teachers that are more competent may exit the profession or certain schools due to increased accountability practices (Feng et al., 2010). We must also consider the ongoing debate over whether teacher sorting occurs more frequently between- or within-schools, but so far the evidence is mixed (Hanushek & Rivkin, 2012; Kalogrides et al., 2013; Luschei & Jeong, 2019). This phenomenon cannot be investigated using the TIMSS dataset, as there is frequently just one classroom sampled per school.

Up until now we have discussed nationally-focused studies investigating inequitable teacher sorting cross-sectionally. A handful of national studies from the US have examined teacher quality gaps over time and show either stable or decreasing trends (Adamson & Darling-Hammond, 2011; Boyd et al., 2008; DeAngelis et al., 2010; Goldhaber et al., 2018). There are also few studies that investigate sorting from an international perspective, though none take this perspective over time. Akiba et al. (2007) report differences between high- and low-SES students, particularly for students in Syria, Chile, Taiwan, the United States and Hong Kong using TIMSS data from 2003. Using the Programme for International Student Achievement (PISA) data, Han (2018) finds that school SES is associated with teacher shortages as well as the proportion of certified teachers in a school. Using data from the Teaching and Learning International Survey (2013), Luschei and Jeong (2019) examine whether the phenomenon of teacher sorting is global, and find evidence of either within- or across-school teacher sorting in several of 32 participating systems.

Teacher quality

Underpinning our discussion of the teacher qualification gap is the ongoing debate regarding the measurement of teacher quality and competence. Teachers are consistently found to be one of the most important school-level inputs for achievement, particularly for disadvantaged students (Darling-Hammond, 2000; Goe, 2007; Nilsen & Gustafsson, 2016; Wayne & Youngs, 2003). However, despite decades of research, there remains debate as to which teacher characteristics matter most, especially from a cross-national perspective (Blömeke & Delaney, 2012; Blömeke & Olsen, 2019). There is limited evidence supporting teacher personality traits and characteristics such as self-efficacy and preparedness for test scores (Goe, 2007; Nilsen & Gustafsson, 2016). However, these teacher characteristics may be important for behavioural outcomes associated with long-term success. Chetty et al. (2014) find that being assigned to higher value-added teachers have long term impacts for students' earnings and higher education graduation rates. Using national data from the US, Jackson (2018) finds that the effects of teacher characteristics are ten times larger for behavioural outcomes such as intention to attend college than on test scores. Other scholars, however, argue that observable teacher

characteristics have failed to account for significant variation in student achievement and should be rejected altogether in favor of value-added models (Hanushek & Rivkin, 2012). Value-added models are not a panacea, however, as evidenced by a recent paper by Bitler et al. (2014). Perhaps more importantly, they are limited by the differential assignment of teachers to students based on ability and in many cases cannot identify which aspects of a teacher are most important. International large-scale assessments do not allow researchers to create value-added measures of teacher effectiveness. Additionally, putting value-added models aside, amenable teacher characteristics such as self-efficacy and preparedness (Goe, 2007) are likely to be influenced by teaching contexts, and therefore present an endogeneity problem when attempting to determine whether students in disadvantaged contexts receive less qualified teachers because of their socio-economic background.

Due to the number of empirical issues raised above, we move our focus towards teacher qualifications. While certain teacher qualifications have been shown to be the most stable and consistent predictors of student achievement, they are also the subject of debate. The first and most frequently cited qualification is teacher experience, which has been shown to matter predominantly in the first three to five years, while recent work has shown its potential importance much longer into a teacher's career (Goe, 2007; Papay & Kraft, 2015; Podolsky et al., 2019; Rice, 2003; Rockoff, 2004). We therefore construct the category of 'novice' teachers, which includes all teachers in their first five years on the job. We use five years to ensure maximum statistical power and minimize sampling error in order to increase the representativeness of teachers in TIMSS dataset (in contrast to using one, two, or three years of experience for novice designation, for instance). The next important qualification includes teacher mathematics education. Several studies have confirmed the importance of teacher a teacher's knowledge in mathematics and science (Baumert et al., 2010; Goe, 2007; Sancassani, 2021; Hanushek et al., 2014). Such qualifications may also predict teaching practices such as cognitive activation and classroom management (Nilsen et al., 2018). There is a theoretical distinction (and plausible differences in teacher effects) to be made regarding some teachers having pedagogical training in mathematics and those who have a post-secondary degree in mathematics with no teacher training. These are referred to pedagogical content knowledge (PCK) and content knowledge (CK), respectively (Shulman, 1986). We create the category of 'out-of-subject' teacher for the primary reason that it is the most validly comparable category across education systems due to differences in requirements and curricula for teaching degrees, thus limiting our focus to teachers with no mathematics education (in other words, without CK or PCK). While there is limited evidence for the importance of teacher certification and level of education (Goe, 2007; Goldhaber et al., 2000) we exclude the former due to its absence from the TIMSS dataset, and the latter as the variation in teachers with and without a bachelor (or masters) degree would be little in a vast majority of education systems. Last, there is generally a consensus that teacher competence matters most for mathematics and science (Goe, 2007). This is due to the fact that while students may practice reading or language skills at home, learning in mathematics and science is generally confined to the classroom. For this reason, we focus on mathematics teachers only. The vast majority of studies related to teacher effectiveness focus on upper middle or secondary school, but only a handful of studies focus

on differential importance of teacher competence at various grade levels (Goe, 2007; Nilsen & Gustafsson, 2016). Teacher qualifications in particular seem to matter most in the later grades (Goe, 2007).

Against this background, we focus our efforts on two observable teacher characteristics which are not amenable to the teaching context, and assume that teachers in their first five years on the job ('novice teachers') or without training in mathematics or mathematics education ('out-of-subject') will have lower competence levels. We use these categories to answer the following research questions:

1. In which education systems do a significantly higher share of socioeconomically disadvantaged students have teachers with lower qualification levels based their most recent TIMSS cycle?
2. How has this phenomenon of 'inequitable teacher sorting' changed over time across and within education systems?

