ISOTIS

INCLUSIVE EDUCATION AND SOCIAL SUPPORT TO TACKLE INEQUALITIES IN SOCIETY

# Achievement inequalities and the impact of educational institutions

**Jesper J. Rözer and Herman G. van de Werfhorst**

THIS PROJECT HAS RECEIVED FUNDING FROM THE EUROPEAN UNION'S HORIZON 2020 RESEARCH AND INNOVATION PROGRAMME UNDER GRANT AGREEMENT NO. 727069

# Achievement inequalities and the impact of educational institutions.

**Jesper J. Rözer and Herman G. van de Werfhorst**

**Document Identifier**
D1.4

**Version**
1.1

**Date Due**
10 August 2019

**Submission date**
10 August 2019

**WorkPackage**
WP1

**Lead Beneficiary**
UvA

## PARTNERS INVOLVED

| Number | Partner name | People involved |
|---|---|---|
| 4 | University of Amsterdam | Jesper J. Rözer (j.j.rozer@uva.nl)<br>Herman G. van de Werfhorst |

## CONTENT

## LIST OF ABBREVIATIONS

| | |
|---|---|
| IEA: | International Association for the Evaluation of Educational Achievement. |
| OECD: | Organisation for Economic Cooperation and Development. |
| PIAAC: | Programme for the International Assessment of Adult Competencies |
| PIRLS: | Progress in International Reading and Literacy Study. |
| PISA: | Programme for International Student Assessment. |
| TIMSS: | Trends in International Mathematics and Science Study. |

## LIST OF FIGURES

## LIST OF TABLES

## EXECUTIVE SUMMARY

Reducing inequalities in educational achievement, like those by socioeconomic status (SES) and migration background, is not only a moral obligation, but also shows that much potential is unused in modern societies. In this report we reviewed the literature on how successful a broad range of educational policies are in reducing these inequalities – including educational expenditure, class size, teacher salaries, teaching hours, teacher quality, coverage of pre-primary school enrolment, age of compulsory education, tracking age, and vocational specificity. In addition, we tested how successful these policies are in tackling inequalities in high income countries using all available large scale student assessment data (i.e. the PIRLS, TIMMS, PISA and PIAAC) and a variety of methods in which we look both at between country differences, within country changes over cohorts, and in which we follow cohorts over their life course.

The results show that there are particularly large inequalities by socioeconomic background, while net gaps between first and second generation migrants and the majority population (i.e. third generation and higher) are substantially smaller. In several liberal and selective migration countries, such as the United Kingdom, Australia and the Gulf region, (children of) migrants even perform on average better than the majority populations, net of socioeconomic status. Moreover, the gaps between first generation migrants and the majority population is often approximately halved by the second generation, irrespective whether they first performed worse or better than natives. Across the life course both types of inequalities are already relatively large at grade 4, indicating that much occurs before grade 4. Inequalities are stable or may even decline while children are in primary and secondary school, but are stable or increase again thereafter. The last 10 years there has been a slight increase in the inequalities by socioeconomic background (comparing cohorts). First and second generation migrants, however, seem to slowly catch up with the majority population over time, especially at grade 4 and grade 8.

Governments implement a wide variety of policies related to the educational system to use the full potential of (all) students. Yet, the effectiveness of many of these policies to reduce inequalities have been debated in the literature. In line with this, for many policies we found null findings and mixed support. In addition, we found that many policies which are often meant to reduce inequalities did not do so. Sometimes it appears that policies aimed at reducing inequalities widened educational inequalities. Probably this is because advantaged groups make more use of these policies and know better how to bend the current structures in their advantage. Consequently, policy makers should not expect too much from general policies affecting all students, but may better implement policies that are specifically targeted at disadvantaged groups.

In our models four policies often – but not always – increased instead of decreased educational inequalities: the number of students that attended early education, teaching hours, and the experience and education of teachers. The literature shows that a couple of measures can be taken to reduce these negative side-effects. To ensure that early education reduces educational inequalities two things appear to be crucial: equal access and the guarantee of high quality care for disadvantaged children. High quality care goes beyond ensuring highly qualified staff for disadvantaged groups, but also that the staff is sensitive for their needs, something that is often lacking. The literature shows that longer school days are most likely to tackle inequalities when they are used to let disadvantaged children catch up, for example by repeating old material instead or learning something new in these additional hours. Increasing the experience and education of teachers is most likely to reduce performance gaps if it is assured that all children have equal access to and take up of those teachers, and that these teachers are sensitive for the needs of children with a low socioeconomic or migration background. This asks for more attention for the needs of children with a migration and low socioeconomic background in educational programs for (highly educated) teachers, and in extra training. These teachers might benefit from hands-on-tools to tackle educational inequalities. Besides these unintended effects, there is ample evidence in the literature that later tracking is likely to reduce educational inequalities. This is largely confirmed by our results. Thus, to reduce educational inequalities governments could consider increasing the age in which students are tracked by their abilities. Related to this, the literature shows that disadvantaged groups disproportionally go to vocational education, while within vocational education vocational subjects are often taught at the expense of general skills, such as language and mathematics. These are crucial skills for life-long learning and flexibility. As such, and partly supported by our results, a vocational orientation increases the risk of widening performance gaps in reading and mathematics. Thus, governments may want to ensure that within vocational education general skills such as language and mathematics get sufficient attention.

All in all, this report underscores that there are probably many unintended side-effects of educational policies, and that policies could be implemented in such a way that allow disadvantaged groups to equally benefit from these policies. Furthermore, this reports draws attention to structural factors, such as a later age of tracking and its vocational specificity. Despite that these policies might be more difficult to implement and may initially face a lot of criticism from citizens and teachers who have to get used to new practices, they might be crucial in tackling educational inequalities.

## 1. INTRODUCTION

Inequalities in educational achievement, like those by socioeconomic status (SES) and migration background, stand high on the political agenda because they are thought to not only reflect differences in abilities, but also in opportunities. Reducing these inequalities is not only seen as a moral obligation to improve the lives of these children, but also shows that much potential is unused in modern societies.

Inequalities in educational achievement vary greatly across countries, within countries over time, and within cohorts over their life course (as shown in a previous ISOTIS report: Rözer and Van de Werfhorst 2017). Yet, countries differ widely on what are identified as crucial aspects of educational systems, like their educational expenditures, teachers' salaries and quality of teachers, age of tracking, and the size of the vocational programs, and it is likely that these differences can explain part of the education inequalities (Van de Werfhorst and Mijs 2010).

To study how national policies shape educational inequalities, information about a great number of students across a broad set of countries is needed. In recent years, comparative studies of student assessments have enriched the opportunities to learn about the performance of students in many societies. Influential datasets include The Progress in International Reading and Literacy Study (PIRLS), the Trends in International Mathematics and Science Study (TIMSS), the Programme for International Student Assessment (PISA), and the Programme for the International Assessment of Adult Competencies (PIAAC). These comparative assessments make it possible to study academic outcomes ranging from grade 4 to adulthood on mathematics, literacy and science.

While the official organizations running these projects – most notably the OECD and IEA – report extensively about the performance of students in all the societies that are studied for each of the data projects separately, little effort has been made to combine assessments for descriptive and comparative purposes. Furthermore, the studies that did (e.g., Brown and Micklewright 2004; Hanushek and Wössmann 2006; Brown et al. 2007; Dronkers, Van Der Velden and Dunne 2012; Ruhose and Schwerdt 2016; Checchi and Van de Werfhorst 2017; Dämmrich and Triventi 2018) often involved two, but not more assessments, and the focus was often on overall performance instead of inequalities in performance. In addition, when inequalities in educational performance where studied the number of countries remains often relatively small. For example, Dämmrich and Triventi (2018), only examined the impact of one educational institution (i.e. tracking) for 9 countries for mathematics and for 11 countries for reading.

In this study we describe inequalities in student performance on test scores combining assessments from grade 4, grade 8, 15-year olds, and older adolescents and young adults. We focus on the impact of social origin and their migration background (that is whether the child or their parents is born abroad). The effects of migration background will be controlled for the socioeconomic origin of the parents, ruling out many differences by countries of origin, and showing the "net" effects of migration. Afterwards, and in addition to previous ISOTIS reports (Rözer and Van de Werfhorst 2017; Skopek et al 2018), we associate these overall patterns to a wide set of key policies related to educational systems: educational expenditure, teacher/pupil ratios, teacher salaries, teaching hours, teacher quality, coverage of pre-primary education, start and end age of compulsory education, tracking age, and vocational specificity. This way, we can test how these policies affect educational inequalities. We focus on high income countries only – as defined by the World Bank – because educational policies may work very differently in high and low income countries. To further rule out cross-country variation we control for GDP per capita and a set of individual level characteristics such as students gender and age. Together our data covers more than four million students from 52 countries. Thus, we improve up on previous research by comparing a greater set of institutions simultaneously for a greater set of countries, applying an innovative method (following cohorts over time) and comparing this to two methods which are often used in the literature (looking at variation between and within countries). The following descriptive research questions guide our analysis:

1. How do test scores differ **by socioeconomic origin and migration background** across countries, within countries over cohorts, and over the life course?
2. And what is the association between **national educational system characteristics** and these patterns?

## 2. THEORY AND LITERATURE REVIEW

### 2.1. INEQUALITIES IN EDUCATIONAL ACHIEVMENT

In this report we look at two types educational inequalities: 1) those by socioeconomic background (which will be measured by the education of the parents and the number of books at home) and 2) by migration background (which will be measured as whether the child or either one of the children is born abroad). We will discuss these two types of inequalities below.

**INEQUALITIES BY SOCIOECONOMIC BACKGROUND.** There are several reasons why children of high socioeconomic background perform comparatively well in school. First, genes may play a role. Parents with a high socioeconomic status may have more innate characteristics that are associated with school performance which they can genetically pass on to their children, such as their IQ, height and race (Boudon 1974; Preston and Campbell 1993; Tucker-Drob and Harden 2012; Mare 2014). Part of the genetic effects may result from interactions between genes and socioeconomic status when higher socioeconomic status promotes the genetic influence on achievement. Second, parents with a high socioeconomic status often possess many skills that they can transmit to their children (De Graaf, De Graaf and Kraaykamp 2000; Fan and Chen 2001; Domina 2005; Conger and Donnellan 2007; Levine et al. 2018). This transmission often works unconscious and subtle. For example, Hart and Risley (2003) estimated that at the age of three, prior to the beginning of schooling, children from upper socioeconomic status families have heard more than 30 million more words than children from lower socioeconomic status families, which creates an important advantage in one's educational career. Third, parents with a high socioeconomic status can directly use their skills to help their children with their homework. Fourth, parents are likely to familiarize their children with the proper cultural attitudes and etiquette in way that allow their children to feel more comfortable at school and consequently perform better (Bourdieu and Passeron, 1977; Dee 2005). Fifth, family stress may play a major role (Conger and Donnellan 2007; Evans and Kim 2013). Stress is often higher among lower class families due to a lack of resources and economic hardship. This may immediately impact the performance of children because several regions of the brain that are supporting cognitive development are sensitive to stress hormones. For example, stress lowers the development of the hippocampus and prefrontal circuitry, resulting in lower memory capacity and cognitive and emotional regulation (Hackman 2009; Noble et al., 2012; Noble et al., 2015). More indirectly, family stress may consume parents time, explaining why upper class parents are found to have a stronger involvement in their children's school life (Domina 2005). And maybe even more worrisome, stress may result in unsuitable or even harmful parenting practices, resulting in children's maladjustment, which may translate in poorer cognitive performances (McLoyd 1998; Skopek and Passaretta 2018). Sixth, to help their children, upper class parents more often question school authority, and more effectively mobilize other parents and professionals to contest the judgments of school officials that are disadvantageous for their children (Horvat 2003). Additionally, parents with a high socioeconomic status can use their knowledge, access to professionals, and financial resources to get their children into the right school and academic track, and help them stay there (De Graaf 1986; Breen and Goldthorpe 1997). In turn, the children may learn more in the classes they end up with. Seventh, the class climate is often argued to be better when the average socioeconomic level is higher, with as a result that high socioeconomic children learn more (Hoxby 2000). Combined, these various factors help the children of parents with a high socioeconomic status to perform better in school and to reach more often a high educational level than the children of parents with a lower socioeconomic status.

A large body of research confirms that children form upper class families have higher educational achievements than children from lower class families (for a review see e.g. Skopek and Passaretta 2018). Meta-analyses report average correlations of about 0.3 between socioeconomic background and student achievement (White 1982; Sirin 2005; Ewijk and Sleegers 2010. cf. Skopek and Passaretta 2018). However, they also indicate that there is considerable variation between students, and the method of the studies. Several studies indicate that social background plays a more important role for reading than for mathematical competencies (Cooper, Nye and Charlton 1996; Bol, Witschge, Van de Werfhorst and Dronkers 2014). In addition, trend studies reveal that socioeconomic gaps have increased the last 50 years in many western societies (Reardon 2011; Chmielewski 2019).

To provide a better indication when SES gaps emerge and evolve over the life course, recent studies have adapted (pseudo) longitudinal designs. For example, Dämmrich and Triventi (2018) and Rözer and Van de Werfhorst (2017) have combined students assessment data – like the PIRLS, TIMMS, PISA and PIAAC – to assess how inequalities within a cohort change when this cohort matures. Both studies already found significant differences between SES groups that arise early in the life course, that is before grade 4 (approximately age 10). Comparing the mathematics and reading performance in respectively 8 and 11 countries, Dämmrich and Triventi (2018) found that differences between children with few and many book at home slightly increased between grade 4, age 15 and young adulthood. Rözer and Van de Werfhorst (2017) used a more extensive approach and used information from over 100 regions and compared children both by their number of books at home and the education of their parents. Relying mostly on data of the education of the parents, they found that socioeconomic inequalities first declined between grade 4 and age 14/15, before they started to increase from around age 16/17.

Longitudinal analyses that actually can trace the same students over time are generally in line with the results from the pseudo panel analyses (e.g., Lee and Burkam 2002; Feinstein 2003; Farkas and Beron 2004; Cheadle 2008; Morgan, Farkas and Hibel 2008; Caro et al. 2009; Fernald, Marchman and Weisleder 2013; Potter and Roksa 2013; Bradbury et al. 2015). What most of these longitudinal studies add is that they often start following children before grade 4, sometimes even before birth. For example, using EEG Noble and colleagues show that at the start of life there are no differences visible in brain capacity between children of low and high SES background (Brito, Fifer, Myers, Elliott and Noble 2016), but that the brains of high SES children typically mature ‘better’ – that is resulting in greater surface area and cortisol thickness that are associated with cognitive performance – due to a more stimulating cognitive environment. These differences become already visible in the first months after giving birth (Noble et al., 2015). In line with the pseudo-longitudinal analyses of Rözer and Van de Werfhorst (2017), these studies show that the SES gaps tend to grow at a slower pace, or even decline while children are at school. This made Skopek and Passaretta (2018, p. 1) tentatively conclude that “schooling decreases social inequality in learning”. As they argue, schooling, which quality is often relatively homogenous across social groups, may compensate for the large differences in the quality of the home environment of high and low status families. In line with this, research shows that student achievement gaps tend to be relatively stable during the school year, but rise considerable during vacations (Alexander, Entwisle and Olson 2001; Downey, Von Hippel and Broh 2004).

**INEQUALITY BY MIGRATION BACKGROUND.** Net of their socioeconomic status, there are several factors that limit the educational achievement of migrants and their children. If parents were born abroad and are not fluent in the host language, this will form a disadvantage for the child's performance (Van De Werfhorst and Van Tubergen, 2007). This will have a direct effect on the language proficiency of the children. Indirectly, poor language skills on the side of parents may impede teacher-parent interactions, while on the side of the children may cause that they have difficulties in following the teacher and interacting with class-mates (Crozier and Davies 2007). In a similar vein, parents born abroad are likely not familiar with the educational system and content of the host society, and as a result may be less likely to help their children maneuver through the educational system, setting them in the disadvantage to fully profit from it in terms of performances (Schneider and Coleman 1993; Van De Werfhorst and Hofstede 2007; Van De Werfhorst and Van Tubergen 2007; Jonsson 2013; Pfeffer 2008). For example, parents may not know the school content of subjects such as language or the national history by which they will be less likely to help their children with these subjects, or parents may poorly advice on the track the child may go to. In addition, as children are likely not socialized with the cultural norms of the educational system in the host society, they may feel less comfortable at school, and at the extreme may even not embrace the goals of the educational system – although, in contrast, it is often found that immigrant students have higher aspiration levels than one would expect on the basis of their educational achievements (Heath, Rothon and Kilpi 2008; Geven 2019; Salikutluk 2016). On top of this, immigrant children face the risk of being discriminated, or expect to be discriminated later life, and this may further demotivate them. Furthermore, stress levels are often higher among immigrant families because migration histories and integration in a new society can be stressful, while stress limits cognitive performance. On the side of teachers, they may expect less of immigrant children because of their disadvantages and therefore not fully stimulate them. In addition, teachers are often culturally distant from migrants, for example due to their education and socializing in the educational system (Dee 2005; Bourdieu and Passeron 1977). As similarity breeds mutual trust and positive expectations this may be a crucial factor why – perhaps especially among highly educated and experienced teachers – migrants perform worse than natives in schools. Finally, the socioeconomic status of the parents may play an important role. Immigrant parents often have a lower education, and have lower earnings and have often accumulated – or inherit – less wealth, such that they have fewer financial resources to support their children (Becker 1994).

However, several factors may also help migrants. Crucially, a 'positive selection' of immigrant parents, and their typical strong drive to improve their life, may lead to higher educational aspirations for some children (Heath, Rothon and Kipli 2008; Jackson, Jonsson and Rudolphi 2012; Heath and Brinbaum 2014; Salikutluk 2016; Van de Werfhorst and Heath 2018). And related to this, immigrants with more potential to increase their life, such as having a high IQ compared to their social standing, may be more likely to migrate (Levels, Dronkers and Kraaykamp 2008). In addition, Timmermans, de Boer, Amsing and Van der Werf (2018) showed that children of immigrants were over-advised by the schools to choose higher tracks, because teachers are aware that immigrant children may underperform and hope that they will do better in secondary education. The pattern of over-advice is however disappearing in the Netherlands, which is interpreted by the authors from a less welcoming culture towards immigrants in the past decade.