Methods

The data for this study come from six waves of the Trends in International Mathematics and Science Study (TIMSS) carried out by the International Association for the Evaluation of Educational Achievement (IEA) from 1999, 2003, 2007, 2011, 2015 and 2019.[1] We take data from a total of 32 countries participating in at least three of these six time points, amounting to a total sample size of 904,309 students and 36,446 mathematics teachers for grade 8. 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.[2]

Teacher- level variables

Experience

Teacher experience was measured by the number of years of a teacher has been teaching altogether. As mentioned, we create the category of 'novice' teachers (those in their first 5 years after qualification).[3] Here, we emphasize the evidence of a 'learning by doing' effect, in which teacher effectiveness improves substantially the first few years on the job. Based on past research, we argue that when taught by novice teachers, students are likely to retain and accrue less knowledge (Podolsky et al., 2019; Rice, 2003; Rivkin et al., 2005; Rockoff, 2004).

---

[1] We exclude data from 1995 due to the inconsistency of certain variables related to teacher characteristics (mathematics education) and student socioeconomic status (parental education).

[2] 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. If countries do not have adequate participation rates, they can be excluded from the results (see Table B in Appendix for sample sizes and participation rates for each country). Most countries (Hong Kong, Morocco, Israel, USA) remain in the high 70 or 80 percent rage, with just England at 54% in 2003. We therefore recommend caution around interpreting estimates from these countries.

[3] We check whether teacher qualification gaps are dependent on how 'novice' was coded by comparing the gaps with those based on novice teachers as defined as having up to 1 year of experience and having up to 3 years of experience. The teacher qualification gaps were in the vast majority unchanged.

Mathematics education

The category 'out-of-subject' teachers was determined by which subject teachers reported to study during their post-secondary education. There are vast differences in the requirements to become a math teacher across countries. Teachers citing 'mathematics' or 'mathematics education' or both thus may have some level of formal training in mathematics (see Shulman, 1986; Blömeke & Olsen, 2019) depending on the requirements for teaching across educational systems. We therefore designate a teacher as 'out-of-subject' if they have none of these. Several studies have mentioned the importance of CK and PCK particularly for mathematics (Baumert et al., 2010; Goe, 2007; Shulman, 1986; Wayne & Youngs, 2003). Some of the strongest evidence comes from Baumert et al (2010) for mathematics, and Sancassani (2021) for science, who employs a within student approach and finds a significant causal effect for teachers with CK on student achievement.

While some studies propose a composite measure of teacher quality, our analysis did not provide evidence in support of such a measure, as factors loaded negatively onto the construct in some education systems and positively in others. In our view, these measurement properties limit the usefulness of a composite approach, particularly with regard to their relevance for policymakers.

Student-level variables

Number of books in the home

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. It is often described as a proxy for 'cultural capital' and generally one of the strongest predictors of student achievement (Hanushek & Woessman, 2011). While this variable is widely used in large-scale assessment research, some have questioned its validity. Some of the strongest criticisms come from Engzell (2019), who alert researchers to a number of endogeneity issues; namely, that children who are more studious tend to accrue more books, and that low achieving students are more likely to misreport lower numbers of books. There is also a lower level of agreement between parent and child reporting on this indicator in particular for students tested at ten years old (though at grade 8 level this risk is much lower) (Jerrim & Micklewright, 2014). Nevertheless, as cultural capital is an important part of our construction of socioeconomic status, we include it within our pool of SES indicators, with these cautions in mind.

Parental education

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. We have made them into one category for all cycles in the study, reducing the categories to seven. Parental education generally has higher child-parent agreement than books in the home, but this too varies across countries (Jerrim & Micklewright, 2014). While the International Standard Classification of Education (ISCED) is a widespread measure of SES for economists and sociologists, it is also not without

complications related to cross-national comparability. For instance, the meaning of having a parent with a bachelor's degree in a Western OECD country may differ substantially from the meaning of having a parent with such a degree in a country with a lower level of economic development. While estimating SES scores within countries and years may mitigate these issues, they can never be entirely removed.

Socioeconomic status composite score
We create a composite socioeconomic measure (factor score) for each country-year comprised of student reported number of books in the home, student reported mother's level of education and student reported father's level of education in a confirmatory factor analysis. Measuring socioeconomic status through a composite score is a widely used approach to measuring socioeconomic status construct, but there is some debate as to the validity of combining the indicators. Ideally, such a composite SES indicator should reflect a family's general level of education, income and occupation. TIMSS does not include information about parental occupation or income across all cycles, but includes a 'Home Educational Resources' (HER) scale comprised of home possessions, parental education, and the number of books in the home, which is available in later cycles but not all of them.[4] We do not use home possessions in our scale as they varied over each of the cycles and there is considerable debate about the comparability of this measure across countries and time (Pokropek et al., 2017).[5]

Top/bottom SES percentiles
Following Chmielewski (2019), we employ the 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). We take students who score below the 33rd and above the 66th percentile of the SES scale to represent those in low- and high-socioeconomic status families. We use thirds instead of quartiles so as to ensure maximum statistical power and minimum sampling error for each country-year estimate. Over and above examining gaps from the most recent TIMSS cycle, we also examine pooled teacher qualification gaps at 33/66 and 10/90 percentiles.

Analysis
In order for TIMSS data to be country representative, inferences must be made at the student level (Rutkowski et al., 2010). TIMSS is also one of the few international large-scale assessments in which teachers can be directly linked to students. We use multivariate imputation by chained equation using the 'mice' package (van Buuren & Groothuis-Oudshoorn, 2011) in RStudio to account for missing data in the student socioeconomic variables. We include mathematics/science achievement (plausible values) and student possessions as auxiliary variables in the imputation model, and create five imputed datasets for each country-year, resulting in a total of 5 imputed datasets × 189

---

[4] However, TIMSS calculates SES scores with *all* countries in the model, while we calculate SES scores for students within country-years.