Despite that migrants also face some advantages, previous studies show that in most societies immigrants on average are disadvantaged when it comes to educational attainment (Heath and Brinbaum 2007; Schneeweis 2011; Rözer and Van de Werfhorst 2017; Passaretta and Skopek, 2018a), although this is not the case for all minority groups (e.g. Kao and Tienda,1995; Heath and Brinbaum, 2014b). Well known are the Asian students that outperform majority populations in many western societies. At the same time, many studies show that differences in parental (socioeconomic) background can account for much of the lower achievement of children of immigrants (e.g., Magnuson, Lahaie and Waldfogel 2006; Tillman, Guo and Harris 2006; Brinbaum and Cebolla-Boado 2007; Heath and Brinbaum 2007; Dustmann, Frattini and Lanzara 2012; Becker, Klein and Biedinger 2013; Ichou and Van Zanten 2014; Rözer and Van de Werfhorst 2017; Passaretta and Skopek, 2018a). However, net of socioeconomic economic differences, small differences often remain visible. For example, in Europe disadvantages often remain for children of Turkish, North African, Southern European and Carribian background (Riphahn 2003; Fekjaer 2007; Phalet, Deboosere and Bastiaenssen 2007; Stören and Heiland 2009; Jonsson Rudolphi 2010), and in the United differences remain for children from Hispanic origin (Warren 1996; Rumberger and Thomas 2000; Portes and Rumbaut 2001; Lutz 2007). Yet, a recent report found that the migrant-native gap in a variety of competence domains if fully explained by difference in familial socio-economic status, and migrants tend even to outperform natives with similar social origin in many countries, such as Germany (Passaretta and Skopek, 2018b), the Netherlands (van Huizen, 2018) and the United Kingdom (Skopek and Passaretta, 2018b).

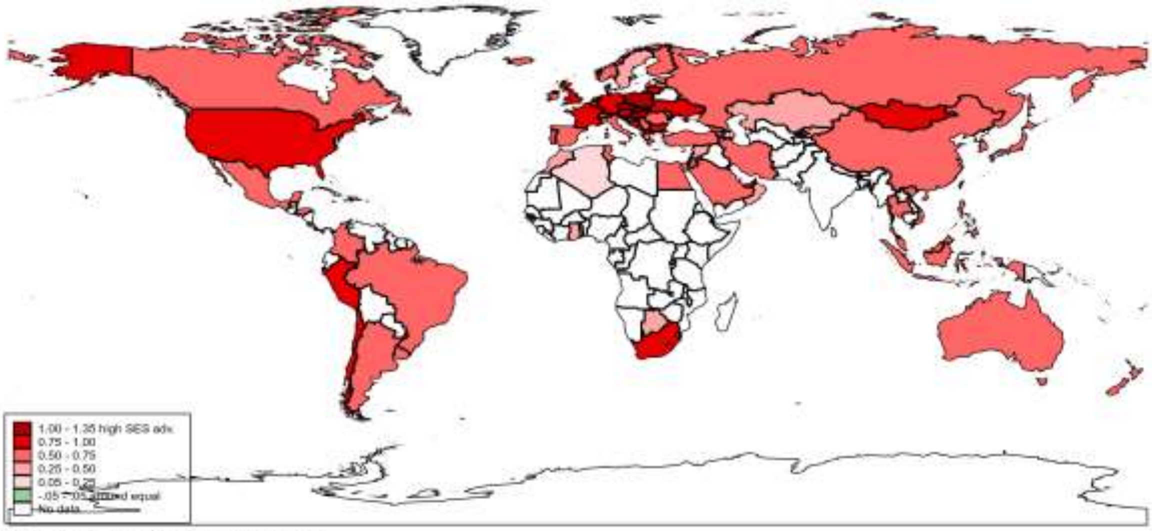
It appears that when there are differences between children with and without migration background, these are often already present in early education (Biedinger, Becker and Rohling 2008; Dicks and Lancee 2018; Passaretta and Skopek, 2018a). This suggests that the home context plays an important role. The main difference between the native majority and students with a migration background lies in acquiring language skills of the host society (Magnuson, Lahai and Waldfogel 2006; Crosnoe 2007; Becker et al. 2013; Passaretta and Skopek, 2018b), although also in early education there are already differences visible in cognitive performance (Becker et al. 2013). The gaps in language skills tend to decline while immigrant children attend school, because from then onwards they are fully exposed to the language of the host society (Magnuson et al. 2006; Crosnoe 2007; Becker et al. 2013; Skopek and Passaretta, 2018b). In contrast, cognitive skills may widen because of so-called 'lock-in' effects, meaning that pupils who lag behind at a certain age may subsequently fall behind, for example because they lack the knowledge needed to understand the subject that are taught (Heckman 2006; Magnuson et al. 2006; Rumberger and Tran 2006; Crosnoe 2007; Lahaie 2008; De Feyter and Winsler 2009; Reardon and Galindo 2009; Dicks and Lancee 2018). Hence, these cognitive gaps may widen.

## 2.2. THE IMPORTANCE OF NATIONAL SYSTEMS, INSTITUTIONS AND POLICIES

While on average there might be inequalities in educational achievement by socioeconomic and migration background, as shown in previous ISOTIS reports (Rözer and Van de Werfhorst 2017; Passaretta and Skopek, 2018a), there are large differences across countries (see Figure 1). Part of the explanation of cross-national differences may be found in educational policies and institutions. In this report we make a distinction between two different types of policies: 1) those with respect to educational input, and 2) those who imply reforms of the educational system.

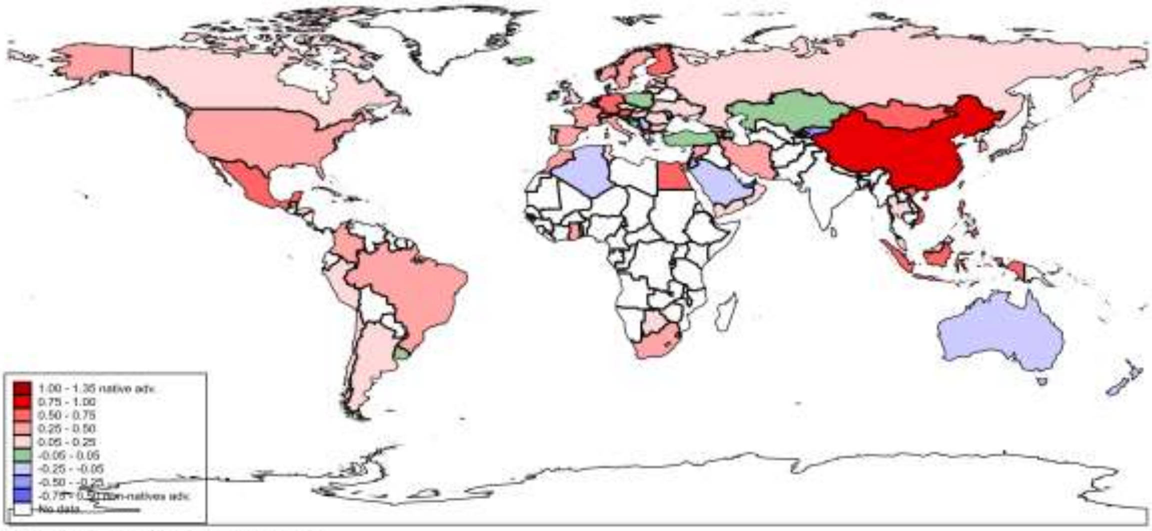
*Figure 1. Socioeconomic inequalities in mathematical skills across the world.*

Panel a: by parental education




Note: all available data per country combined

Panel B: by migration background.




Note: all available data per country combined

Source: Rözer and Van de Werfhorst (2017).

### 2.2.1. EDUCATIONAL INPUT

**EXPENDITURE.** One of the most popular cries to improve educational systems is to increase the budget. More money can increase educational performances and decrease educational inequalities, for example because it can be used to support disadvantaged children and schools. However, it is the question whether this actually occurs, because advantaged parents can use their resources to affect policies that work in their interest. A general sociological theory on educational inequality is the theory of "maximally maintained inequality" (MMI) which states that inequalities will only reduce at educational attainments where saturation has been achieved among the advantaged classes (Raftery and Hout 1993). Also, the Effectively Maintained Inequality thesis of Lucas (2001) states that qualitative inequalities are maintained even at saturated levels of attainment, because middle class parents are able to create institutions that promote their own children's advantage.

In this light, it is much debated whether there is a positive relationship between educational expenditure and the quality of education. This started with the famous Coleman report (1966) which showed that most variance in student achievement was explained by family background characteristics and community level variables, while school resource variables explained no or little variance. Literature reviews of Hanushek (1981, 1986, 1997) from the 1990s till late 1990s came to a similar conclusion. However, around the mid-1990s and early 2000s, this was contested by Greenwald, Hedges, and Laine (e.g., 1996), and Krueger (2003), who used more powerful meta-analytic techniques that take effect sizes into account. Currently, this debate is still not over. For example, using more recent data students' performance data like those available in the PISA, Hanushek and Wößmann (e.g. 2003, 2017) showed that between and within countries there is no or only a very modest relationship between educational expenditure and educational achievement. In contrast, using school reform data, others found positive effects of increases in student expenditure on student achievements, especially for children of low-income families (e.g., Papke 2005; Roy 2011; Jackson, Johnson and Persico 2016; LaFortune et al. 2018; Jackson, Wigger and Xiong 2018).

One possible explanation for the small effects is that money may be spent on schools that already have a large budget. Consequently, the marginal gains may be limited. In this respect LaFortune, Rothstein and Schanzenbach (2018) show that from the 1990s many US states have deliberately implemented policies that increased funding for schools in low-income districts – which was often forced by court. They show that, in line with previous research, spending on these schools have a large positive effect on student achievement (e.g., Guryan 2001; Papke 2005; Roy 2011; Jackson et al. 2016). However, counterintuitively, these reforms do not raise the performance of low-income and minority students on the state level, because, as LaFortune et al. (2018) argue, these students are less concentrated in these areas than one might expect.

Hence, educational expenditures may by itself do little to decrease educational inequalities, and it may matter a lot where the budget is used for. Besides the budget and location of the school, research has especially focused on four popular policy interventions: smaller classes, higher salaries for teachers, better quality teachers, and more teaching hours.

**CLASS SIZE.** Reducing class sizes may be one way to reduce educational inequalities. Smaller classes help teachers to tailor their lessons and attention towards a smaller groups of students which can increase the quality of instruction (e.g., Smith and Glass 1980; Achilles 1999; Blatchford, Basset and Brown 2011). In addition, smaller classes help teachers to give certain students the little additional attention they need. And, smaller classes may be favorable to students' performance because students may sooner socialize to the school culture in smaller settings, making them feel more comfortable, which in turn would increase their achievements. Especially children with a migration background or low socioeconomic background may benefit from smaller classes because smaller classes help students to be socialized more rapidly into the school culture – which they are less familiar with – and because they more often need the little bit of extra attention (Gustafsson 2003).

Research on the effect of class size on the overall student performance is heavily mixed, and may show weaker effects than popular accounts do suspect (e.g. Hoxby 2000; Blatchford et al., 2007; Urquiola 2006; Wößmann and West 2006; Wößmann 2016; Shen and Konstantopoulos 2017). Despite this, effects for disadvantaged groups seem to be more favorable. Although effect sizes are relatively small, substantial empirical evidence supports the idea that students who struggle in school and those from ‘unfavorable’ backgrounds benefit most from class-size reductions (Angrist and Lavy 1999; Krueger 1999; Blatchford et al. 2003; Maasoumi, Millimet and Rangaprasad 2005; Bosworth 2014). A recent study showed for Sweden that class size even has a causal effect on earnings in the later career (Fredriksson et al. 2013), with effect sizes passing the cost-benefit analysis of smaller classes.

**TEACHER SALARIES.** Another measure to decline inequalities is to raise teacher salaries. This can either be in the form of overall teacher salaries or in the form of incentive or performance-based pay. Both can have an immediate effect on the quality of teachers because it may provide a motivation for teachers to work harder. Incentive or performance-based pay forms a direct incentive, while the effect of raising the overall teacher salaries may work more subtle as well, as higher salaries not only increases teacher’s income, but also how teachers are valued in society. This respect might be an additional motivation for teacher to perform well, and might also form a reason for students and parents to treat teachers with respect, making teaching easier (Dolton and Marcenaro-Gutierrez 2011). In the long term, due to higher salaries and more respect, more people may want to become a teacher. This can facilitate the recruitment of more talented individuals, and make the profession more selective (Clotfelter et al. 2008; Chetty, Friedman and Rockoff 2014). In addition, it may set a competition for higher quality in motion – especially when teachers are paid by their performance – that improves the achievement of students (Dolton and Marcenaro-Gutierrez 2011; Chetty et al. 2014).

Support has been found for the positive effects of raising the salaries of teachers. For example, Clotfelter et al. (2008) found that an annual bonus reduced teachers turnover, Hendricks (2014) found that teachers pay and turnover were strongly associated, and Steele, Murnane, and Willett (2009), Leigh (2012) and Rikman, Wong and Winters (2017) found that higher salaries increased the aptitude of future teachers. In addition, Lee and Barro (2001), Boarini and Lüdemann (2009) and Dolton and Marcenaro Gutierrez (2011) found that pupils perform better in countries with higher salaries. Yet other studies are more critical. For example, several studies suggest that the power of higher salaries to attract better teachers (Hanushek, Kain and Rivkin 1999; Hanushek 2003; Loeb and Page 2000; Glazerman and Seifulah 2012), to reduce the turnover of teachers (Steele, Murnane and Willett 2009), and to increase students’ performance (Hanushek 2006; Glazerman and Seifulah 2012) is limited.

Increasing the salaries of teachers might be especially effective in reducing performance inequalities within education. An abundance of evidence, especially from the United States, shows that the lowest performing teachers work in schools that are disproportionally attended by disadvantaged children, and that the lower salaries in these schools may form one reason why the better performing teachers do not work in these schools (Lankford, Loeb and Wyckoff 2002; Scafidi, Sjoquist and Stinebrickner 2007; Clotfelter et al. 2011; Jackson et al. 2016; LaFortune et al. 2018). Yet, direct evidence that higher salaries in a country or larger region raises the performance of disadvantaged children remains limited.

**TEACHER QUALITY.** Higher salaries are often implemented to attract higher quality teachers. Although this effect is often found to be weaker than many suspect (see above), high quality teachers may be of great importance for increasing performances and decreasing inequalities. High quality teachers should be better able to explain the learning content and also to be better able to give the individual student the specific attention and care (s)he needs.

There is ample evidence that teacher quality greatly matters for students achievements (e.g. Nye, Konstantopoulos and Hedges 2004; Hanushek and Rivkin 2005; Aaronson, Barrow and Sander 2007; Kane, Rockoff, and Staiger 2008; Clotfelter, Ladd and Vigdor 2010; Chetty, Friedman and Rockoff 2014; Hanushek, Piopiunik and Wiederhold 2018). However, teacher credentials generally have only a minor effect (Rockoff 2004; Harris and Sas 2011), suggesting that qualification level is a suboptimal indicator of teacher quality. Besides subject knowledge, especially experience (e.g. Gerrtisen et al. 2010; Harris and Sas 2011; Papay and Kraft 2011), and pedagogical skills are of importance (Rivkin, Hanushek and Kain 2005).

One of the major limitations towards reducing inequalities through teacher quality is that high quality teachers often work in schools with the most advantaged and best performing students (Borman and Kimball 2005; Flores 2007; Clotfelter et al. 2010; Mansfield 2015; LaFortune et al. 2018). Hence, inequalities can be reduced by raising the overall level of teachers, such that also less advantaged children are taught by high quality teachers (e.g. Desimone and Long 2010; Montt 2011; Hanushek et al. 2018). On top of this, it seems plausible that disadvantaged children would benefit from high qualified (and highly paid) teachers most. This may be the case because they know better what these children need, have more patience with them, are more highly committed, or have better pedagogical skills to train them. This reasoning is supported by a handful of studies. For example, Aaronson et al. (2007) found that teacher quality is especially important for lower ability students.

**TEACHING HOURS.** Another measure that may decline educational inequalities is the number of teaching hours. Intuitively, increasing school teaching hours should increase student achievement because there are more learning opportunities. Time that, for instance, could be used to increase the debt of the subjects (Farbman and Kaplan 2005). Things that children might not learn outside school (OECD 2016; Heubener, Kuger and Marcus 2017). However, increased teaching hours could lead to a misuse of the additional time and negative impact on student motivation (Patall et al. 2010). Students also need to rest and need to have time to progress the things they have learned at school. In addition, fewer teaching hours may release pressure from already busy teachers, and the teachers may use the additional time to prepare classes or follow courses which might eventually benefit education. Hence, theoretically, more teaching hours does not per definition increase student achievement. While comparative research often finds small, and sometimes not significant, effects of teaching hours on student achievement (e.g., Lee and Barro 2001; Scheerens 2014; Mandel and Süssmuth 2011), other studies that take out further heterogeneity between students, find stronger positive effects (e.g., Bellei 2009; Goodman 2014; Cortes, Goodman and Nomi 2015; Lavy 2015; Rivkin and Schiman 2015; Aucejo and Romano 2016; Heubener et al. 2017).

Both higher- and lower-performing students might particularly profit from additional teaching hours: higher-performing as they learn more efficiently, and lower-performing students as they need more time than better-performing students to process new learning content. When the additional time is used to provide extra time to process the information given lower performing students might benefit most. But when instead the additional hours is used to teach new subjects, disadvantaged children may be less likely to reap the fruits of the additional hours. In this instance, the school system may ask too much of them and favour the performance of high-performing students (Heubener, Kuger and Marcus 2017).

Unique for children with a migration and low socioeconomic background is that their parents are on average less effective in passing on information than higher educated and majority parents, meaning that they might benefit most from the additional hours (Dewey, Husted and Kenny 2000; Alexander et al. 2001). At the same time, they are least familiar with the school culture, causing that they might find the additional time most stressful. Hence, opposing arguments can be given whether additional time is beneficial for reducing inequalities.

In line with these opposing arguments, research found mixed evidence whether additional instruction time benefits the bottom or top most. Some studies found that the top profits most from additional teaching hours (e.g. Cattaneo et al. 2017; Heuber et al. 2017), some studies found no significant effect (Banerjee et al. 2007), and others found that the bottom or most disadvantaged socioeconomic groups profit most from additional teachers hours (Harn, Linan-Thompson and Roberts 2008; Carlsson et al. 2015; Jez and Wassmer 2015). Furthermore, several studies show positive effects for students with a migration background, for instance by pooling the TIMMS and PISA (Schneeweis 2011), or by pooling the PISA and PIRLS (Ammermüller 2013). All in all, it remains uncertain whether the least performing students benefit most from additional teaching hours, and there is some – but still little – evidence that socioeconomic disadvantaged and children with a migration background benefit most from additional teaching hours.

### 2.2.2 STRUCTURAL DIFFERENCES

Besides the direct inputs into schools and classes, educational systems also differ widely in how they are structured. These structures are often historically grown, but might have large impact on educational performances and inequalities. We discuss four structural differences: enrollment in pre-primary education, compulsory schooling age, age of tracking, and vocational specificity.

**PRE-PRIMARY EDUCATION.** Early childhood education is shown to be of great importance for child development. Education at young ages is greatly important for cognitive performance because the brain is most 'plastic' early in life, implying in more popular terms that learning is easier (Sowel 2003). Brain growth is primarily driven by caregiver input, like rich language environments, adult–child interaction, and caregiver responsiveness (Nelson et al. 2007; Fernald et al. 2013; Raver et al. 2013). In addition, as "learning begets learning", a good start may make further learning easier (Heckman 2000; Carneiro and Heckman 2003). Therefore, the younger children are, the higher the returns of investment in schooling potentially are (Berlinski, Galiani and Gertler 2009; List, Samek and Suskind, 2018; Skopek and Passaretta 2018).

Children with few educational resources – in terms of socioeconomic and migration background – may profit most from pre-primary education for several reasons (for meta-analyses see Camilli, Vargas, Ryan and Barnett 2010; Nores and Barnett 2010; for a recent review see Kulic et al. forthcoming). First, their home-environment offers often a lower quality of learning stimulation, implying that professional teaching has greatest added value for those children (Dearing, Berry and Zaslow 2006; Duncan and Magnusson 2011; Reardon 2011; Raudenbush and Eschmann, 2015). Additionally, the lowest performing students may benefit most from competent peers, for instance in their language performance (Ribeiro, Zachrisson and Dearing 2017; Hanushek et al. 2003). Second, the cognitive development of children with few educational resources is often lower because of lower quality nutrition and being exposed to more environmental toxicants, such as smoke (Bobak et al., 2000; Georgieff 2007; Layte and Whelan, 2009; Prado and Dewey 2014). Early education can offer more healthy nutrition and environment and can stimulate parents to offer it to their children. Third, family stress is an important explanation for inequalities by socioeconomic and migration background, and early education can reduce the amount of stress a child is exposed to by offering a safe and structured environment, and by taking some care away at hand of the parents (Zachrisson and Dearing 2015). Besides lowering the stress at home, structured and safe early education may also foster stable routines at home, and improve the quality of care that parents provide to their children (Hsin and Felfe 2014). Fourth, the familiarization with the school context is often poorer among immigrant and low class families. Through early education children are socialized with the school context. Therefore, the relative gains of pre-primary education may be greatest for poor children and children with a migration background (Esping-Andersen et al. 2012).

However, there are also several reasons why advantaged children may profit most from pre-primary education, especially those with a high socioeconomic status and no migration background. First, native and high socioeconomic parents are often reported to make greater use of pre-primary education because they are more able to afford it and because they more often need it as they more often work (Zachrisson, Janson and Nærde 2013; Petitclerc et al. 2017; Skopek et al. 2017; Cornelissen et al. 2018; Kulic et al. forthcoming). Hence, children of high educated parents that profit most from an extension of pre-primary education until a point of saturation is reached at the side of high educated parent, that is the point in which educated parents start to make greater use of additional coverage. Second, for a plethora of reasons children from advantaged families may profit most from the care they get; they are likely to go to better early education with better facilities and more competent teachers and peers that stimulate them; they may get the most of attention from the teachers; their parents may better cooperate with the teachers facilitating better learning conditions for their children; and their parents may cause that their children get most out of the interactions at early education, for instance by letting their children feel comfortable and extending the care to their daily routines (e.g. Dearing, Kreider and Weiss 2008; Ribeiro et al. 2017; Cook, Dearing and Zachrisson 2018). Despite that care is typically of high quality at home, especially due to the advantages that are more difficult to offer at home, such as daily interactions with competent peers, advantaged children might even learn more – and/or other skills

– in early education than at home.

Numerous studies show that pre-primary education has the potential to reduce early inequalities in brain development and cognitive performance, both in terms of migration and socioeconomic background (e.g., Bauer and Riphahn 2003; Feinstein 2003; Kamerman et al. 2003; Leuven et al. 2004; Goodman and Sianesi 2005; Schütz et al., 2008; Schneeweis, 2011; Felfe, Nollenberger and Rodríguez-Planas 2015; Cebolla-Boado, Radl and Salazar 2017; Van Huizen and Plantenga, forthcoming; for a review see Raudenbush and Eschmann 2015; Cornelissen et al. 2018; Kulic et al. forthcoming). Leuven et al. (2010) showed this with a rather ingenious study design. They exploited the fact that in The Netherlands primary-school children who are born just after the summer vacation experience eleven additional weeks of pre-schooling than students who are born just prior the summer vacation (who were no longer entitled to pre-school facilities during the summer). Using this as an instrumental variable, they examined the impact on disadvantaged children, measured by a combination of ethnicity and education of the parents. They showed that eleven weeks of additional education in pre-schools increased language and math scores for disadvantaged students, while no impact was found for advantaged children. Yet, at the same time, children from high socioeconomic status and native parents also more often make use of pre-primary education (Skopek et al. 2017; Petitclerc et al. 2017). Combining these results it remains the question what the impact is of raising the number of children that attend pre-primary education (cf. Kulic et al, forthcoming).

**COMPULSORY SCHOOLING AGE.** Lengthening the phase in which students are obliged to follow education can happen at two points: by starting earlier or being able to leave later. Starting compulsory schooling earlier often implies that the take-up of pre-primary education goes up (defined as education offered before the age of 6). Especially higher educated and native parents make use of pre-primary education, and therefore starting compulsory schooling earlier may especially increase the amount of pre-primary education taken by socioeconomically disadvantaged children and children with a migration background. In turn, as the home environment of these families is often less cognitive stimulating than those of native and high socioeconomic families, these children may also profit most (Felfe et al. 2015). Hence, pulling the starting age of compulsory schooling forward may reduce educational inequality by socioeconomic and migrations background.

Next to lowering the compulsory age at which children have to go to school, governments can extend the compulsory age at which children can leave education. The extension of compulsory schooling age affects cognitive development later in life (Angrist and Krueger 1991; Acemoglu and Angrist 2000; Oreopoulos 2007; Pischke 2008). Its main impact is by keeping individuals ‘on board’ who otherwise should have been dropped-out of the educational system. This is of significant importance because the returns to schooling are substantial for those individuals leaving school at or near the dropout age (Angrist and Krueger 1991; Acemoglu and Angrist 2000; Oreopulus 2007).

A few studies examined the impact of extending compulsory age on performance on test scores. Extending compulsory schooling may reduce inequalities by increasing the attained schooling of some of the lowest performing students who often come from a low social class and have a migration history. Yet it appears that the effects are limited in size (Oreopoulos 2007; Stephens and Yang 2014). This is because of two reasons. First, (disadvantaged) children who remain in school because it is compulsory may fail to gain a diploma despite being in school longer. Second, more high-skilled and more motivated students probably benefit more from an additional year of schooling, despite that it is likely that they remain in school irrespective of the compulsory age (Raudenbush and Eschmann 2015). This illustrates the limited extent to which schooling at later ages ceases to reduce skill dispersion and inequalities.

**TRACKING**. During their educational career, students are often 'tracked' into different classes or schools based on their (expected) performance. We focus on the age of tracking. In more selective systems, students are tracked into streams sometimes as early as age 10. In comprehensive systems, students remain much longer in untracked classes, often up to age 16.

The motives for tracking are twofold. First, the sorting of students allows to sort them according to their talents and interests, which would help to optimize their skills and knowledge and to optimally prepare them for their future (Schafer and Olexa 1971; Oakes 1985; Brunello and Checchi 2007; Van de Werfhorst 2014, 2018). In practice, tracking enables to offer students distinctive programs, often along the lines of vocational and academic education, which should be aligned to their potentials and the skills they need in the future (Trautwein et al. 2006). Second, the creation of more homogeneous classes is argued to increase efficiency. Homogeneous classes allow for a more focused curriculum and allows teachers to tailor lessons to students' specific needs (Hallinan 1994; Brunello and Checchi 2007). Teachers have to worry less about losing the slowest learners or boring the fastest ones (Hanushek and Wößmann 2006).

Despite these alleged advantages, there are also certain risks of (early) tracking. Importantly, students may end up in the wrong track. This is particularly likely at an early age as abilities are difficult to predict at an age as early as age 11 or 12 (Manning and Piscke 2006). Furthermore, students might perform worse in tracked systems, especially those at the bottom, because the quality of education is often lower, fewer resources are allocated to it, a low track might signal low abilities which would work as a self-fulfilling prophecy, peer culture is often of lower quality (e.g. misconduct and motivation), and teaching is often not adapted to the specific (homogeneous) group of students in the lower tracks. Because of these risks, most studies show that early tracking has no or a small negative effect on mean performance of students, while effects are particularly weak or even negative for low performing students (e.g., Manning and Pischke 2006; Brunello and Checchi 2007; Van Houtte and Stevens 2008; Horn 2009; Van de Werfhorst and Mijs 2010; Bol and Van de Werfhorst 2013; Heisig and Solga 2015; Lange and Werder 2017).

To the extent that (early) tracking disperses skills, it may increase existing socioeconomic inequalities because low SES students often end-up in the lower tracks. Scholars have pointed several reasons for this. First, the earlier children are tracked, the more their family background will still have an impact on their performance (Dustmann 2004; Horn 2009). Second, both parents and teachers sooner expect children from high status parents to succeed in higher tracks because they are more likely to form as a role-model and to help their children in school. Third, high status parents are more likely to strive for a high track than low status parents to avoid social demotion. This effect is often stronger than the positive value low parents give to social upward mobility (Van de Werfhorst and Hofstede 2007;

Hillmert and Jacob 2010; Le Donné 2014; Van de Werfhorst 2018). Fourth, high status parents often use different child-rearing strategies from low status parents, thereby reproducing ideas of entitlement and privilege, resulting in choosing higher educational tracks. Fifth, and related to the former, to help their children, (upper-)middle class parents more often question school authority, and more effectively mobilize other parents and professionals to contest the judgments of school officials that are disadvantageous for their children (Horvat 2003). Additionally, parents with a high socioeconomic status can use their knowledge, access to professionals, and financial resources to get their children into the right school and academic track and help them stay there (De Graaf 1986; Breen and Goldthorpe 1997). Sixth, and especially important for the age of tracking, parents differ in the far-sightedness of their choices, particularly in early phases of the educational career of their children. Parents and children of more advantaged social backgrounds are more 'myopic' than parents and children of less advantaged social background, causing that inequalities are especially reproduced in early tracked systems (Lucas 2001; Van de Werfhorst 2018).

In line with these reasons numerous studies have shown that early tracking is associated with stronger socioeconomic inequalities in academic performance (e.g., Oakes 1985; Ammermüller 2005; Breen and Jonsson, 2005; Marks 2005; Marks, Cresswell and Ainley 2006; Brunello and Checchi 2007; Horn 2009; Van de Werfhorst and Mijs, 2010; Bol et al. 2014; Le Donné 2014; Chmielewski and Reardon 2016; Burger 2016; Ruhose and Schwerdt 2016; Van de Werfhorst, 2018). Two studies have shown that the association between social background and performance in more selective regimes was already present in primary school (Waldinger 2006; Jakubowski 2010). These studies cast doubt on the idea that tracking amplifies socioeconomic inequalities. However, in contrast to this, Schubert and Becker (2010) found that socioeconomic inequalities did increase in the highly selective Germany system between the fourth grade and age 15. In addition, in a rather ingenious study design, using several experiments in which the degree of tracking was varied, Batruch et al. (2018) recently found that strong versions of tracking cause that teachers find lower tracks more suitable for lower than higher SES pupils. Hence, it still seems likely that early tracking increases socioeconomic inequalities.

In contrast to children with low status parents, children of immigrants tend to have higher aspirations and make more ambitious educational choices (Heath, Rothon and Kilpi 2008; Jackson, Jonsson and Rudolphi 2012; Heath and Brinbaum 2014; Salikutluk 2016). Despite of this, children with a migration background may not benefit from their optimism because they are often not placed in the higher tracks. One important reason for this is that early tracking gives disadvantaged students less time to catch up and get acquainted with the host society and to ultimately demonstrate their abilities (Kirsten and Granato, 2007; Alba and Waters 2011; Spörlein and Schlueter 2016). In general, teachers also may expect less of immigrants due to stereotypes and a perceived lack of support in their environment, and also give immigrants worse advices than natives with similar grades (Kao and Thompson 2003; Lüdemann and Schwerdt 2013). In addition, immigrant parents are less knowledgeable on how to defend their children from a mis-specified advice from teachers (Kao and Thompson 2003). This low track placement may discourage immigrants (and their children) and may provide a worse learning environment than a higher track placement (Kao and Thompson 2003). As a result, immigrant children profit less often from the steeper gains found in higher tracks (Spörlein and Schlueter 2016).

Due to this, early tracking is not found to be beneficial for immigrants. Cross-sectional (e.g., Cobb-Clark, Sinning and Stillman 2012; Griga and Hadjar 2013; Borgna and Contini 2014; OECD 2015) and repeated cross-sectional (Teltemann and Schunk, 2016) designs even seem to suggest that early tracking increases differences between tfirst and second generation immigrants and the native majority. Yet, other studies found few effects, either when applying a mixed effects location design (Spörlein and Schleuter 2016) or a difference and difference design (Ruhose and Schwerdt 2017). All in all, it is safe to assume that early tracking at least has no positive effects in terms of inequalities between the native majority and immigrants and their children.

**VOCATIONAL SPECIFICITY.** Another aspect of educational systems is its vocational specificity. The main feature of vocational education is that it prepares for specific job-related tasks rather than aiming at the provision of general skills that can be used in a variety of settings, like language or mathematic skills. By their focus on job related skills, vocational programs often combine schooling with work experience on the job, for example in the form of apprenticeships. This type of education is often associated with lower levels of education where the emphasis lies on learning work-specific skills. In this respect it has a close connection to tracking. Yet tracked systems greatly differ in the size of their vocational systems. For example, both The Netherlands and Germany tracks students early, but the size of the vocational system is larger in Germany than the Netherlands. Also, in Denmark a large apprenticeship system exists, even though the system tracks students comparatively late. Hence, research on the vocational specificity of education systems focuses mostly on upper secondary vocational education.

The main reason to implement a (large) vocational system is that it prepares students for future employment (Van de Werfhorst and Mijs 2010). Theoretically, the talents of students that opt for a vocational specific education match with the educational program, while the close connection to the labor market ensures a smooth transition to the labor market resulting in low unemployment (Shavit and Müller 1998; Müller and Gangl 2003; Breen 2005; Biavaschi et al. 2013; Rözer and Bol 2019). Furthermore, a vocational system often goes hand in hand with ability tracking, which would make teaching easier and more efficient (Brunello and Checchi 2007). Another advantage of a vocational system is that students who would have dropped out of general education, may find their place in vocational education, while the close connection to the labor market may motivate students to perform well (Wößmann 2009; Tessaring and Wannan 2010). Therefore, vocational education is often seen to function as a ‘safety net’ (Shavit and Müller 2000). Besides, in some countries high grades during pre-vocational education may increase the chances to enter a (prestigious) vocational program, providing a motivation for students to do their best (Soskice 1994). Furthermore, it has been argued that a high level of competition among firms offering apprenticeships can increase its quality (Heckman 2000). These reasons suggest that vocational systems increase the overall performance of students, particularly by lifting up the performance of low performing students.

On the other hand, vocational education may learn job-related skills at the costs of general skills in literacy in mathematics (Kerckhoff 2001; OECD 2010: Heisig and Olga 2015; Hanushek et al. 2017). This is most clearly visible when vocational education is combined with apprenticeships, because during apprenticeships often little attention is paid to academic skills. Moreover, placement in vocational tracks may discourage further learning and may form a barrier for further mobility in the educational system (DiStasio 2017; Buchmann and Park 2009; Shavit and Müller 2000; Bol and Van de Werfhorst 2013).

To the extent that vocational education disperses skills, it may increase existing inequalities because low class children and ethnic minorities more often go to vocational education. This matches with a perspective on vocational education that sees differentiation in educational systems as a way to institutionalize social distances. In this view, a vocational system would reproduce existing inequalities (Bol and Van de Werfhorst, 2013).

Research on the impact of vocational education has, however, mainly focused on its labor market outcomes (e.g., Shavit and Müller 1998; Müller and Gangl 2003; Breen 2005; Biavaschi et al. 2013; Forster and Bol 2018; Rözer and Bol 2019). In comparison, very little research has been done on the impact of vocational education on skills. Hanushek and Wössmann (2011) and Hanushek et al. (2017) show that individuals with general education have on average higher literacy scores than those with vocational education; however, the scores overlap considerably, indicating that both studies share a common support in teaching this skill. Furthermore, Heisig and Solga (2015) show that a greater emphasis on vocational skills in upper-secondary education increases skills for both less- and intermediate-educated adults, but more so for intermediate-educated adults dispersing the gaps between them. Given this dispersion and that children of low socioeconomic parents and immigrants more often end up in vocational education, one would expect that vocational education also increase these types of inequalities. However, Bol and Van de Werfhorst (2013) show that a vocational system does not widen SES gaps, taking into account that these systems also track earlier. Thus, much remains uncertain about the impact of the vocational specificity of an educational system.