[5] We do use home possessions to predict missing values in our imputation model for each country-year, however.

country-years $= 945$ imputed datasets.[6] To determine the final teacher estimates from each imputation, we average estimates from Eqs. (1) and (2) below across the five imputed datasets and pool the standard errors following Rubin (1987) and Gonzalez (2014).

In order to estimate the socioeconomic status scores for each student and for each imputed dataset we conduct a factor analysis with polychoric correlations to estimate factor loadings and extract the single factor scores. This has been done using the function 'fa' in the RStudio 'psych' package. Next, we group students into top and bottom thirds and deciles within each country on this scale to determine their relative SES position, applying the student weights.

To determine the teacher qualification gaps for each imputed dataset, we estimate the proportion of students in high- and low- socioeconomic contexts with novice and out-of-subject teachers in each year and each country, applying the mathematics teacher weights to account for students with more than one mathematics teacher. The formula is as follows:

$$TQD = \frac{1}{\sum_l \sum_t TQD_{lt}} \sum_{l=1}^{32} \sum_{t=1}^{6} \tag{1}$$

$$TQAD = \frac{1}{\sum_l \sum_t TQAD_{lt}} \sum_{l=1}^{32} \sum_{t=1}^{6} \tag{2}$$

where: $TQ_{D/AD}$: the proportion of novice or out-of-subject teachers for socioeconomically disadvantaged (D) or advantaged (AD) students, $l =$ country, $t =$ year.

Both teacher quality estimates $TQ_{(D)}$ and $TQ_{(AD)}$ equate to the percentage of students with novice and out-of-subject teachers in country l at year t and is an estimate between 0 and 1. We do this for each year, SES group (33/66 and 90/10) and for each teacher qualification indicator, clustering the standard errors by school using the 'ClubSandwich' (Pustejovsky, 2020) package with the heteroscedasticity-consistent 'CR0' variance estimator in Rstudio.[7] To determine the teacher qualification gap, we calculate $TQ_D - TQ_{AD}$.[8] For 33/66 gaps, we include point estimates for each country-year in Additional file 1: Appendix Table C1 and C2.

To examine linear trends in teacher sorting over time, we again use the 'ClubSandwich' (Pustejovsky, 2020) package in RStudio to estimate the standard errors with the same variance estimator 'CR0'. We predict positive or negative changes in teacher sorting with the following logistic model:

$$l_{ig} = \log \frac{p}{1-p} = \beta_0 + \beta_1 x_{1ig} + \beta_2 t + \beta_3 x_{1ig} t \tag{3}$$

---

[6] As the 'mice' package allows for multivariate imputation, continuous variables are imputed using linear regressions, dichotomous variables are imputed using logistic regressions, and so on.

[7] 'CR0' accounts for the stratified two- stage sampling structure of TIMSS data.

[8] For 33/66 gaps, we include point estimates for each country-year in the appendix Table C1 and C2. Due to significantly reduced sample size for 90/10 gaps, we calculate only the pooled estimates.

where l is the log odds that a teacher is new/out-of-subject for student i in school g, x_1 is the low-SES term, t is time,[9] and x_1t their interaction term indicating change over time. β_3 is therefore our coefficient of interest. As the results are intended to be descriptive and examine the association between teacher qualifications and student socioeconomic contexts, we do not use additional controls in the model. Furthermore, we restrict the analyses to linear trend lines. Due to the risk of sampling error and the relatively small number of time points, it is not possible to determine with accuracy non-linear trends.[10] Here, we follow the same imputation approach as outlined above. Estimates are reported as odds ratios (the exponents of the unstandardized beta coefficients). Here, an odds ratio of 1.05 denotes that a one unit increase (time in years between TIMSS cycles) is associated with a 5% increase in the odds that a low SES student will receive a teacher with lower qualifications.

Results

Estimating the magnitude of inequitable teacher sorting across countries

For descriptive statistics and missing data proportions, see Tables A and B in the Additional file 1: Appendix. In order to investigate countries in terms of the extent of inequitable teacher sorting over the past two decades (i.e. an average across the past 6 TIMSS cycles), we use data from each wave of TIMSS to determine a pooled estimate for each country. We do this for both teacher quality indicators and student socioeconomic groups.[11] Please refer to Figures A1 and A2 in the Additional file 1: Appendix for a visual depiction of the pooled gaps. We now focus our attention on the extent of inequity in teacher sorting in the most recent TIMSS cycles of participating education systems.

Figure 1 displays estimates for students in the 33rd and 66th socioeconomic tertiles in terms of their exposure rates to novice and out-of-subject teachers from the most recent participating TIMSS cycle. For countries showing positive gap magnitudes, a greater share of low-SES students have teachers with lower qualifications than their high-SES counterparts. For countries showing negative gap magnitudes, a greater share of high-SES students are exposed to teachers with lower qualification levels. We report confidence intervals for these estimates for both groups based on the standard errors adjusted for school clustering. Gap magnitudes—the proportion of socioeconomically more affluent students with under-qualified teachers subtracted from the proportion of socioeconomically disadvantaged students—are depicted by the grey bars.

Before discussing the details of these plots, it is important to highlight an immediate conclusion from these figures. For both sorting by experience and mathematics education, the teacher qualification gaps do not reach statistical significance in a majority of countries. While novice teacher sorting is clearly more prevalent, less than half of the participating education systems display gap magnitudes over 5 percentage points. By and large, this indicates that many countries are successful in their allocation of teachers by experience and subject matter education in mathematics,

---

[9] The time variable is defined as follows:$1999 - 0$; $2003 - 4$; $2007 - 8$; $2011 - 12$; $2015 - 16$; $2019 - 20$.

[10] We include odds ratios for only the 33/66 group due to limited sample size in the 90/10 groups for each system. However, we plot the changes over time for both groups in Figs. 3 and 4.