## 3. DATA AND METHODS

### 3.1 COMBINING STUDENT ACHIEVEMENT DATA

In this study we combine information from the PIRLS, TIMSS, PISA, and PIAAC to assess inequalities in student achievement across countries, over time and across the life course. As shown in Table 1, these surveys assess mathematics and numeracy, and literacy and reading skills. They sometimes also include information about students' science performance, but to not further extend and complicate the analyses, this information is left out. The surveys include information about migration background, and about socioeconomic background in the form of parents' educational level and number of books in the home. These surveys measure migration background by asking whether the child and his/her parents are born in the country of test. Hence, they do not allow to make a further differentiation, for example between immigrants from Western and non-Western countries.1

*Table 1. Surveys, year of assessment, and available variables.*

| | SURVEY | YEAR | STUDENT | | | | |
|---|---|---|---|---|---|---|---|
| | | | Math | Read | Migration | Education of parents | Books at home |
| **Grade 4** | PIRLS | 2001 | | | | | |
| | | 2006 | | | | | |
| | | 2011 | | | | | |
| | | 2016 | | | | | |
| | TIMSS | 1995 | | | | | |
| | | 2003 | | | | | |
| | | 2007 | | | | | |
| | | 2011 | | | | | |
| | | 2015 | | | | | |
| **Grade 8** | TIMSS | 1995 | | | | | |
| | | 1999 | | | | | |
| | | 2003 | | | | | |
| | | 2007 | | | | | |
| | | 2011 | | | | | |
| | | 2015 | | | | | |
| **Age 15** | PISA | 2000 | | | | | |
| | | 2003 | | | | | |
| | | 2006 | | | | | |
| | | 2009 | | | | | |
| | | 2012 | | | | | |
| | | 2015 | | | | | |
| **Adult** | PIAAC | 2012 | | | | | |

Note: green means that the variables are available in the surveys.

1 Our dataset and method is similar as the one in a previous ISOTIS report (Rözer and Van de Werfhorst, 2018), and consequently the description of the data and method overlaps. The main differences is that in this report we examine the impact of educational institutions on achievement gaps. Furthermore, we updated our dataset by including the most recent rounds of every survey.

Despite being considered the best source for large-scale cross-country comparisons of student achievement, also these surveys have several limitations which have to be mentioned (e.g. Figazzolo 2007; Mortimore 2009; Dämmrich and Triventi 2018). Even while great effort has been put in ensuring cross-national equivalence of measures, the tests might still have some limitations in the translation of concepts and questions across countries and cultures. Moreover, it has been noted that in some countries there might have been a risk of voluntary under-sampling of low-achieving students to perform better in the country comparisons (Mortimore 2009).

Among the countries available we selected the high income countries only – as defined by the World Bank – to make them more comparable. In total we use information from 52 countries[2] and 4,297,271 respondents who are nested in 21 surveys. However, depending on the available measures for countries, the sample sizes are often substantially smaller. This is indicated by the respective analyses.

## 3.2 MEASURES

### 3.2.1 TEST PERFORMANCE

We are primarily interested in differences in mathematics and numeracy, and literacy and reading skills. These skills are assessed with several tests in the respective surveys. The focus of the PIRLS and TIMSS studies – run by the IEA – and of the PISA and PIAAC studies – run by the OECD – is different, however. PIRLS and TIMSS are grade-based assessments aimed to test performance of subjects in the way these are taught in schools, while PISA and PIAAC are age-based assessments founded on the principle to measure life skills that are useful in the further life course. Nevertheless, whether the focus is on school-based skills or life skills, both types of assessments have been used to assess not only the performance of individual children, but also to report about the quality of the educational system in producing human capital, and social and ethnic inequalities therein. Research providing extensive discussion of the similarities and differences between these tests makes us confident that the skills have indeed a common dimension, and are to a reasonable extent comparable (Brown et al. 2007; Hannushek and Wössmann 2012; Lennon and Tamassia 2013; Gal and Tout 2014).[3]

A related difficulty is whether the scores are comparable between countries and over time. In the original data, the individual test scores are standardized such that they have a common mean and standard deviation across all participation countries (PIRLS, TIMSS, and PIAAC) or across OECD countries (PISA). However, these scores are not directly comparable, because the pool of countries on which they are calculated differs between surveys and within surveys sometimes across years. This problem is often overcome by standardizing the scores across all countries in the analyses using z-scores, such that all outcomes are measured on the same scale (e.g., Brown et al. 2007; Jerrim and Choi 2013; Dämmrich and Triventi 2016; Van de Werfhorst 2018). However, as the set of countries differs across surveys and across years, this still results in different scales when all survey-years are

---

[2] These include: Argentina, Australia, Austria, Bahrain, Belgium, Canada, Chile, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hong Kong, Hungary, Iceland, Ireland, Israel, Italy, Japan, Korea, Republic of, Kuwait, Latvia, Liechtenstein, Lithuania, Luxembourg, Macao, Malta, Netherlands, New Zealand, Norway, Oman, Panama, Poland, Portugal, Qatar, Saudi Arabia, Singapore, Slovakia, Slovenia, Spain, Sweden, Switzerland, Taiwan, Trinidad and Tobago, United Arab Emirates, United Kingdom, United States, and Uruguay.

[3] Nonetheless, there may be differences between the tests and within the same test over time. For example, the length of test booklets was reduced in the TIMSS between 2003 and 2007. However, we expect that these differences across tests and within tests over time have few consequences for the (standardized) gaps by socioeconomic and migration background.

pooled. Therefore, we standardize within each combination of country and assessment (and by grade in the multiple-grade TIMSS data).[4] Thus, we examine relative positions within a country-survey year (for a similar approach see Andon et al. 2014; Chmielewski and Reardon 2016; Chmielewski, 2019). We are aware that the measurement models underlying the various assessments differ, and that the assessments were not designed to be compared in the way we do. Yet by standardizing within country-survey year, we are able to associate policies to relative inequalities across SES and migration groups.

### 3.2.2 SOCIOECONOMIC BACKGROUND

One of the most common indicators for socioeconomic background is the education of the parents. Education, however, is coded in a variety of ways in the respective surveys, ranging from having eight answer categories (e.g. PIRLS 2011) to three answer categories (PIAAC). Therefore, we had to standardize them. This is done by creating, as good as possible[5], three equal groups, consisting of respectively the highest, middle and lowest educated parents within a country within a survey. In this way, we treat education as a positional good.[6]

Another way of measuring socioeconomic background is by using the number of books students report there are at home. Three different answer categories are used across the surveys and waves.[7] However, because answer categories sometimes overlap it is hard to come to common categories. Therefore, we decided to standardize the variable within a country-wave, using z-scores.

The education of parents and the number of books at home are often based on student's reports. The question is to what extent these are reliable reports, and to what extent these reports can be used to compare individuals over countries, over time, and of different age. Several studies show that students are well capable to report (father's) educational level (Lien 2001; Jerrim and Micklewrigh 2014). Furthermore, children's reports on the education of their parents seem to be reliable for cross-country comparisons (Jerrim and Micklewrigh 2014). Moreover, at least at older ages (13/15), the reports by students of different ages is comparably reliable (Lien 2001), and probably this is the case for younger children as well (West 2001 looked at descriptions of parents' occupation of 11-year olds). Reports on the number of books, however, are less reliable. Jerrim and Micklewright (2014) show that reports are not reliable and useful for cross-country comparisons, and that there is low agreement between children and parents.

Because parental education is a more common and stronger proxy for student's socioeconomic background than the number of books students have at home, and because the reports for the education of the parents are probably more reliable as well, we use the education of the parents as the main indicator for the student's socioeconomic background. However, we also report on the number of books, because this measurement is available across all surveys and waves.

---

[4] We use first plausible values to calculate the inequalities by socioeconomic and migration background. While this results in accurate slope estimates, this might slightly underestimate the standard errors (typically between 1 and 6 percent). We do not use multiple plausible values though because we combine different tests, which makes it difficult to use them.

[5] By minimizing the sum of absolute differences between the percentages in the cells high, medium and low with 1/3.

[6] There are two attractive alternatives. First, by using the same educational level as a basis (e.g. treating a bachelor always as high). However, this resulted in unrealistic fluctuations in the percentage of respondents with high, middle and low educated parents, even across the waves within countries. Second, standardizing is an attractive alternative, but in practice results that scores represent something different across the waves and surveys. For example, in country A a score of 1 may represent a master program in 2000, but only upper secondary education in 2012.

[7] A with categories 1) 0-10; 2) 11-25; 3) 26-100; 4) 101-200; 5) >; B with categories 1) 0-10; 2) 11-25; 3) 26-100; 4) 101-200; 5) 201-500; 6) > 500; and C with categories 1) none; 2) 1-10; 3) 11-50; 4) 51-100; 5) 101-250: 6) 251-500; 7) >.

### 3.2.3 IMMIGRATION BACKGROUND

In all surveys it is asked whether the student and his/her mother and father are born in the testing country. We make a distinction between first (sometimes also called one-and-a-half)[8] and second generation migrants. Children are considered a first generation when both they and their parents are not born in the country in which the children follow education. Children are considered a second generation migrant when they are born in on the country in which they follow education but their parents not. Children are considered to be part of the native majority when they and their parents are born in the country of test. Unfortunately, there is no information available in a substantial number of survey-waves about the country of origin, so we group together children with a migration background.

### 3.2.4 EDUCATION SYSTEMS AND POLICIES

We include a wide range of educational policies and institutional variables. Table 2 presents an overview about their sources and how they are matched to the achievement data. For all national level variables we applied linear interpolation to reduce the number of missing values. Below we discuss the construction of the variables in greater length.

Table 2. Overview of independent variables.

| Variable | Source | Information used at | | | |
|---|---|---|---|---|---|
| | | Grade 4 | Grade 8 | Age 15 | Adult |
| Educational expenditure | UNESCO | Primary | Secondary | Secondary | Secondary |
| Teacher/pupil ratio | UNESCO | Primary | Secondary | Secondary | Secondary |
| Teacher salaries | OECD | Primary | Lower sec | Upper sec | Upper sec |
| Teaching hours | PIRLS, TIMMS | G4 | G8 | G8/A15 | G8/A15 |
| Teacher educ. (% master) | PIRLS, TIMMS, PISA 2015 | G4 | G8 | G8/A15 | G8/A15 |
| Teacher exp. in years | PIRLS, TIMMS, PISA 2015 | G4 | G8 | G8/A15 | G8/A15 |
| Pre-prim. enrol. (at age 3) | World Bank | | | | |
| Compulsory edu.: start age | Reform data, Eurydice | | | | |
| Compulsory edu.: end age | Reform data, Eurydice | | | | |
| Tracking age | Reform data, Eurydice, etc. | | | | |
| Vocational spec. (at age 16) | UNESCO | | | | |

Notes: G4 = Grade 4; G8 = Grade 8; A15 = Age 15.

**EDUCATIONAL EXPENDITURES.** Education expenditures is derived from UNESCO and is measured as the percentage of GDP spend on education.[9] Information on primary school expenditures are matched to students in grade 4, and information on secondary school expenditures to children at grade 8 and age 15 and older.

[8] First-and-a-half generation migration is often used to describe individuals who are migrated as children or adolescents. Hence, this term might be most accurate for our “first generation” migrants below say the age of 18. However, because we measure them similar as adolescents who first arrived, and to not further complicate the terminology, we stick to “first generation migrants”.

[9] We relied on information from Unesco instead of the World Bank and the OECD because it includes most data. A further distinction between lower and upper secondary education could be made. However, these variables include many missing values (42 and 41 percent).

**TEACHER/PUPIL RATIO.** Class size is measured as the average pupil/teacher ratio in a country. This is derived from UNESCO. Again data on primary education is matched to grade 4 children, and on secondary education to children in grade 8 and age 15 and older.

**TEACHER SALARIES.** Information about the salaries of teachers is derived from the OECD. To make them comparable across countries we divided them by the GDP per capita. In some countries teachers earn more than twice than the GDP per capita (e.g. in Germany and Denmark in the early 2000s), while in others they earn less than the average GDP per capita (e.g. in Iceland and France in the early 2010s). Information about the salaries of teachers in primary education is matched to respondents in grade 4, information about the salaries of teachers in lower secondary education is matched to respondents in grade 8, and information about the salaries of teachers in upper secondary education is matched to respondents age 15 and older. Note that we only include baseline differences in remuneration, not data on incentivized teacher pay.

**TEACHING HOURS.** Information about the total teaching hours a year is derived from the PIRLS and TIMMS data. Students at grade 4 get the average instruction hours per year (divided by 100) in their country for grade 4 (based on the PIRLS and TIMMS grade 4 data), students at grade 8 and older get the average instruction hours in their country for grade 8 (based on the TIMMS grade 8 data).

**TEACHER'S EDUCATIONAL LEVEL.** A first indication of the quality of the teachers is derived by taking into account the level of education of the teachers. This information is derived from PIRLS, TIMSS, and PISA (2015) data. Answer categories differed over surveys (e.g. PIRLS) and survey-years (e.g. PIRLS 2001), but in every survey it was asked whether teachers had a master degree. Therefore, for every country-year we calculated the percentage of teachers with a master degree. For students at grade 4 we use information from grade 4 (PILRS and TIMMS), for students at grade 8 we use information from grade 8 (TIMMS), and for students age 15 and older we use information from grade 8 and if available from age 15 (PISA 2015). If both information at grade 8 and age 15 is available we used the average of both measures.

**TEACHERS' EXPERIENCE.** A second indication of the quality of teachers is their years of experience. This information is also derived from the PIRLS, TIMMS, and PISA (2015) data. They are matched to the respondents similar as the educational level of teachers: for students at grade 4 we use information from grade 4 (PILRS and TIMMS), for students at grade 8 we use information from grade 8 (TIMMS), and for students age 15 and older we use information from grade 8 and if available from age 15 (PISA 2015).

**PRE-PRIMARY EDUCATION.** Information about enrolment in pre-primary education is derived from the World Bank. We use the percentage of students who are enrolled in pre-primary education when the respondents where 3 years old, because around this age they attend pre-primary education. For example, when 35 percent of all children went to pre-primary education in 2000, students who were three years old in 2000 get this score.

**COMPULSORY SCHOOING AGE.** We relied on several sources to construct a variable of compulsory schooling age. Information till 2000 comes from the reform dataset of Braga, Checchi and Meschi (2013). As a second source we relied on the Eurydice reports on “compulsory education in Europe” and “the structure of educational systems”. From 2007 these report annually report on the compulsory age in most European countries. For the period between 2000 and 2007 we searched the internet whether there were any reforms. In doubt, we included a missing value. It appeared that many reforms focused on a broader change of the educational system, with the side effect that the compulsory age shifted a year in the official report (e.g. in Poland 2016). Students get the score that applied to their cohort. For example, when students born in 2000 in The Netherlands had to attend school from 4 years old, they get this score.

**AGE OF TRACKING.** We put great effort in getting information about tracking from a wide variety of countries over a substantial time period. We use seven sources to construct a measure of tracking. Our primary dataset is the Eurydice dataset about how educational systems are organized. We complement these data with the reform dataset of Braga, Checchi and Meschi (2013), and if data is missing with those from the International Bureau of Education (Unesco 2007, 2012), Education at a Glance (OECD 1998, 200-2018), and a paper of Österman (2017). If data was incompatible between sources we checked them with other sources, most often the websites of the official agencies in the respective countries or country specific reports. Students get the score that applied to their cohort. For example, when students born in 2000 in The Netherlands where tracked at age 12, they get this score.

**VOCATIONAL SPECIFICITY.** Following Bol and Van de Werfhorst (2013), the prevalence of vocational enrolment indicates the percentage of students who are enrolled in upper secondary vocational programs. These data are derived from UNESCO. Information is used when individuals where 16 years old. For example, when in 2016 40 percent of all students enrolled in upper secondary education followed a vocational track students who were then 16 (and thus are born in 2000) get this score.

### 3.2.5 CONTROL VARIABLES

GDP per capita in purchasing power parity in 100k dollar is derived from the Penn World Trade tables. We took the log to control for possible diminishing returns of wealth. This is matched per country-year. Thus, when in The Netherlands in 2000 the GDP per capita was 50k, Dutch respondents in 2000 get this score.

## 3.3 METHOD

We rely on two methods to analyze our data:

1. Multilevel regression models on each survey (i.e. the PIRLS, TIMMS, PISA and PIAAC), comparing inequalities between countries and inequalities within countries over *cohorts*.[10]
2. Pseudo panel analyses, following cohorts over their *life course*.

### 3.3.1. MULTILEVEL MODELS WITH BETWEEN AND WITHIN NATIONAL LEVEL VARIABLES

In a first step we describe inequalities between countries and within countries over cohorts. Therefore, we estimate multilevel models with between and within national level variables. We calculated the group mean for each national level variable and the deviation from that mean, respectively. These models are similar to hybrid multilevel models (Allison 2009; Schunck 2013).

We regress the achievement scores of students on their socioeconomic or migration background (showing the educational inequalities) and add a cross-level interaction between these inequalities and the several educational institutions (group mean centered and the deviations from these means).[11] The cross-level interactions indicate the extent to which the inequalities by socioeconomic and migration background differ between (average values of) countries and (deviations from these averages) within countries. We control for age, gender and GDP. We nested students within countries and control for year dummies.

Although within effects rule out all stable between country differences (such as cultural differences), these effects may still be biased by unobserved heterogeneity. This may also be the case with the cross-level interactions (Giesselmann and Schmidt-Catran 2018). In practice, policies are often implemented together and by not controlling for these other policies our interactions may pick up the effects of these other policies. However, controlling for several policies at the same time greatly reduces our sample size and is therefore not done. Furthermore, the cross-level interactions do pick up both within and between country variation, and thus are not purely a between and within estimator.[12] In addition, as these interaction mix a between and within effect it might still be affected by both within and between country heterogeneity (Giesselman and Schmidt-Catran 2018).

In the literature on hybrid models, the individual level variables in the cross-level interaction often have no random effect, which is also not necessary to perfectly mimic fixed effects models (e.g. Allison 2009; Giesselman and Schmidt-Catran 2018). However, in the multilevel literature this is often recommended (Heisig and Schaeffer 2017). Because models including these random effects regularly not converged we excluded them in all models for consistency.