[11] Point estimates for 33/66 gaps for each country-year may be found in Table C1 and C2 in the appendix, and point estimates for pooled 90/10 gaps may be found in Table D in the appendix.
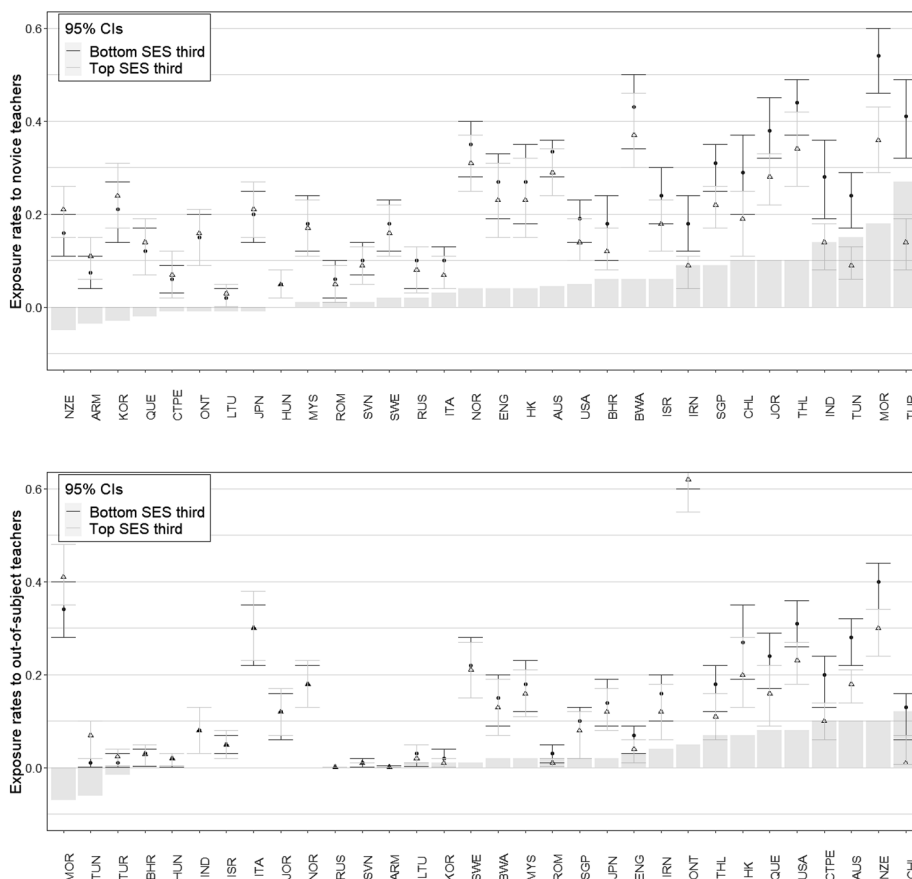
Fig. 1 Student exposure rates to novice and out-of-subject teachers from the most recent TIMSS cycle (2011–2019). Education systems' most recent cycle is 2019 with the exception of Armenia (2015), Botswana (2015), Indonesia (2011), Slovenia (2015), Thailand (2015), and Tunisia (2011). Grey columns show teacher qualification gap magnitudes (the difference between low- and high-SES student exposure rates). n = 175,746

and that lower-SES students face a much more modest disadvantage with regards to their access to qualified teachers than the literature suggests. Referring to the pooled gaps in Figure A1 of the Additional file 1: Appendix, a similar pattern of countries is observed for sorting across both teacher qualification dimensions (i.e. Turkey, Tunisia, Morocco, Iran, Thailand for novice teacher sorting and Chile, New Zealand, Australia and Chinese Taipei for out-of-subject sorting). These figures give a better idea of the extent to which education systems have struggled with inequity in sorting over the past two decades. Figure A2 (in the Additional file 1: Appendix) shows the pooled qualification gaps between students at more distant ends on the socioeconomic spectrum. The exposure rates of these students show statistically significant inequities in many more education systems for novice teacher sorting (i.e. Singapore, Israel, USA, and Romania), but not for sorting by mathematics education.

To reiterate, Fig. 1 (panel 1) shows that while novice qualification gaps are positive in many countries, they are in most cases small and not statistically significant. This is in part due to the TIMSS sampling procedure and the very large standard errors produced

when adjusted for school-level clustering. Nevertheless, the median novice qualification gap across the countries is around 3 percentage points, meaning that just three percent more low-SES students have newly qualified teachers in comparison to the higher-SES group.[12] There is also considerable heterogeneity in terms of overall teacher qualification levels across the education systems. In Morocco, more than half of the students in the low-SES group are taught by novice teachers, compared to just five percent of students in Romania or Hungary. There are however some education systems for which novice teacher sorting is dramatic. Turkey, Tunisia, Morocco and Indonesia show very large gap magnitudes between 15 and 27 percentage points. Given the average returns of the first 5 years to teaching experience from past research (Rice, 2003; Rivkin et al., 2005; Rock-off, 2004), a 15 percentage point difference in exposure to novice teachers would yield a significant disadvantage in learning opportunities for lower-SES students.

Sorting by mathematics education is much less common across educational systems. The median out-of-subject qualification gap (Fig. 1, panel 2) is around 1 percentage point, also demonstrating that many countries do not show inequity in their allocation of these teachers. In addition, many countries have almost no out-of-subject mathematics teachers in the workforce, such as Russia, Lithuania, Slovenia, Romania, Korea, and Hungary. In Ontario, however, over 2 thirds of students have a mathematics teacher without training in mathematics or mathematics education. Figure 1 (panel 2) demonstrates that just Australia and New Zealand reach conventional levels of statistical significance, with Chile and Chinese Taipei showing larger qualification gap magnitudes as well. Our results do not imply that teacher sorting is not a problem in education systems which do not reach statistical significance. As we have pointed out, this is in many cases largely due to the uncertainty introduced by the TIMSS sampling structure. A teacher qualification gap magnitude of between 5 and 10 percentage points (as in Ontario, Thailand, Hong Kong, Quebec, and the USA) is still nevertheless noteworthy and given the evidence on subject matter education, likely to systematically disadvantage lower- SES students (Sancassani, 2021).