### 3.3.2. AUTOREGRESSIVE MODELS ON PSEUDO PANEL DATA

In a second step we modelled how inequalities developed when children get older. While we have no genuine panel data to track students through the different educational systems, it is possible to track

[10] Note that this also means that we compare scores across time, as new cohorts enter the data over time.
[11] Therefore, we should also take the mean and demeaned values of socioeconomic and migration background.
[12] Therefore, we should also have taken the mean and demeaned values of migration background and socioeconomic status. However, this would also affect the interpretation of the results (e.g., what happens when both the national level variables and the percentage of individual in a country with a migration background change over time).

cohorts of students over time.

A challenge in creating a pseudo panel is to group individuals within cohorts, such that they can be followed over time, and hence when they become older. Following the survey from which we have most data, i.e. the TIMSS, we created cohort groups of 4 years. Table 3 provides an overview of cohort-age combinations and the surveys we use to estimate the respective inequalities for mathematics. Similar cohort-age combinations are used for reading and literacy. Noteworthy, we split the PIAAC in age groups from 16-17 till 30-33 to keep following the cohort of the TIMSS. Furthermore, because the TIMSS is held every four years, while the PISA is held every three years, two waves of PISA fall within one cohort. Consequently, when looking at changes within cohorts these two surveys are merged.

Based on these pseudo panel data we can describe inequalities in mathematics and numeracy, and literacy and reading skills within a cohort over the life course. For example, we can start tracking the cohort born between 1983 and 1986 in 1995 when they are in grade 4, and can observe how inequalities develop when they are in grade 8, are 15 years old, and when they are approximately between 26 and 29 years old (see Table 3).

*Table 3. Cohort/age combinations for mathematics and the respective surveys.*

| **GRADE / AGE** | **SURVEY** | **BIRTH COHORT** | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | 1979-1982 | 1983-1986 | 1987-1990 | 1991-1994 | 1995-1998 | 1999-2002 | 2003-2006 |
| Grade 4 (~10y) | TIMSS | | 1995 | | 2003 | 2007 | 2011 | 2015 |
| Grade 8 (~14y) | TIMSS | 1995 | 1999 | 2003 | 2007 | 2011 | 2015 | |
| Age 15 | PISA | | 2000 | 2003/2006 | 2009 | 2012 | 2015 | |
| Age 16-17 | PIAAC | | | | | 2012 | | |
| Age 18-21 | PIAAC | | | | 2012 | | | |
| Age 22-25 | PIAAC | | | 2012 | | | | |
| Age 26-29 | PIAAC | | 2012 | | | | | |
| Age 30-33 | PIAAC | 2012 | | | | | | |

Note: some surveys took place in several years.

A common approach for analysing panel data is to use fixed effects or autoregressive models in which individuals are followed over the life course. With pseudo panel data, these individual records, however, are not available. In the pseudo panel literature (e.g. Deaton 1985) it is suggested that cohorts can replace the records of individuals to come to accurate predictions.

An assumption is that the average of the cohorts are unbiased.[13] A common solution is to ignore this assumption when the number of respondents within a cohort is large. As a rule of thumb, 100 respondents are often regarded as sufficient (e.g. Deaton 1985), but Devereux (2007) showed that sometimes more than 2000 respondents within cohorts are needed to have unbiased estimates. For the PIRLS, TIMSS, and PISA, we almost always have more than 2000 cases per cohort/survey, but for the PIAAC sometimes “not more” than 300.[14]

To tackle this problem, we use a two step-approach – or estimated dependent variable approach –, in which we directly include the uncertainty in the estimates in the model – that form the “pseudo panel data” – by using weights (Hanushek 1977; Lewis and Linzer 2005). Advantages of this approach

[13] This assumption is especially critical when lagged affects are included, as this doubles the problem because this assumes that the lagged effect is an unbiased estimate.
[14] A related difficulty is that it is assumed that respondents within cohorts represent the same population over time. In our study, this is not the case because we, for example because at some observations respondents in a cohort can be born (on average) in a slightly different year. However, because changes across (a small amount of time) are small (see the result section), we expect that this bias is limited.

is that it uses all information in the data, avoids the heteroscedasticity of the error terms of the model and predicted values, is more efficient than traditional OLS/WLS models on a pseudo panel data set (Lewis and Linzer 2005), and is similar to a multilevel slope as outcomes model (De Leeuw and Kreft 1986), but less computationally intensive (Jaime-Castillo 2015).

In the first step, a linear regression model is run for every survey-cohort-year, in which the mathematics, science and reading and literacy skills are predicted by either parental education, the number of books at home, or migration background. In these models we control for the age and gender of the respondents. Additionally, we control for migration background when estimating the effects of parental education and the number of books, and we control for parental education and the number of books when we estimate the effects of migration background. The estimates and standard errors of the estimates are saved. As these are bivariate models the estimates represent the average differences in skills between children with high, middle, and low educated parents, with few and many books (1 standard deviation difference in the number of books), and natives and first and second generation migrants. The variances represent the uncertainties in the estimates.

In the second step, estimates obtained in the first step become the dependent variable to be explained by a set of aggregate predictors. Weights are used to take into account that the data are estimates. These weights are based on two components: i) standard errors of the estimates in step 1, and ii) the sum of variance of the residuals in a model in which the estimates are predicted by the other variables in the model, corrected for the fact that this includes the variance of step 1. The second component is calculated as follows:

1. $\sigma^2 = \frac{E(\sum_i v^2) - \sum_i w^2 + tr(x'{x}^{-1} x' g x)}{n-k}$

In which $E(\sum_i v^2)$ is the sum of squared residuals from a model in which y (from step 1) is predicted by the variables in the final model, $\sum_i w^2$is the sum of variances from step 1, tr is the trance function, x represent the variables, g is a diagonal matrix with the variance of the predicted dependent variables as the ith diagonal element, and n and k present the number of variables and cases, respectively (see Jaime-Castillo, 2015 for a full explanation). The weights are calculated by dividing 1 over the squared root of component 1 ($w_i^2$) + component 2 ($\sigma^2$):

2. $weight = \frac{1}{\sqrt{w_i^2 + \sigma^2}}$

Now equation 1 can be estimated with OLS regression, with the predicted values replacing the dependent variable, and including the weight from equation 4.

Autoregressive models are used in the second step to estimate how the estimates (representing the inequalities by parental and migration background) vary by the respective policies. These models are similar to cross-lagged models. Therefore, estimates at an older age (the dependent variable) are predicted by estimates at a younger age (the lagged effect) and the policies at a younger age. We include country-cohort fixed effects, such that we look at the changes within cohorts in a country. This is written out in equation 3:

3. $estimate_{\ t=i} = a + b_1 estimate_{t=i-1} + b_2 national\ policy_{t=i-1} + \ b_3 \log GDP + u + e.$

In which the estimates present the estimates from the first step, t presents the age/grade groups, u is the country-cohort fixed effect, and e is the error term. Note that this equation applies to the second step of our equation, thus for the data on the country-cohort-year level. For example, the performance gap between children with parents with few and many books at home at grade 8 (i.e. the estimate of the number of books in reading in grade 8 from step 1) is predicted by the educational expenditures in primary education (e.g. in grade 4), while we control for the performance gap in grade 4 and GDP per capita, and include country-cohort fixed effects. This shows how the educational expenditures in grade 4 affect the change in inequalities by the number of books at home between grade 4 and grade 8. Obviously the weights from equation 2 are applied.

Linear interpolation – of the estimates, weights, and other variables – is used to deal with potential missing values and the unbalance of the panel. Missing values in the pseudo panel dataset do occur by design often after age 15, because we have only information from one wave of the PIAAC data in which adults above 15 years old are included. Especially for the group above 21 years of age this can be potentially problematic because including them implies missing values between age 15 and the age they have (e.g. age 21). Therefore, to control for the impact of the linear interpolation we compared the estimates between a model with linear interpolation, without linear interpolation, and a model in which only respondents below 21 years of age where included. Results for the estimates until age 21 where similar for the model excluding individuals above 21 and a model with linear interpolation, but differed for the model without linear interpolation. This is an important indication that the model with linear interpolation is more accurate and reliable than a model without.

### 3.3.3 INTERPRETATION AND RELIABILITY OF BETWEEN, WITHIN AND PSEUDO PANEL MODELS

Typically within-estimates are seen as more reliable than between effects because they rule out all between country variance (Allison 2009). However, they only look at direct changes within a policy, and not at long term effects. Therefore, between effects are often also considered important (e.g. Schröder 2016). In addition, and specific to our data, there are measurement errors for the dependent and independent variables that might switch over time and between surveys. For example, we match rather unspecific data on educational expenditure in secondary education to data on grade 8 and older. This affects the accuracy of comparing the effects over time, and increases the risk that we pick up the correlation of measurement bias at two different time points in the within and pseudo panel models.

In general, we consider within effects as a slightly better test, but take the between effects into account as well. Pseudo panel models allow to follow cohorts over the life course and may come closest to true panel data. This is a strong advantage. We apply autoregressive models to these data. Within these models we look at changes in the dependent variable, but not at changes in the independent variables. Consequently, we do not rule out all between country variance. Overall, outcomes may be most reliable when all estimates point in the same direction.

Table 4 presents how we value each method for each policy by the grade and age for which we can use each method. Educational expenditures, pupil/teacher ratio's, teacher salaries, teaching hours, and the education and experience of the teachers directly affects the students in a classroom. Therefore, we expect these policies to affect children most when they are (certainly) still in school, that is in grade 4, grade 8 and age 15. As explained above, we consider the within and pseudo panel estimates (slightly) more reliable than the between-estimates. However, especially for those measures in which we have inaccurate measures between grade and measurement, such as matching secondary education data to grade 8 and older, the within-estimates might be less reliable. Pre-primary education and the start age of compulsory education apply most when they start school, and thus in our data as early as possible, that is in grade 4. With the pseudo panel data, the earliest transition we observe is from grade 4 to grade

8, and we have thus no information what affects the gaps at grade 4 (or ideally even earlier). Therefore, in this instance, the within-estimates for grade 4 (i.e. comparing grade 4 across survey years) are most reliable followed by the between-estimates for grade 4 (i.e. comparing countries within survey years) although earlier data was thus highly preferred (and used in a complementary ISOTIS report of Skopek et al 2018). Finally, the end age of compulsory education, the tracking age, and the vocational specificity do not apply yet at grade 4.

*Table 4.*

| | **Between** | | | | **Within** | | | | **Pseudo** | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | G4 | G8 | A15 | Adult | G4 | G8 | A15 | Adult | G8 | A15 | Adult |
| Educational expenditure | | | | | | | | | | | |
| Pupil/teacher ratio | | | | | | | | | | | |
| Teacher salaries | | | | | | | | | | | |
| Teaching hours | | | | | | | | | | | |
| Teacher educ. (% master d.) | | | | | | | | | | | |
| Teacher experience | | | | | | | | | | | |
| Pre-primary school enrol. | | | | | | | | | | | |
| Compulsory edu.: start age | | | | | | | | | | | |
| Compulsory edu.: end age | | | | | | | | | | | |
| Tracking age | | | | | | | | | | | |
| Vocational specificity | | | | | | | | | | | |

Notes 1: G4 = Grade 4; G8 = Grade 8; A15 = Age 15.
Notes 2: dark green=strongest test; green=moderate test; light green=weakest test; grey=does not apply or very weak test.

# 4. RESULTS

## 4.1 PERFORMANCE GAPS BETWEEN COUNTRIES, COHORTS AND OVER THE LIFE COURSE

### 4.1.1 PERFORMANCE GAPS BETWEEN COUNTRIES

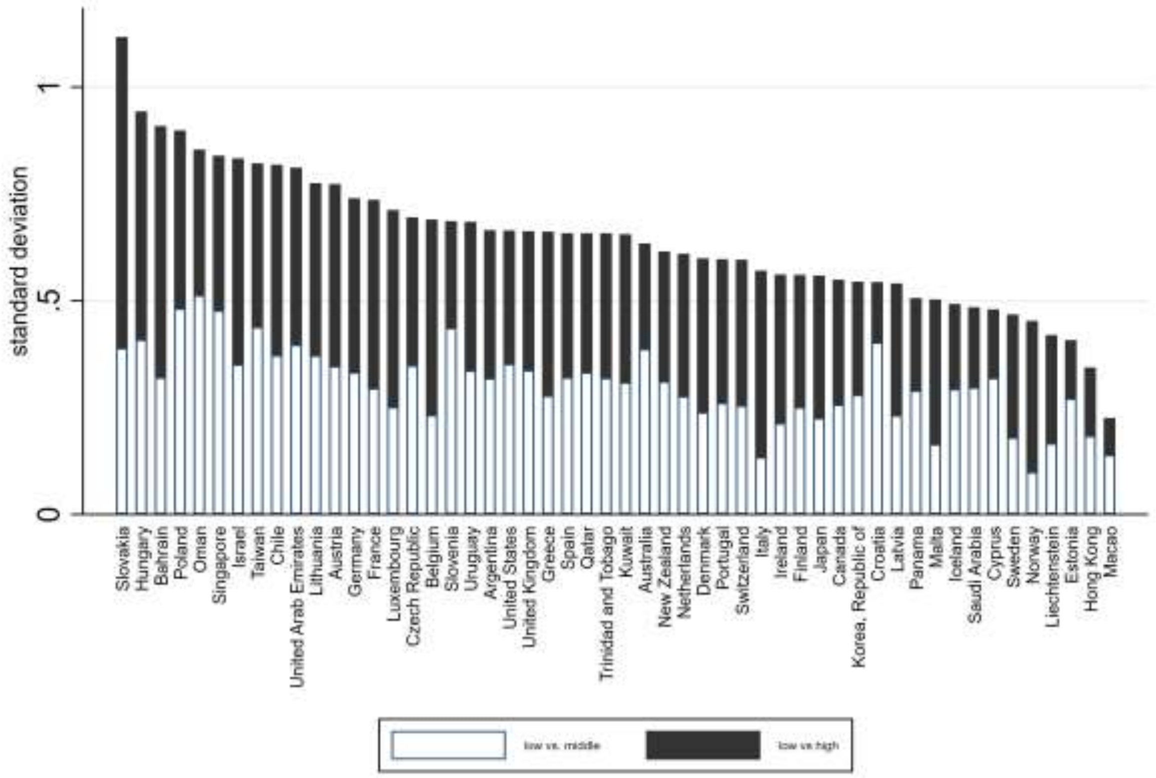
Before we test the association between inequalities and national policies, we describe the gaps in reading and mathematics performance by socioeconomic and migration background over countries, over time, and across the life course. We start with describing the gaps across countries. Therefore, we calculated for each country the average differences between respondents by parental education, number of books at home, and migration background for each survey year (e.g. PIRLS 2001, or PISA 2012) and afterwards took the mean of these gaps. As the scores for each country are based on different survey-years they are not completely comparable. Nevertheless, they give a good indication in which countries the inequalities are larger – especially because most countries participated in most survey-years. Figure 2 presents the results.

Panel A presents the gaps in reading and mathematics performance by parental background. As one would expect, the gaps in reading and mathematics performance between children with low and high educated children are larger than those between children of low and middle educated parents. Gaps in reading performance range between 0.23 standard deviation between children of low and high educated parents in Macao to 1.12 standard deviation in Slovakia. Gaps for mathematics performance are smallest in Macao (.23 standard deviation), and the largest gap can be found in Hungary (1.13 standard deviation). Performance gaps are especially large in many Eastern European countries such as Hungary and Poland and Slovakia, and in financially unequal countries such as Israel and Chile. In contrast, gaps in reading and mathematics performance are relatively small in the egalitarian Scandinavian countries and in China's city states such as Macao and Hong Kong.

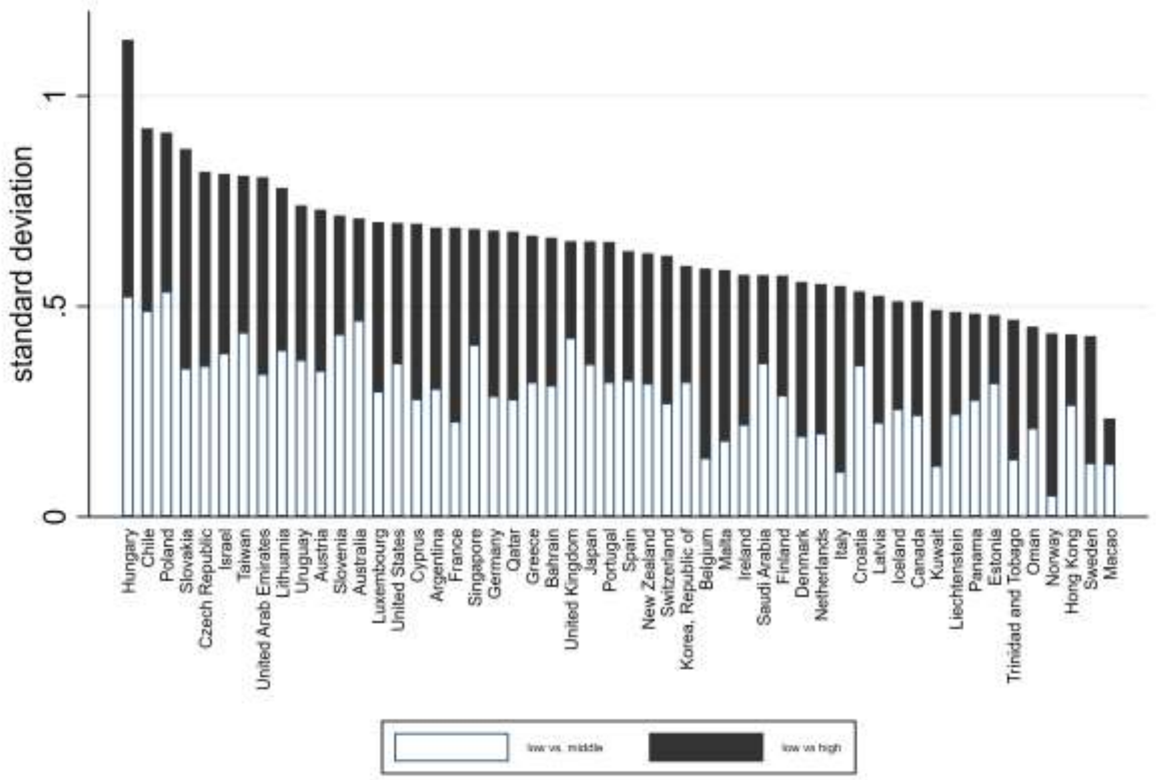
Panel B presents the performance gaps by the number of books at home. As the number of books at home is standardized, these gaps represent the effect of having 1 standard deviation more books at home. Gaps in reading performance range between 0.06 standard deviation in Kuwait and 0.38 standard deviation in Hungary. Gaps in mathematics performance range between 0.08 standard deviation in Kuwait and again 0.38 standard deviation in Hungary. Because of different scales, gaps in reading and mathematics performance by parental education and books at home are difficult to compare. Nevertheless, the impression is that they are larger by parental educational than by the number of books at home. Even when we take a two standard deviation difference in number of books at home, the gaps are smaller by the number of books at home than those between children with low and high educated parents. Performance gaps are again large in the Eastern European countries, but also in the Central European countries such as Germany, Austria and Switzerland. In contrast, gaps are relatively small in the Gulf region, the several Asian countries in the data, and in Trinidad and Tobago (which is the only African country in the data).

*Figure 2. Educational inequalities across high income countries (in standard deviations)*

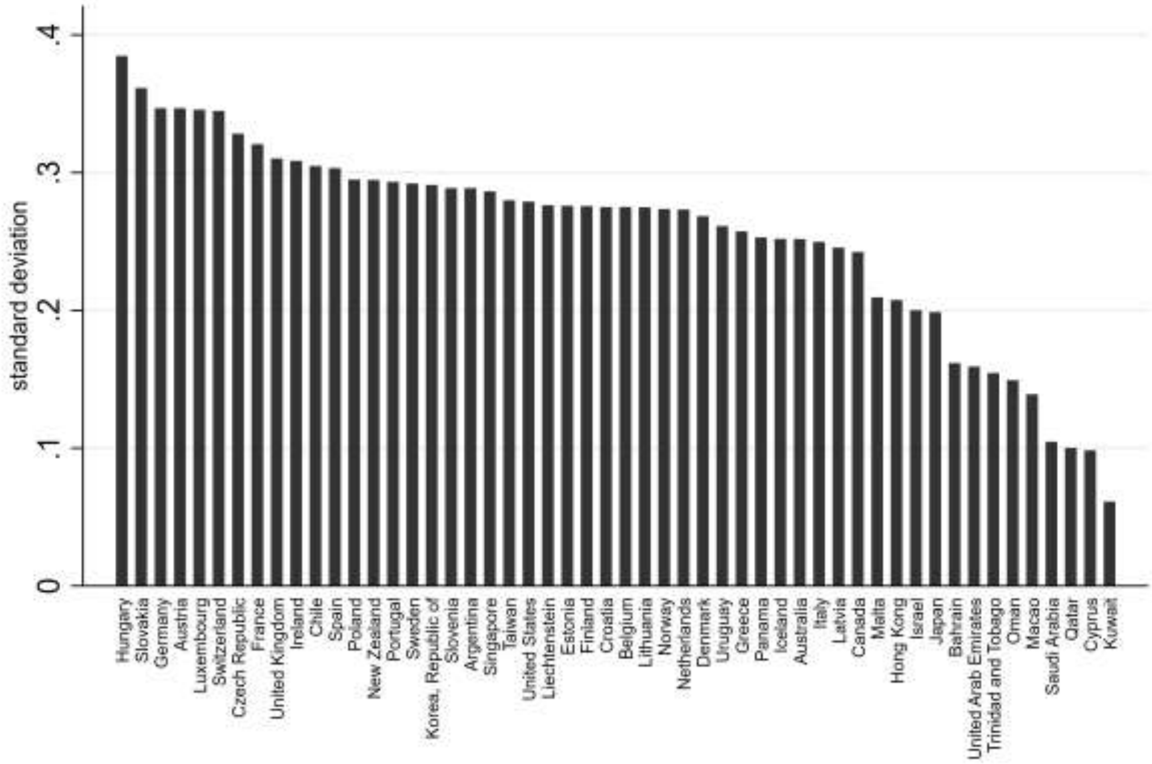
Panel A1: by parental education (in reading)



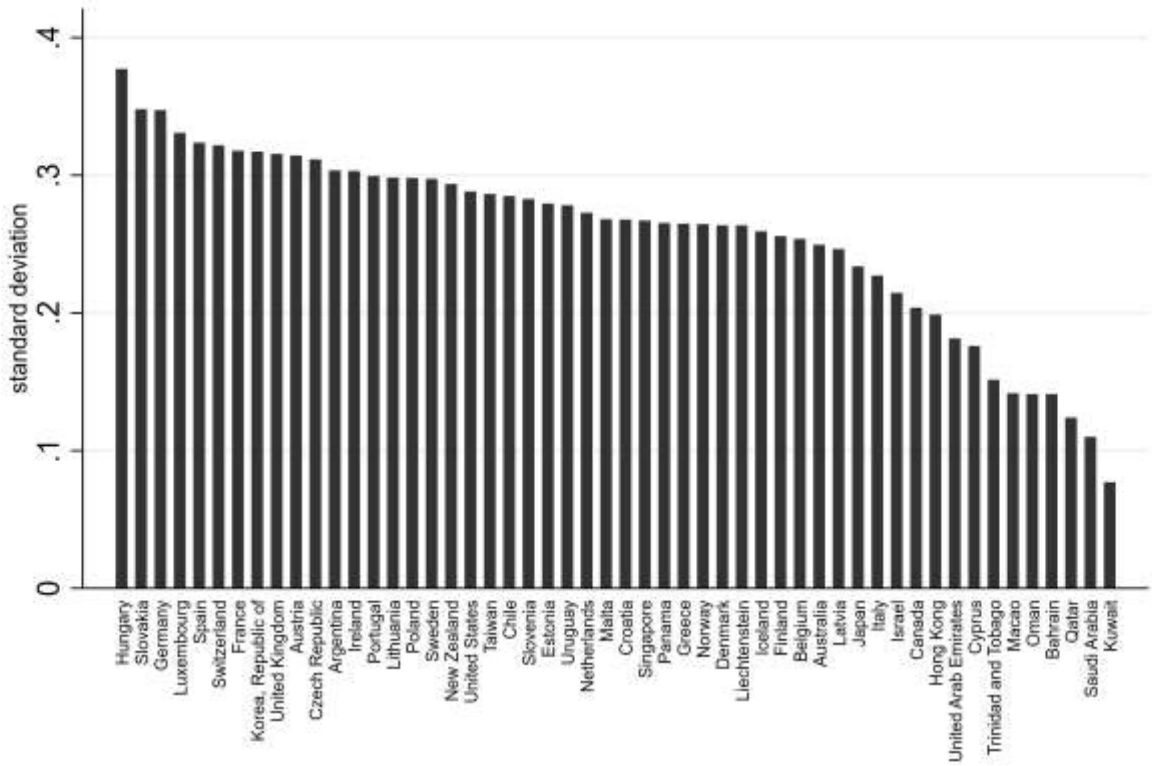

Panel A2: by parental education (in mathematics)



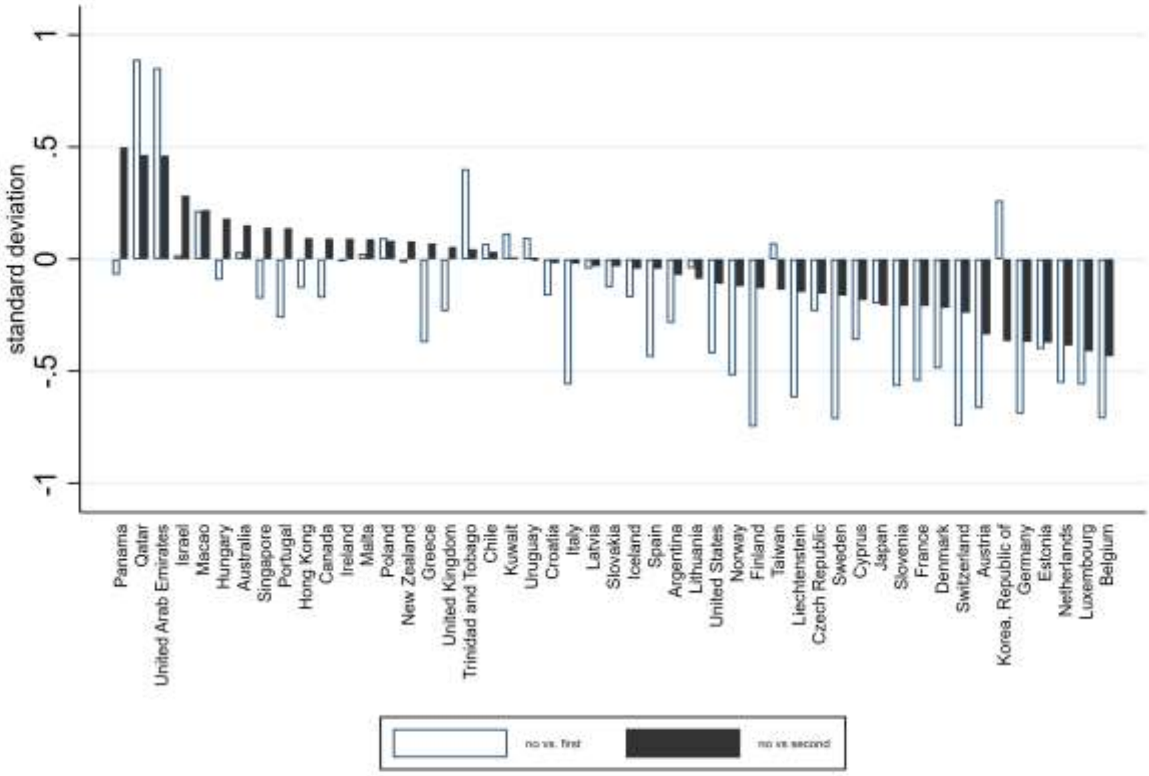
Panel B1: by number of books at home (in reading)



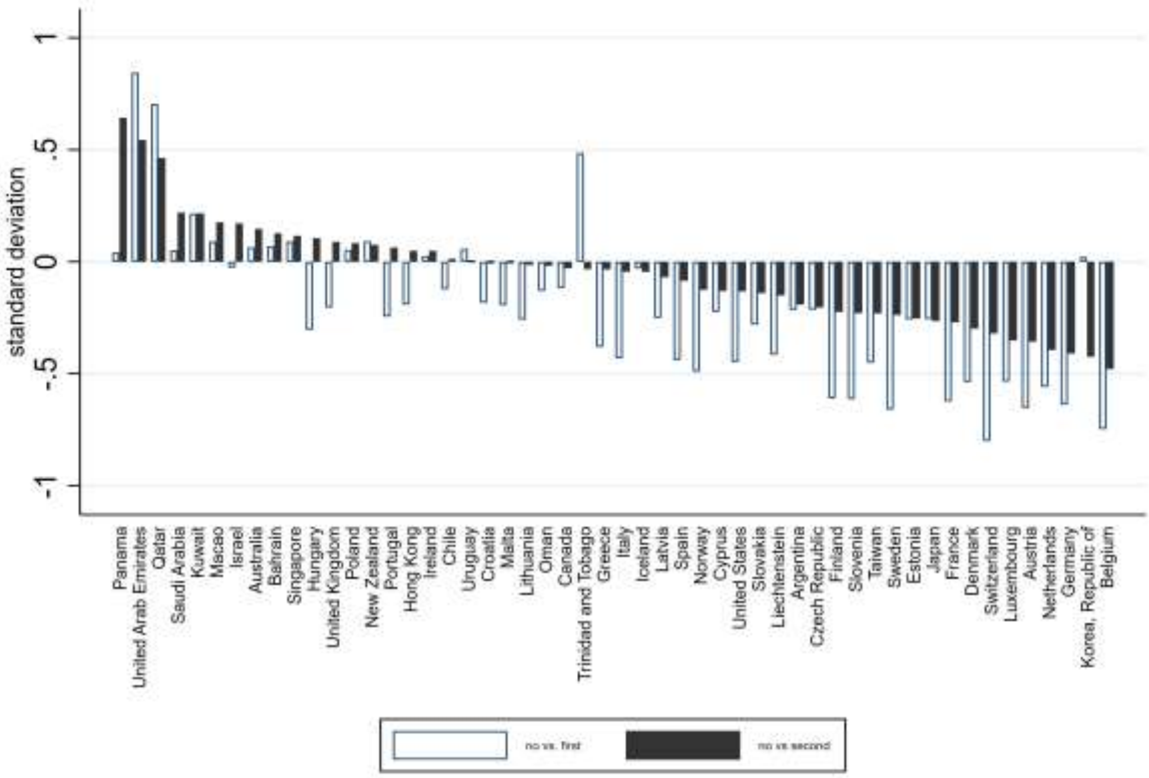

Panel B2: by number of books at home (in mathematics)



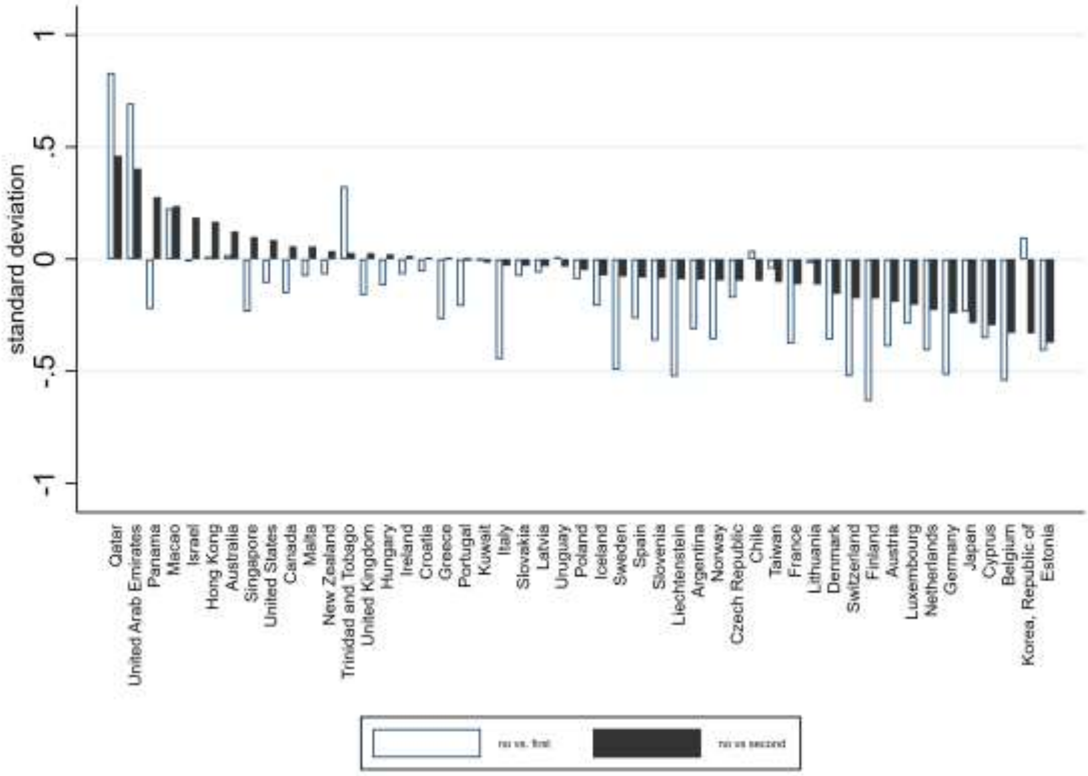
Panel C1: by migration background (in reading)



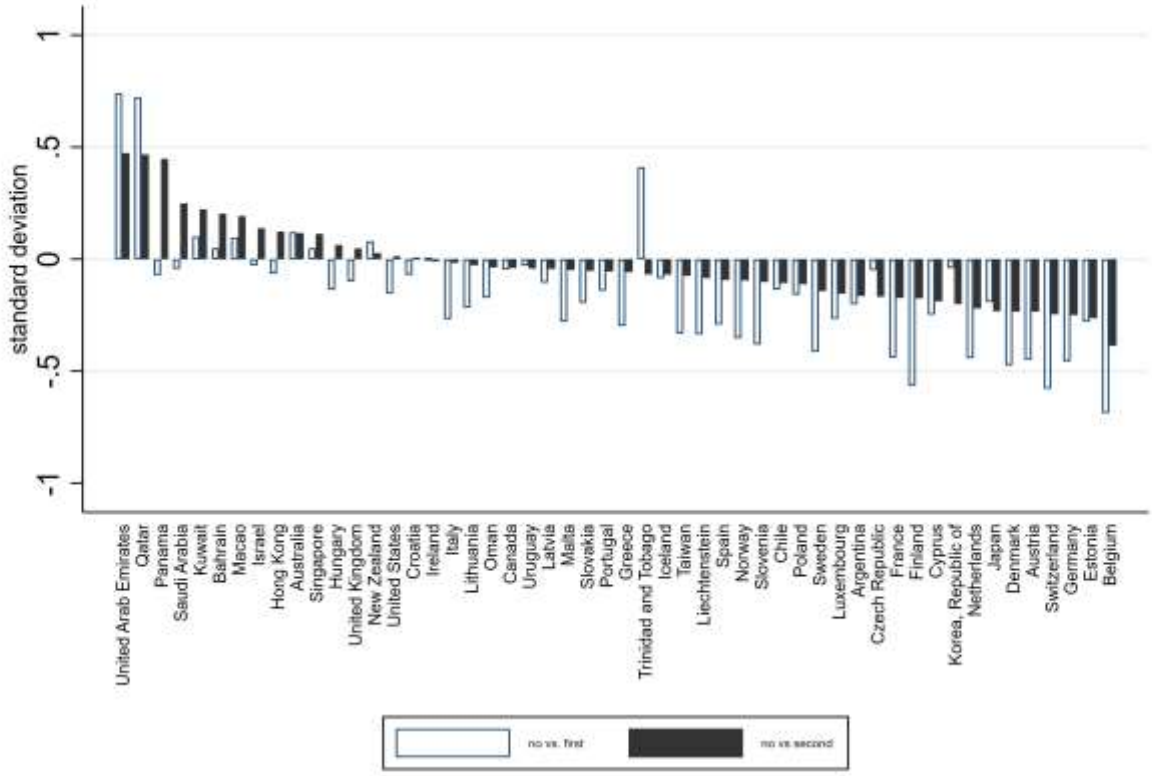

Panel C2: by migration background (in mathematics)



Panel D1: by migration background, controlled for parental education and books at home (in reading)




Panel D2: by migration background, controlled for parental education and books at home (in reading)



Panel C presents the performance gaps by migration background (without controls for SES). Results are split by first and second generation migrants because effects sometimes go in the opposite direction. For example, in Korea first generation migrants perform better than the native majority in reading, while second generation migrants perform considerably worse than the native majority. In case of Korea this has to do with the small number of first generation migrants in some survey-years. Pooling all data directly, allowing that some survey-years with more data weight stronger, diminishes in case of Korea the difference. In other instances results remain the same, indicating that they reflect true differences between first and second generation migrants. The most plausible reason for these differences is that in these countries first and second generation migrants are different immigrant groups coming for instance from different origin countries.