It is also worth calling attention to the heterogeneity in gap magnitudes across the teacher qualification dimensions. Save for Chile, not a single education system depicts inequities across both experience and subject matter education. This is in line with previous findings (Luschei & Jeong, 2019), and emphasizes that the determinants of sorting differ.

Estimating within-country trends of teacher sorting (1999–2019)

Turning our focus to the next research question, we plot the trends for countries participating in TIMSS in at least three time points between 1999 and 2019. In Fig. 2, we zoom in to two particular education systems—Chile and the Republic of Korea—and display the evolution of sorting by mathematics education. Dashed grey regression lines represent SES deciles, solid grey regression lines represent bottom SES deciles, and blue and red regression lines represent top and bottom SES tertiles, respectively. The vertical distance between the lines represents the teacher opportunity gaps for each given year. This up-close comparison between the two education systems highlights several features which we wish to emphasize. First, education systems may have a consistently low

---

[12] As mentioned, novice teacher qualification gaps based on other cut-offs did not change the cross-national pattern of results.
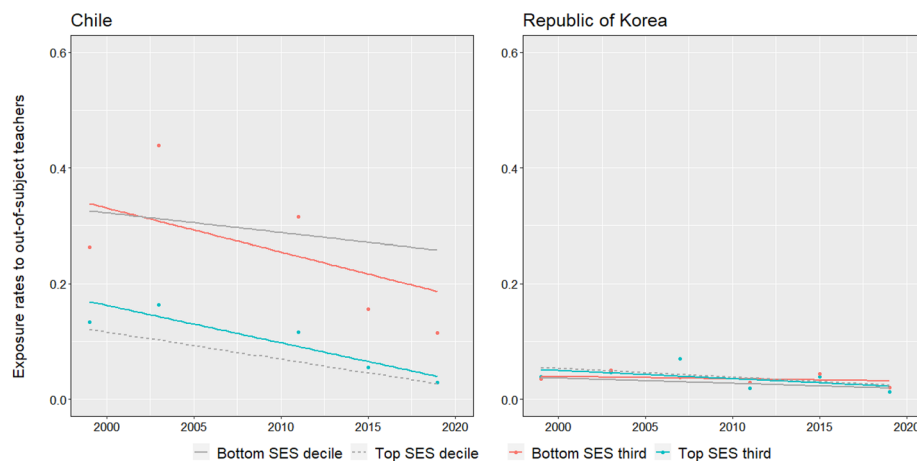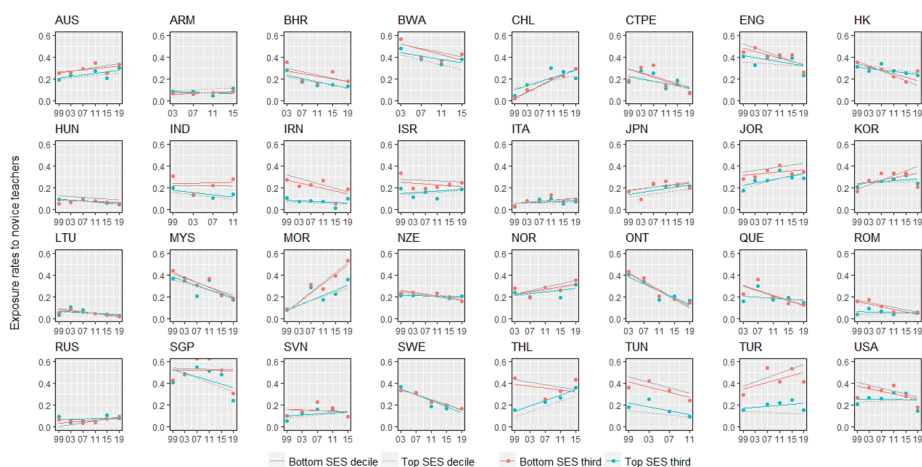
Fig. 2 20 year linear trends in out-of-subject teacher sorting for Chile and the Republic of Korea (1999–2019)

proportion of teachers with no mathematics education as in the case of South Korea. Since 1999, no more than around 5 percent of students (regardless of SES) have been allocated out-of-subject teachers. This demonstrates at once a high level of equity as well as a high overall qualification level of the South Korean teacher workforce. On the other hand, Chile shows greater variation across the years, with the overall exposure rates to out-of-subject teachers trending downwards since 1999. Despite this, Chile shows a consistently high (or even widening) gap magnitude across the TIMSS cycles.

Next, trends for all education systems are displayed in graphical form in Figs. 3 and 4. While many education systems vary from cycle to cycle in terms of the overall exposure rates to novice teachers, most show little or no variation over time in terms of the teacher qualification gap magnitudes. Certain educational systems show large but relatively stable gap magnitudes, such as Turkey, Tunisia, and Indonesia. A clear pattern of increasing gap magnitudes is displayed in Morocco and Singapore. In some cases, more subtle increases in inequity exist. For example, although Sweden points to an overall decline in novice teachers, the country shows increasingly pronounced sorting until 2015 when the trend starts to reverse. However, the largest qualification gaps in Sweden are still quite small and unlikely to reach statistical significance (at about 4 percentage points). As Sweden is among the most socioeconomically equitable societies in the world, this trajectory is nevertheless noteworthy and in line with reports of increasing school segregation in the country (Yang Hansen and Gustafsson, 2016).

Similar to the novice teacher sorting trends, almost no changes are displayed when we consider out-of-subject sorting. Australia and Quebec display stable gap magnitudes, with Quebec showing an overall decline in the proportion of students with recently qualified teachers. Again, few education systems appear to show a very clear upward trend in inequity (Chile, New Zealand and Chinese Taipei).

Taken together, these plots show that most educational systems do not show increasing signs of inequity in teacher allocation, even at the more extreme ends of the socioeconomic spectrum (90/10 deciles). In addition to this positive picture, in most cases, there do not appear to be increases in the overall share of students with teachers with lower qualification levels, regardless of their socioeconomic background. The proportion

Fig. 3 Estimated linear trends in novice teacher sorting, 1999–2019
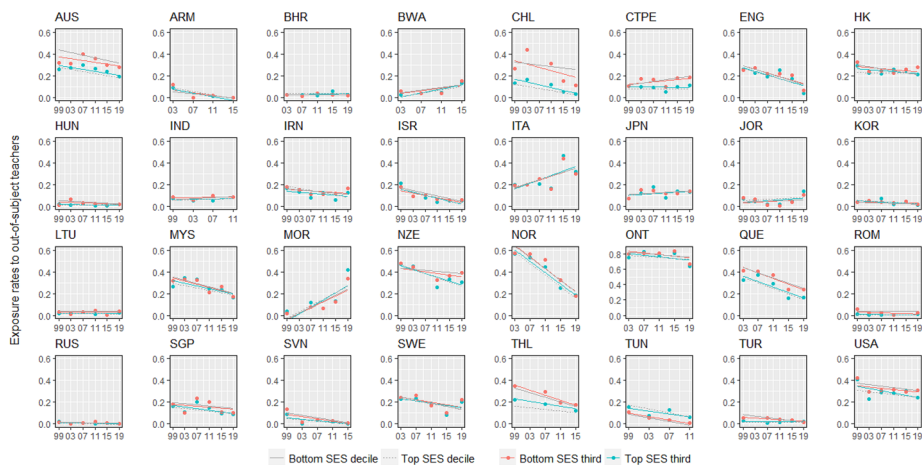
Fig. 4 Estimated linear trends in out-of-subject teacher sorting, 1999–2019

of students with newly qualified or out-of-subject teachers is not increasing across the countries included in our sample.

In a next step, we estimate trends in the above-plotted teacher qualification gaps to determine whether countries are displaying significantly growing or narrowing inequities in teacher allocation. Table 1 displays odds ratios for logit models regressing the likelihood of having a novice or out-of-subject teacher on the interaction between the low-SES student group and time as compared to the higher-SES student reference group (between each TIMSS cycle).

Here, we find more positive news to substantiate the aforementioned figures. Only a handful of countries show statistically significantly increasing opportunity gaps. For sorting by experience, just Chile, Morocco, and Singapore reach statistical significance at the 5% level.[13] In fact, more education systems show significant downward trends

---

[13] A handful of additional countries show statistically significant increases at the 10% significance level.

Table 1 Estimated linear trends for novice and out-of-subject teacher sorting by education system (1999–2019)

	Logistic regression model				N	T
	1. Novice teachers		2. Out-of-subject teachers			
	OR	SE	OR	SE		
Australia	1	(.01)	1	(.01)	32,814	6
Bahrain	1.02+	(.01)	.99	(.03)	15,823	5
Armenia	1.04	(.03)	1.09	(.10)	14,073	4
Botswana	.98	(.01)	.99	(.01)	13,816	4
Chile	1.04**	(.01)	1.03*	(.02)	17,708	5
C. Taipei	.99	(.01)	1.02+	(.01)	18,986	6
Hong Kong	.98*	(.01)	.99	(.01)	16,695	6
Hungary	1	(.04)	.96	(.04)	16,942	6
Indonesia	1.04+	(.02)	1	(.03)	13,237	4
Iran	1	(.01)	.99	(.01)	21,479	6
Israel	.99	(.01)	1.02	(.01)	17,930	6
Italy	1.01	(.01)	1	(.01)	16,016	6
Japan	.99	(.01)	1	(.01)	21,327	6
Jordan	.99	(.01)	.99	(.02)	24,944	6
S. Korea	1.02+	(.01)	1.01	(.02)	20,489	6
Lithuania	.99	(.02)	1	(.03)	14,900	6
Malaysia	1	(.01)	.99	(.01)	24,373	6
Morocco	1.04***	(.01)	1	(.02)	27,275	5
New Zealand	.99	(.01)	1.02**	(.01)	18.046	5
Norway	1	(.01)	1	(.01)	15,315	5
Romania	.95*	(.02)	.96	(.04)	14,378	5
Russia	1.03	(.03)	1.04	(.05)	16,886	6
Singapore	1.02***	(.01)	1.01	(.01)	21,076	6
Slovenia	.96***	(.01)	.99	(.02)	12,923	5
Sweden	1.02+	(.01)	1	(.01)	16,061	5
Thailand	.94***	(.01)	.98*	(.01)	15,337	4
Tunisia	1	(.01)	.99	(.04)	12,231	4
Turkey	1	(.01)	.98	(.02)	18,418	5
USA	.98***	(.01)	1.01+	(.01)	36,834	6
England	1	(.01)	1	(.01)	14,846	6
Ontario	.99	(.01)	.99	(.01)	13,509	5
Quebec	.96***	(.01)	1	(.02)	14,359	5

OR, odds ratio per 4 year TIMSS cycle; SE, cluster-robust standard error (clustered by school); N, total number of observations; T, number of time points included in analysis

* $p < .05$, ** $p < .01$, *** $p < .001$

in novice teacher sorting inequity, including Hong Kong, Romania, Slovenia, Thailand, USA, and Quebec. While we do not test the 90/10 gaps, Figs. 3 and 4 confirm that in a vast majority of cases, they follow the exact same trend patterns as 66/33 gaps.

Similarly, out-of-subject teacher sorting is stable in almost all systems, with just two countries displaying statistically significant upward trends (Chile and New Zealand).[14]

---

[14] USA and Chinese Taipei show statistically significant upward trends at the 10% significance level.

However, in this case, just one educational system trends towards more equitable teacher allocation (Thailand).

Discussion

It is widely claimed that students with disadvantaged socioeconomic backgrounds have less competent and qualified teachers (OECD, 2018; Strietholt et al., 2019). Using relatively straightforward descriptive analyses, this paper addresses two important gaps in the international literature on teacher sorting. First, it provides up-to-date knowledge about the extent of teacher sorting cross-nationally using data where students can be linked to their teachers. Second, this paper provides the first cross-national evidence for changes in socioeconomic teacher sorting over time.