While it is often assumption that children with a migration background perform worse in education than those without, the results show that in many countries migrants perform better. In selective immigration countries such as Australia, and in countries that attract many high educated migrants such as Panama and Ireland, children of immigrants often perform better in reading and mathematics than children of the native majority. In contrast, in many Western and Central European countries that attracted (in the past) many guest workers and refugees, such as in Belgium, Germany and The Netherlands, first and second generation migrants score considerably worse than the native majority. Yet, within these countries gaps in reading and mathematics performance by parental education are still often at least of the same size and sometimes even twice as large than those by migration background. While a part of the differences by migration background is explained by taking socioeconomic background into account, the overall effects of migration background provided very similar country rankings as without controls for parental education and the number of books in the household (see Panel D).

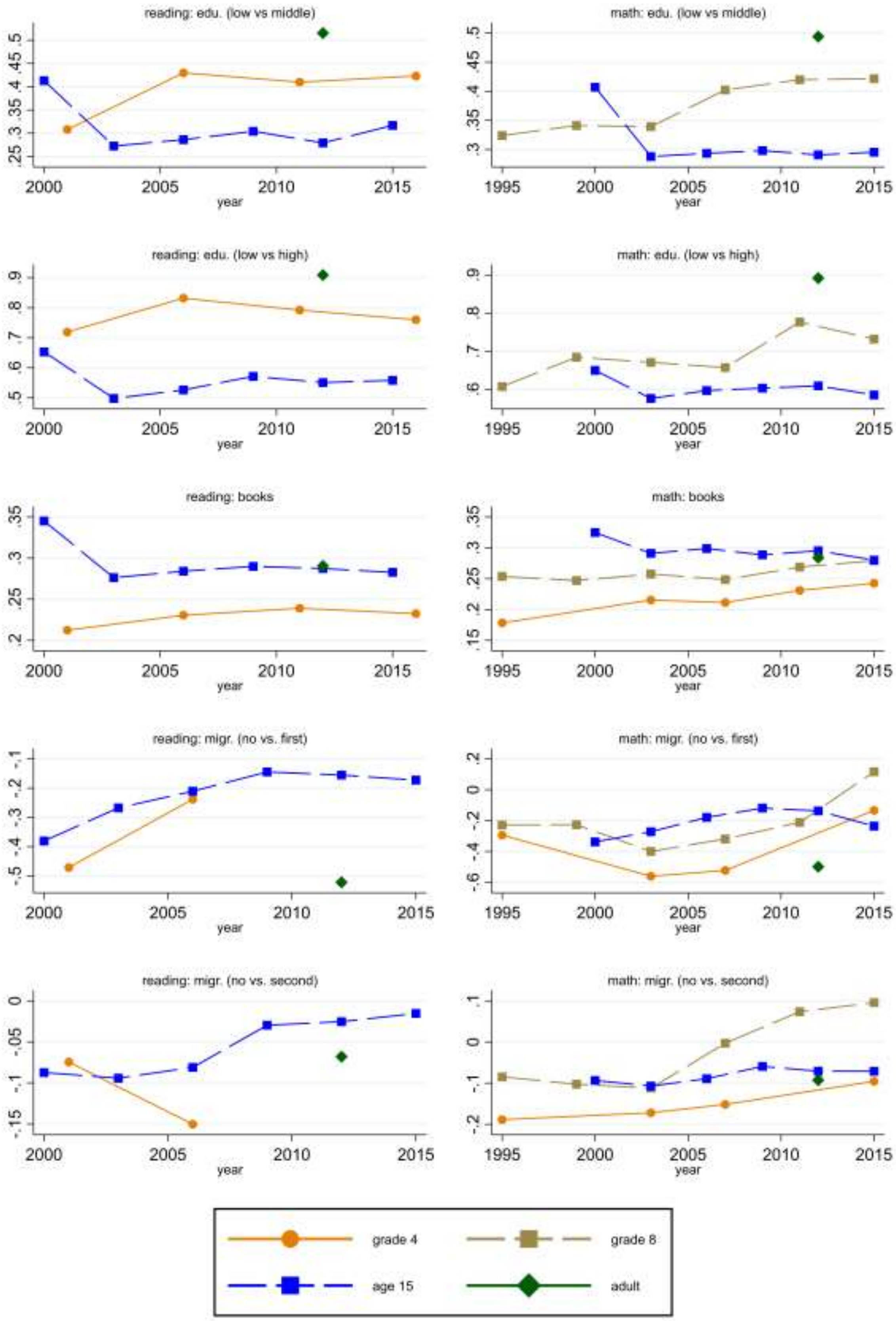
### 4.1.2 PERFORMANCE GAPS WITHIN COUNTRIES OVER COHORTS

Next we describe how the gaps in reading and mathematics performance changed within countries over cohorts. Therefore, we again calculated for each country the average differences between respondents by parental education, number of books at home, and migration background for each survey-year (e.g. PIRLS 2001). Afterwards, we plotted how the performance gaps developed over cohorts by survey. Figure 3 presents the results. The left column presents the results for reading, and the right column those for mathematics.

The first two rows present the results by parental education, comparing children with low educated parents with children with middle high educated parents, respectively. Gaps in reading and mathematics performance are naturally larger between children with low and high educated parents than between children with low and middle educated parents. The gaps are typically twice as large. Moreover, they are almost always larger the older the children are. At adulthood the differences between children with low and middle educated parents is around 0.5 standard deviation, and those between children with low and high educated parents around 0.9 standard deviation. These gaps can be called large.

For adulthood we have only one time point, but for the other grades/ages we can assess how the inequalities changed over time. The gaps in reading performance at grade 4 increased slightly between 2000 and 2006, but remained stable afterwards; gaps in mathematics performance at grade 8 remained stable between 1995 and 2003, increased until 2012, and stayed stable to 2015; and gaps in reading and mathematics performance at age 15 declined between 2000 and 2003, and remained stable afterwards.

*Figure 3. Educational inequalities over cohorts by survey (in standard deviations)*



The third row presents the results by the number of books at home. Although less pronounced, trends are in line with those by parental education; for grade 4 and grade 8 the gaps increased between 1995 and 2015, while for age 15 between 2000 and 2003 there is a slight decrease in the gaps followed by a stable trend.

The fourth and fifth row presents the results by migration background, comparing natives with first and second generation migrants, respectively. In contrast to the scores by parental education and the number of books at home, the higher the line, the better first and second generation migrants perform compared to the native majority. Trends differ by grade and between first and second generation migrants. For grade 4, gaps in reading performance declined between the native majority and first generation migrants declined between 2001 and 2006, while they increased between the native majority and second generation migrants. And gaps in mathematics performance at grade 4 between the native majority and first and second generation migrants declined between 2000 and 2009, and increased slightly between 2009 and 2015.

For grade 8, we only information about mathematics performance. The gaps between students with and without a migration background increased between 1995 and 2003, but declined steeply between 2003 and 2015. In 2015 children of immigrants even perform slightly better on average in mathematics than the native majority. There is a similar trend for the gaps in performance in reading and mathematics at age 15, with migrants gradually performing better compared to natives between 2000 and 2015. For adults we cannot observe a trend because we only have data from one year. Comparing the gaps in reading and mathematics performance of adults in this year with the gaps in performance at age 15, grade 8, and grade 4, it appears that gaps in adulthood are large, especially for first generation migrants. This is probably because a large proportion of the first generation migrants in the adult sample followed (part of) their education in the country of origin. Gaps lay between 0.5 and 0.6 standard deviation, which is substantial.

All in all, it seems that inequalities by socioeconomic status (by parental education and books at home) are from 2003 onwards stable at best, and probably increasing, irrespective of people's age (or grade they are in). With respect to migration background, from 2000 onwards inequalities are probably declining for individuals at all ages, except for reading performances at grade 4.

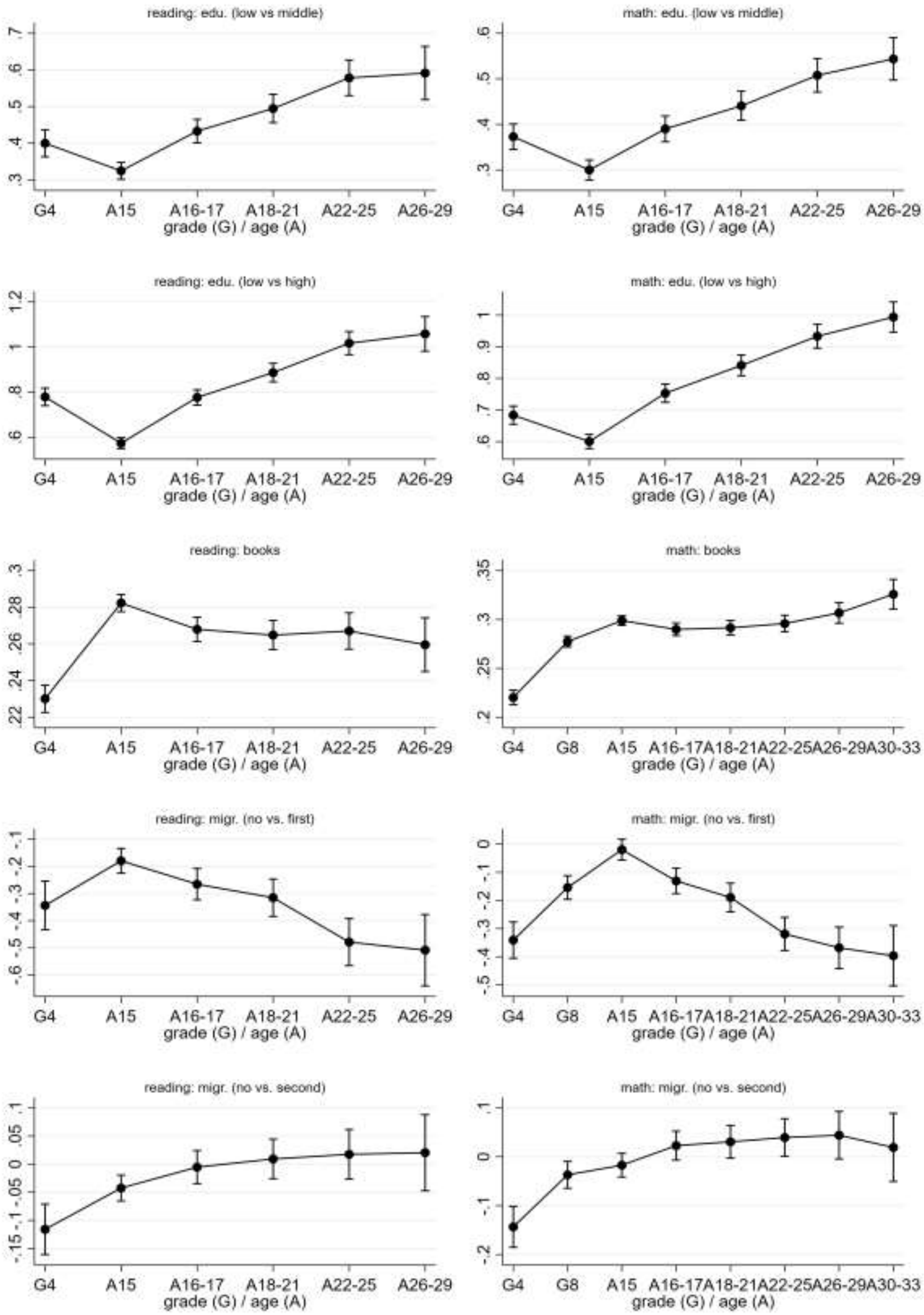
### 4.1.3 PERFORMANCE GAPS WITHIN COHORTS OVER THE LIFE COURSE

Next, we describe how gaps in reading and mathematics performance develop over the life course. Therefore, for each cohort within a country[15] we calculated the gaps in reading and mathematics performance by parental education, number of books at home, and migration background. Afterwards, we predicted those gaps by grade/age and control for country-cohort fixed effects. Thus, we model how on average the gaps change by grade/age over country-cohorts. Predicted values are presented.[16] Figure 4 presents the results.

The first two rows present the changes by parental education, comparing children with low and middle educated parents and children with low and high educated parents, respectively. Both for reading and mathematics we see a similar pattern: performance gaps declined while children attend school (between grade 4 and age 15), but increase steeply afterwards.

[15] See the method section how they are defined.
[16] This is done by the margins command of Stata 15.1. To calculate the margins Stata uses the constant to calculate the y-value when x is 0. However, the fixed effects command presents as a constant the average values of the fixed effects. Instead, for a correct representation of the margins, this should be the average value of the fixed effects when x=0. We adjusted all values of the margins for this bias.

*Figure 4. Educational inequalities over the life course (in standard deviations)*



The third row presents the changes by the number of books at home. In contrast to the changes by parental education, the gaps in reading and mathematics performance increase between grade 4 and age 15. Afterwards gaps in reading performance seem to decline slightly and are more or less table from age 16, while gaps in mathematics performance increase slightly.

The last two rows present the changes by migration background, comparing children without a migration background with first and second generation migrants, respectively. In contrast to the results by parental education and migration background, the higher the figures, the smaller the gaps (i.e. the better migrants perform). First generation migrants seem to catch up with the majority population when they are in school (till age 15), but the gaps tend to widen again afterwards. Second generation migrants also seem to catch up when they are still in school (till the age of 16/17), but in contrast to first generation migrants, their performance compared to the majority population does not decline afterwards.

All in all, it seems that inequalities by socioeconomic and migrations background decline while children are in school (except by the number of books at home). After leaving school, from around the age of 16/18, gaps remain stable or increase.

## 4.2 THE IMPACT OF EDUCATIONAL INSTITUTIONS

### 4.2.1 PERFORMANCE GAPS BY PARENTAL EDUCATION

We examined how the several aspects of educational systems condition the previous performance gaps in reading and mathematics; do they increase or decrease them? Table 5 summarizes the results for parental education for the multilevel models with between and within country effects of the national policies (henceforth between and within models) and the autoregressive models on the pseudo panel that form the second step of our two-step approach (henceforth pseudo panel models).

The between models show how much larger the effect of parental education is in countries that score 1 point higher on the policies (e.g. who spend more on education). The within effects show how much larger the effect of parental education is when the policies change 1 point over time within a country (e.g. when the educational expenditures increases). The pseudo panel models model the transition into a certain age. For example, the grade 8 pseudo panel models show what happens when children move from grade 4 to grade 8. There was no information about scores in grade 4 in mathematics. Models that involved information in grade 4 could therefore not be estimated, i.e. grade 4 within and between models and the (transition from grade 4 to) grade 8 pseudo panel model.

Because the table summarizes so many effects, we do not discuss each separate effect, but instead focus on the broader picture. Note that we evaluate national level policies, not individual level effects. For example, performance gaps may widen when more teachers in a country become high educated, while within a country performance gaps are smallest in the classes with the highest educated teachers.

A number of broader conclusions can be drawn:

- For most policies there are ***mixed results***, both across methods as well as within methods over grades and age.
- Yet, there are a number of policies which are, sometimes surprisingly, systematically associated to an increase in the performance gaps in reading and mathematics: a larger share of ***experienced*** *and* ***highly educated teachers,*** *a* larger number of children that enroll in ***pre-primary education,*** *an* ***earlier start of compulsory schooling****,* and ***earlier tracking.***

Below we zoom in on the separate effects for gaps in reading and gaps in mathematics performance, and explain how we come to the broader conclusions. All outcomes of the models (132 in total), including estimates and standard errors, can be found in the online appendix (www.isotis.org).

**GAPS IN READING PERFORMANCE.** Panel A summarizes the results for gaps in reading performance between children with low and middle educated parents. There are ***many mixed*** and not significant results. Thus, applying one method on one survey (say PIRLS) may result in different conclusions than using the same method on a different survey (say TIMSS) or using the same survey (say PIRLS again) using another method. Yet, a couple of policies stand out.

A first policy that stands out is ***teacher experience***. This has the strongest effect when children are still in school, thus from grade 4 to at least age 15 (see also the method section), and having this age range in mind we find all significant effects (except for the within model on performance gaps at age 15). All these effects are positive, suggesting that raising the experience of teachers in a country results in larger gaps between children with low and middle educated parents. Especially the effect size of the within model at grade 4 is substantial with an estimate of 0.065, meaning that an increase in the experience of teachers with 10 years increases the gaps between children with low and middle educated children with 0.65 standard deviation. Such as gap represents for example the difference between an average student and a student that belongs to the top 25 percent in her country (assuming a normal distribution). Other estimates are smaller, ranging between 0.009 for the pseudo panel model to 0.020 for the between model at age 15.

Another policy that stands out is ***pre-primary school enrollment***. Theoretically this has the strongest effect when children start school. Thus, in our case at grade 4. For grade 4, the estimates are statistically significant in the within model, but not the between model. This can either be because only a change in pre-primary attendance affects these inequalities (as the within model picks up) but a high level of pre-primary attendance as such not (which the between model is more likely to pick up), or because the within models rules out all time invariant variance and confounders. Because of the latter, we consider the result of the within model slightly more reliable. Simplified, it seems that the more children enroll in pre-primary education, the larger the performance gaps in reading between children with low and middle educated parents become. Yet, the effect size of this estimate is small (b=0.009), meaning that a change from no coverage at all to full coverages of pre-primary education increase the performance gaps with less than a tenth of a standard deviation.