Though recent reports have raised the alarm about SES- based teacher sorting internationally (Luschei & Jeong, 2019; OECD, 2018), our analyses of 32 countries between 1999 and 2019 suggest a more positive picture. This is particularly true for out-of-subject teacher sorting. Just three to four countries show substantial differential exposure rates, including Chile, Thailand, Australia, and Chinese Taipei. Aside from these education systems, even at the most extreme ends of the socioeconomic spectrum, low-SES students were no more likely to receive out-of-subject mathematics teachers than those coming from high-SES backgrounds. This positive news extended to our findings in the trends over time, where just two systems showed an increase at conventional levels of statistical significance (Chile and New Zealand). These findings were somewhat surprising, as mathematics teacher shortages are discussed more and more in relation to increasing and better paid opportunities for the quantitatively inclined outside of teaching (TALIS, 2018). In fact, we find very few countries with *overall* increases in the share of students taught by out-of-subject teachers, which we would expect to find in countries with substantial or increasing shortages. Future research may investigate whether changes in teacher working conditions are compensating for such a potential change (for example, through larger class sizes). Importantly, Thailand is the only country showing decreases in such teacher sorting over time, indicating that where mathematics education inequities are found, they tend to persist. It is here worth re-emphasizing that out-of-subject teacher sorting is a problem in a very small number of education systems, but where it does exist, (Chile, Australia, Chinese Taipei, Quebec) it has persisted over the past two decades.

The findings for students with novice teachers are slightly less positive, with roughly one third of the countries in our sample showing at least some degree of sorting inequity. Low- and middle- income countries tended to be most affected, including Tunisia, Turkey, Iran, Indonesia, and Morocco, but inequities in novice teacher sorting were found between top and bottom student SES deciles in Romania, USA, Australia, Israel and Singapore as well. However, there is positive news here too. Since 1999, inequity increased in just three systems (Chile, Morocco, Singapore), and decreased significantly in six (Hong Kong, Romania, Slovenia, Thailand, USA, Quebec). Many more education systems show differences over time in novice sorting as compared to sorting by mathematics education, highlighting once again that their respective determinants likely differ.

Heterogeneity in teacher qualification gaps and trends

There appears to be limited patterns across groups of education systems by economic development level or geographic region. In the case of teacher sorting based on mathematics education, it was not possible to identify a pattern based on level of economic development or region. As mentioned, the most evident pattern is the prevalence of novice teacher sorting across lower income education systems. The institutional and demographic differences across such systems make it difficult to speculate on such determinants, however. It may suggest that overall working conditions in schools and classrooms with more socioeconomically disadvantaged students are more preferable in higher-income countries, and therefore more experienced teachers stay or even seek out such positions. It may also be the result of system-level educational characteristics, such as hiring practices or school choice conditions (OECD, 2019). Han (2018) finds a link between school autonomy and inequity in teacher sorting, but no difference in this link for higher- or lower- income countries. However, given the intuitive connection between more autonomous hiring practices and inequity in teacher allocation, there is also the puzzling finding whereby some systems with centralized teacher allocation showed clear inequities. For example, such was the case in Turkey. According to Özoğlu (2015), teachers in Turkey receive high 'seniority scores' at a faster rate by being allocated to teach in disadvantaged regions, and subsequently have greater choice over where they want to transfer, leading to higher turnover rates and proportions of inexperienced teachers in such schools. South Korea also has a centralized teacher allocation system, and is frequently heralded for its teacher rotation policy and excellence in the teacher workforce (Han, 2018; Kang & Hong, 2008; Luschei, Chudgar and Rew, 2013). However, although the estimate does not reach conventional levels of statistical significance, South Korea shows a slight increase in novice teacher sorting since 1999.[15] Luschei and Jeong (2019) also find a higher-than-expected level of sorting inequity in South Korea, and offer rising socioeconomic inequality alongside accountability and ability grouping as potential explanations. We do not compare our findings in a more detailed way to those two previous studies on teacher sorting using large-scale international data for various reasons.[16] There is also the case of Chile, which displays high 'equity' in novice sorting until 2019 when the trend reverses. Teachers in more disadvantaged schools are less likely to leave their jobs and retire much later in Chile (Meckes & Bascope, 2012), which does not point to equity but rather a particular teacher mobility pattern in the Chilean educational system.

In 2007, Akiba et al. reported that the USA had one of the highest teacher 'opportunity gaps' in the world. Our findings show that it has been decreasing since 1999 for novice teachers. This is in line with results from other research examining the impact of the No Child Left Behind (NCLB) act of 2001 (Boyd et al., 2008; DeAngelis et al., 2010). Under

---

[15] Despite this, we find average levels of overall teacher experience for top and bottom thirds to be the same (Appendix I). The results for novice sorting are therefore peculiar, as more novice teachers should drag down the overall average of teacher experience in a school, and can only be explained by having higher proportions of more experienced teachers as well. When we look further into the data, this is exactly what we find. Individuals in the top SES decile have just one fourth of teachers with more than 15 years of experience, while over one-half of teachers of students in the bottom SES decile have more than 15 years of experience.

[16] Akiba et al. (2007) do not report which indicators they use to determine student SES, how they accounted for sampling weights or missing data, so it is difficult to compare their analysis with our own. Additionally, Luschei and Jeong (2019) use different teacher quality indicators.

NCLB, teachers were newly required to be fully licensed, have a bachelor's degree and a demonstrated competency in their subjects taught, and states were mandated to eliminate the inequitable teacher distribution. While the novice teacher sorting trend may lend support to the NCLB policy, mathematics education sorting has remained generally stable or even increased. Overall, few educational systems display large teacher qualification gaps and rising sorting inequity. Clear exceptions to this pattern include Turkey, Tunisia, Morocco, Chile, Singapore, Australia, and New Zealand. There is also a question regarding the substantive significance of the teacher qualification gaps. Goldhaber et al (2018) raise the alarm about inequities when the proportion differs by just 1 to 5 percentage points. Many educational systems in our sample show much larger differences, despite not reaching statistical significance.

Limitations

As with all empirical research, this study has some limitations. First, TIMSS employs a stratified sampling procedure which results in greater uncertainty and larger standard errors. Although TIMSS and other international large-scale assessments provide unique opportunities for researchers to study cross-nationally focused questions, there is a high risk of sampling error which is compounded when examining population subsets. There are variations in sample sizes for each country and cycle, as well as differences in the degrees of missing data. Importantly, the data are not representative of teachers but rather students taught by mathematics teachers in the eighth grade. The data, therefore, cannot tell us anything about the national characteristics of teachers (such as mean teaching experience). While we have attempted to highlight and address as many issues related to the quality of the data as possible through the use of multiple imputation (at each country-year point) and conservative indicators of teacher quality to maximize statistical power and cross-national comparability, it is not possible to remove them entirely. Next, the SES measures in our study do not include parental occupation as it is not available from the data. Moreover, the SES- measures are reported by the students and will necessarily include some degree of error. The study also only addresses socio-economic inequities, and not those based on student migration background, language spoken at home, geographic location, or ethnicity. Our findings also provide a picture of teacher sorting based on just two indicators of teacher quality. While we posit that a 'novice' or 'out-of-subject' teacher designation will have a significantly detrimental impact on student outcomes, there are of course other factors contributing to teacher competence which we do not address. For instance, Sweden has reported a gap in access to certified teachers (Hansson & Gustafsson, 2016), and while Sweden shows specialization relatively equitable pattern regarding the indicators in this study, there may be clear inequities in the distribution of teachers based on other measures. Last, the study does not speculate on the determinants of teacher sorting trends across these educational systems, but this would indeed be a worthy topic for future research.

Concluding remarks

Based on the indicators in our study, teacher sorting by student socioeconomic status is an issue in a select group of education systems, and is widening in only about a handful. Such education systems should take this issue seriously, and focus on closing

these teacher 'qualification gaps,' work to improve working conditions, and better attract competent teachers to low-income schools and classrooms, particularly for the much more common phenomenon of novice teacher sorting. The qualification gaps likely translate to concrete asymmetries in levels of instructional quality for low- and high- SES students, and therefore are likely to influence both student test scores and behavioural long-term outcomes (Jackson, 2018; Rockoff, 2004; Sancassani, 2021). However, many of the education systems in our sample show no or very little difference in exposure rates to novice and out-of-subject teachers between low- and high-SES students, even at the most extreme ends of the socioeconomic spectrum. These education systems should not set the question of teacher sorting aside, however, and future research should work to rule out inequities across other indicators of teacher competence. However, policymakers should consider other potential determinants of the SES achievement gap with equal focus.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s40536-022-00125-9.

> **Additional file 1.** Supplementary materials regarding pooled estimates, point estimates from each country-year, and additional percentile comparisons.

Declarations

References
Adamson, F., & Darling-Hammond, L. (2012). Funding disparities and the inequitable distribution of teachers: Evaluating sources and solutions. *Education Policy Analysis Archives, 20*, 1–46.
Akiba, M., LeTendre, G. K., & Scribner, J. P. (2007). Teacher quality, opportunity gap, and national achievement in 46 countries. *Educational Researcher, 36*, 369–387.
Bacolod, M. (2007). Who teaches and where they choose to teach: College graduates of the 1990s. *Educational Evaluation and Policy Analysis, 29*, 155–168.
Baumert, J., et al. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal, 47*, 33–180.
Bitler, M., Corcoran, S., Domina, T., & Penner, E.K. (2014). *Teacher effects on student achievement and height: a cautionary tale*. in NBER Working Paper Series.
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*, 3.

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.

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*. in NBER Working Paper Series.

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

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.

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. (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.

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

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

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.

Downey, D. B., & von Hippel, P. (2004). Are schools the great equalizer? Inequality during the summer months and the school year. *American Sociological Review, 69*, 613–635.

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, 12*, 1–28.

Feng, L., Figlio, D. N., & Sass, T. (2010). School accountability and teacher mobility. *Journal of Urban Economics, 103*, 1–41.

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., 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.

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

Hanushek, E.A. & Woessman, L. (2011). *The Economics of International Differences in Educational Achievement*. p. 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, 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., & 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*. MIT Press.

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.

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.

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

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

Kratz, F., Pettinger, B., & Grätz, M. (2022). At which age is education the great equalizer? A causal mediation analysis of the (in-)direct effects of social origin over the life course. *European Sociological Review, 1*, 16.

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.

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.

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

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.

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

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

OECD. (2019). *Balancing School Choice and Equity. An International Perspective Based on PISA.* Paris: 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.

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.

Podolsky, A., Kini, T., & Darling-Hammond, L. (2019). Does teaching experience increase teacher effectiveness? A review of US research. *Journal of Professional Capital and Community, 4*, 286–308.

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.

Rice, J. (2003). *Teacher Quality: Understanding the Effectiveness of Teacher Attributes.* 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.

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. Wiley.

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

Shen, T., & Konstantopoulos, S. (2019). Estimating causal effects of class size in secondary education: Evidence from TIMSS. *Research Papers in Education, 10*, 1470–1146.

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

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.

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

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 L. Volante, S. V. Schnepf, J. Jerrim, & D. A. Klinger (Eds.), *Socioeconomic Inequality and Student Outcomes: Cross-National Trends, Policies, and Practices* (pp. 17–38). New York: Springer.

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

Walsh, D., McCartney, G., Smith, M., & Armour, G. (2019). Relationship between childhood socioeconomic position and adverse childhood experiences (ACEs): A systematic review. *Journal of Epidemiological and Community Health, 73*, 1087–1093.

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

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

Publisher's Note