Furthermore, we found some support that ***tracking age*** matters. Tracking affects children around grade 8 and age 15 most. The estimates of the age of tracking are not significant in the within model, but they are at age 15 and in adulthood in the between and pseudo panel models. These effects are in line with the majority of the literature and imply that later tracking declines the gaps in reading performance between children with low and middle educated parents. Effects sizes are moderate. For example, the pseudo panels show that in 2007 in countries with an average GDP which tracks at age 11 the disadvantage from students with low educated parents increases between grade 4 and grade 8 with 0.328 standard deviation, but in countries that track at age 16 with 0.310 standard deviation: a difference of less than a tenth of a standard deviation. Note, however, that research has shown that tracking policies more strongly correlate to inequalities by parental occupation than by their education (Van de Werfhorst 2018, 2019).

Vocational specificity only shows positive effects. Yet, most of them occur in the between model which is argued to be least reliably and occurs before this policy is argued to be effective (around grade 8), pointing at a selection effect. For the other policies – educational expenditure, pupil/teacher ratios, teacher salaries, teacher hours, teachers education, compulsory education – (very) few effects are significant or the results are (very) mixed.

Panel B summarizes the results for gaps in reading performance between children with low and high educated parents. While there are again many ***mixed findings, teachers experience, pre-primary school-enrollment,*** and ***tracking*** again seem to be related to larger socio-economic inequalities. Yet, we have to be somewhat careful. For teacher experience two effects have become borderline significant (P<0.10) and there is one additional negative effect in the within model for age 15. It might be that more experienced teachers especially increase inequalities especially at a younger age. In line with this, estimates are largest for the youngest grade, although they are still small. Increasing the experience of teachers with 10 years is associated with a further advantage of children with high educated partners compared to children of low educated parents of around 0.10 (the between model) and 0.16 (the within model) standard deviation. For pre-primary school enrollment the positive effects in the between and pseudo panel lose significance (P-values becomes larger than 0.05); however, there is still a small but significant positive effect (of 0.007) in the within model at grade 4, the model we find the strongest test for this policy. And for the age of tracking we not only find positive effects, but also a borderline negative significant, at the within model at age 15.

Besides these effects, there are a substantial number of positive effects for ***teachers education*** (6 out of 8). These imply that when the percentage of teachers with a master degree increases in a country, the gaps between students with parents with low and high educated parents increases. These effects are considerable, and they increase over the life course. For example, 50 percent point more highly educated teachers, which is not unlikely given the huge variation across countries, results in a further advantage of children with high educated parents with approximately 0.30 (between), 0.25 (within) or 0.20 (pseudo panel) standard deviation. For example, this roughly represents the difference between an average students and one that belong to the top 40 percent in her country.

Also the ***start age of compulsory education*** seems to matter. This policy should affect the youngest children most. In our case, those at grade 4. For grade 4 we find a negative effect, both in the between and within model. These effects suggest that a later start age declines the gaps between children with low and high educated parents, while an earlier start age might increase these gaps.

The other policies show few or (very) mixed results.

**GAPS BY MATHEMATICS PERFORMANCE.** Panel C summarizes the results for gaps in mathematic performance between children with low and middle educated parents. Results are in line with those for reading. Besides ***many mixed and not significant effects***, for ***teacher experience*** and ***pre-primary school enrollment*** we again find a substantial number of positive effects (especially at grade 4 for pre-primary education), indicating that these policies increase the gaps between children with low and middle educated parents slightly. In addition, in line with the earlier results, we find in the between and pseudo panel models – but not in the within models – a negative effect for the ***age of tracking***, implying that gaps are slightly smaller when countries track later. The other policies show few or (very) mixed results.

Panel D summarizes the results for gaps in mathematic performance between children with low and high educated parents. As for gaps in reading (Panel B), ***teachers experience, pre-primary education*** and ***early tracking*** still seem to increase educational inequalities. Yet, more uncertainty around these findings arise (especially around the effect of tracking): for teacher experience there are also two negative effects; for pre-primary school enrollment we again find positive estimates at the earliest observation (that is grade 8) for the between and within model, but not in the pseudo panel model; and for the age of tracking we find one additional negative estimate in the within model.

*Table 5. Overview outcomes hybrid and pseudo panel models on gaps in student achievement sores by parental education.*

Panel A: Gaps in reading performance: low versus middle educated parents

| | Between | | | Within | | | Pseudo | |
|---|---|---|---|---|---|---|---|---|
| | G4 | A15 | Adult | G4 | A15 | Adult | A15 | Adult |
| Educational expenditure | p | p | | p | n | p | n | |
| Pupil/teacher ratio | n | n | | | p | | n | |
| Teacher salaries | | n | n | | n | p | | n~ |
| Teaching hours | | p | | n | p~ | | | |
| Teacher educ. (% master d.) | p | n | p | n | | | | p~ |
| Teacher experience | p | p | p | p | | | p | |
| Pre-primary school enrol. | | p | | p | | | p | |
| Compulsory edu.: start age | | | n | n | | p | | |
| Compulsory edu.: end age | p | p | | n | p | n~ | | |
| Tracking age | | n | n | NT | | | n | n~ |
| Vocational specificity | p | p | p | p | | | | p |

Panel B: Gaps in reading performance: low versus high educated parents.

| | Between | | | Within | | | Pseudo | |
|---|---|---|---|---|---|---|---|---|
| | G4 | A15 | Adult | G4 | A15 | Adult | A15 | Adult |
| Educational expenditure | n | n | | | | | n | p |
| Pupil/teacher ratio | p | p | p | n | | | | |
| Teacher salaries | p | p | n~ | n | | p | | n |
| Teaching hours | n | p | | | p | | | |
| Teacher educ. (% master d.) | p | p | p | n | | p | p~ | p |
| Teacher experience | p | p | p | p~ | n | | | p~ |
| Pre-primary school enrol. | | | | p | p | | | |
| Compulsory edu.: start age | n | n | | n | | p | | |
| Compulsory edu.: end age | p | p | p | | p | | | |
| Tracking age | n | n | n | NT | p~ | | n | n |
| Vocational specificity | p | | p | n | n | p | | p |

Panel C: Gaps in mathematics performance: low versus middle educated parents

| | Between | | | | Within | | | | Pseudo | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | G4 | G8 | A15 | Adult | G4 | G8 | A15 | Adult | G8 | A15 | Adult |
| Educational expenditure | NT | p | p | p | NT | n | n | p | NT | n | |
| Pupil/teacher ratio | NT | n | n | | NT | p | p | | NT | | |
| Teacher salaries | NT | n | n | n | NT | n | n | p | NT | | n~ |
| Teaching hours | NT | p | p | n~ | NT | p | | | NT | | |
| Teacher educ. (% master d.) | NT | p | n | p | NT | p | | p~ | NT | | n~ |
| Teacher experience | NT | p | p | p | NT | p | n~ | | NT | | |
| Pre-primary school enrol. | NT | p | p | | NT | p | | | NT | | |
| Compulsory edu.: start age | NT | | | n | NT | | | | NT | | n |
| Compulsory edu.: end age | NT | n | p | | NT | n | p | p | NT | | |
| Tracking age | NT | n~ | n | n | NT | NT | | | NT | n~ | n |
| Vocational specificity | NT | p | p | p | NT | n | | | NT | | p |

Panel D: Gaps in mathematics performance: low versus high educated parents.

| | Between | | | | Within | | | | Pseudo | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | G4 | G8 | A15 | Adult | G4 | G8 | A15 | Adult | G8 | A15 | Adult |
| Educational expenditure | NT | | n | | NT | n | | | NT | | |
| Pupil/teacher ratio | NT | p | p | p | NT | n | n | | NT | p | |
| Teacher salaries | NT | n | p | n | NT | | n | p | NT | | n |
| Teaching hours | NT | p | p | | NT | p~ | p | | NT | | n~ |
| Teacher educ. (% master d.) | NT | p | p | p | NT | p | p~ | p~ | NT | | |
| Teacher experience | NT | p | | p~ | NT | p | n | | NT | n | p |
| Pre-primary school enrol. | NT | p | n | | NT | p | p | | NT | | |
| Compulsory edu.: start age | NT | n | n | n | NT | n | | p | NT | | n~ |
| Compulsory edu.: end age | NT | n~ | p | | NT | | p | | NT | n | |
| Tracking age | NT | n | n | n | NT | NT | p | | NT | | n |
| Vocational specificity | NT | p | n | p | NT | n | | | NT | | p |

Abbreviations: p = positive effect; n = negative effect; NT = not testable; blank = not significant; (P<.05); ~ borderline significant (P<.10); grey= based on less than 30 cases; G4 = grade 4; G8=grade 8; A15= age 15.

Furthermore, similarly as for reading (in Panel B) it seems that teachers education and the start age of compulsory education seems to matter. With the exception of the pseudo panel models, we find positive estimates for ***teachers education***. Effect sizes are slightly lower than for reading, however. A 50 percent point increase in highly educated teachers is associated with an increase in the disadvantage of children with low educated parents with approximately between 0.30 (between) and 0.20 (within) standard deviation. For example, these gaps roughly correspond to the difference between and average student and a student who belongs to the top 40 percent in her country. Furthermore, we again find negative effects for ***the start age of compulsory education*** at grade 4, implying that an earlier start age increase the gaps between children with low and high educated parents.

Besides these effects, there is some indication that ***teaching hours*** matter. Except for the pseudo panel model, at the ages that children certainly attend school (grade 4, grade 8 and age 15), and thus form the best test, we find a positive effect. The effects sizes of the between and within models are small, however (laying between approximately 0.50 for the between models and approximately 0.30 for the within models). These estimates mean that we need at least 1000 additional teaching hours to decline the disadvantage of children with low educated parents with half or a third of a standard deviation. The pseudo panel models are very demanding as they control for gaps in mathematics performance in grade 8, and might not have picked up these small effects. For the other policies we again find few or mixed results.

### 4.2.2 PERFORMANCE GAPS BY THE NUMBER OF BOOKS AT HOME

Table 6 summarizes the results for the number of books at home. The number of books is standardized, and effect sizes are calculated between children who differ 2 standard deviation. For example, children 1 standard deviation below and above the average. These are children who roughly belong to the bottom and top 16 percent (assuming a normal distribution) in their country, and can be thought as to represent children whose parents have “few” and “many” books at home.

The broader conclusions are largely in line with those for parental education:

- For most policies there are ***mixed results***.
- Yet, there are a number of policies which are most likely to increase the performance gaps in reading and mathematics: the ***experience of teachers***, enrollment in ***pre-primary education***, the ***start age of compulsory education***, and ***later tracking***.
- Furthermore, the number of high ***educated teacher*** is associated with larger gaps in mathematics performance.

Below we zoom in on the separate effects for gaps in reading and gaps in mathematics performance, and explain how we come to the broader conclusions. All outcomes of the models, including estimates and standard errors, can be found in the online appendix (www.isotis.org).

**GAPS IN READING PERFORMANCE.** Panel A summarizes the outcomes for gaps in reading performance. There are many mixed and not significant results. Yet, a couple of policies stand out.

Two of the policies that stand out are ***pre-primary education*** and the ***start age of compulsory schooling***. The best test for the effectiveness of these policies are the grade 4 between and within model (see also the method section). These models show a positive effect for pre-primary education, meaning that a larger enrollment increases the gaps in reading performance by the number of books at home. For the start age of compulsory education we find with the same models a negative effect. This means that when the start age of compulsory education increases, the gaps in reading performance decline, and thus also that when the start age declines the gaps increase. Effect size are moderate though. For the coverage of pre-primary education the effect size is around 0.001, suggesting that a 50 percent point increase in the coverage of pre-primary education increases the performance gap between children that differ two standard deviation in the number of books at home (e.g. a child one standard deviation below and a child one standard deviation above average) with no more than 0.10 of a standard deviation. For the start of compulsory education, estimates are around -0.01 (between) and -0.05 (within), suggesting that for instance changing the compulsory age from 7 to 5 years increase the performance gap between children that differ two standard deviation in the number of books at home with 0.2 standard deviation (within) or less (between). This represents for example the difference between an average student and a student who almost belongs to the top 40 percent of her country (assuming a normal distribution).

Furthermore, for ***teacher experience*** we find a substantial number of positive effects (6 out of 8 effects). Effects are again moderate, being around 0.016 or smaller, meaning that the experience of teachers has to increase with 6 years to increase the performance gaps between children that differ two standard deviation in the number of books at home with 0.2 standard deviation, making the difference between an average student and one that almost belongs to the top 40 percent. It is likely that the within model at age 15 and the pseudo model for the transition into adulthood, which showed not significant effects, did not pick up these effects as they are demanding for the data.

Moreover, around the age that ***tracking*** occurs (i.e. between age 12 and 16), three out of four of these models show a significant negative effect. These indicate that the later countries track, the smaller the gaps between children with few and many books at home will be. The within model for adulthood shows a substantial effect size of -0.05, meaning that when the tracking age is for example changed from age 11 to age 16, performance gaps between children 1 standard deviation below and above the average decline with approximately half a standard deviation. The estimates for the other models, however, are smaller, laying around -0.01.

Teacher salaries and vocational specificity also seem to have consistent results. However, these effects already occur after and before these policies should be most effective, pointing at selection effects. For the other policies we find few or (very) mixed results.

**GAPS IN MATHEMATICS PERFORMANCE.** Panel B describes the results for gaps in mathematics performance. The effect of ***pre-primary school enrollment*** and the ***start age of compulsory education*** are in line with those for reading. The within and between models at grade 4 show again a positive effect for pre-primary school enrollment and a negative effect for the start age of compulsory education, suggesting that the disadvantage for children with few books at home increases when more children attend pre-primary education. Effect sizes, however, are again small, with estimates for the coverage of pre-primary education being around 0.001 and of the start age of compulsory education between -0.01 (between) and -0.03 (within), meaning that coverage has to increase with 50 percent point to result in a reduction in the performance gaps between children that differ two standard deviation in the number of books at home with a tenth of a standard deviation and 2 years of additional compulsory education with a reduction of the gap with approximately a tenth (within) of a standard deviation or less (between).

Furthermore, for ***teacher experience*** we again find a substantial number of positive but small significant effects (7 out of 11, with estimates being around 0.01 or lower). But, in contrast to these findings, we also find one significant negative between effect for age 15, while the effects in the pseudo panel models are not statistically significant. Nevertheless, if there is an effect, it remains most likely that more experienced teachers in a country increase the performance gaps.

Also the effects of ***tracking*** become less pronounced: in the within and between models at age 15 (which form the strongest tests) we still find a negative but small effect, implying that later tracking can decline inequalities; however, the effect in the pseudo panel model is only borderline significant, no significant effect for grade 8 is found in the pseudo panel models, and the between model for age 15 even shows a positive effect. Thus, we have to be somewhat careful with concluding that later tracking can close the gaps between children with few and may books at home.

We also find a substantial number of positive effects for ***teachers education***, especially in the within models which are argued to be most reliable. With effect sizes laying between 0.38 and 0.75 for the within models effects are moderate: increasing the experience of teachers with 10 years increases the performance gaps with between 0.15 and 0.07 standard deviation.

Furthermore, the vocational specificity of an educational system seems to matter. This policy affects children in grade 8 and above most. Three out of six effects in the between and pseudo panel models show a positive effect (and two out of three in the between model). Simplified, a larger vocational system in a country widens the performance gaps between children with few and many books at home. However, with effect sizes of 0.002 or lower, these effects are (very) small: a change from a complete general to complete vocational system does widen gaps between children that differ two standard deviation in the number of book at home with at best two tenth of a standard deviation and more likely less. The results for the other policies are again mixed at best.

*Table 6. Overview outcomes hybrid and pseudo panel models on gaps in student achievement sores by the number of books at home.*

Panel A: Gaps in reading performance

| | **Between** | | | **Within** | | | **Pseudo** | |
|---|---|---|---|---|---|---|---|---|
| | G4 | A15 | Adult | G4 | A15 | Adult | A15 | Adult |
| Educational expenditure | n | p | | | | | | |
| Pupil/teacher ratio | p | | p | n | n~ | | | |
| Teacher salaries | p | | p | | p | p | | |
| Teaching hours | n | n | | n~ | p | | p | |
| Teacher educ. (% master d.) | p | p | p | | | | | |
| Teacher experience | p | p | p | p | n | p | p | |
| Pre-primary school enrol. | p | p | | p | | n~ | p | |
| Compulsory edu.: start age | n | n | | n | | p | n | |
| Compulsory edu.: end age | p | p | p | n~ | | | | |
| Tracking age | n | n | n | NT | | n | n | |
| Vocational specificity | p | p | p | n | p~ | p | | |

Panel B: Gaps in mathematics performance

| | **Between** | | | | **Within** | | | | **Pseudo** | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | G4 | G8 | A15 | Adult | G4 | G8 | A15 | Adult | G8 | A15 | Adult |
| Educational expenditure | n | p | n | | p | n | p | | | | |
| Pupil/teacher ratio | p | p | p | p | n | n | | p | n | | p |
| Teacher salaries | n | n | n | p~ | n | n | p | p | n | n~ | |
| Teaching hours | n | n | p | | n | p | n | | | | |
| Teacher educ. (% master d.) | p | p | n | p | p | p | p | | | | |
| Teacher experience | p | p | n | p | p | p | p | p | | | |
| Pre-primary school enrol. | p | p | n | | p | p | p | n~ | | p | |
| Compulsory edu.: start age | n | n | p | n | n | n | n | p | | n~ | n~ |
| Compulsory edu.: end age | p | | n | | n | | p | n~ | n | p | |
| Tracking age | n | n | p | n | p | | n | | | n~ | |
| Vocational specificity | p | p | n | p | p | | p | p | | p | |

Abbreviations: p = positive effect; n = negative effect; NT = not testable; blank = not significant; (P<.05); ~ borderline significant (P<.10); grey= based on less than 30 cases; G4 = grade 4; G8=grade 8; A15= age 15.

### 4.2.3 PERFORMANCE GAPS BY MIGRATION BACKGROUND

Table 7 summarizes the results by migration background. Positive effects mean that the scores of first and second generation migrants increase more than those of the native majority (i.e. third generation and higher) when a policy is implemented or extended.

A number of broader conclusions can be drawn:

- Where we already found many mixed results by parental education and the number of books at home, we find even more ***mixed results*** by migration background. Thus, we have to be (very) careful with our conclusions, or it must be that there are no unambiguous effects.
- ***Pre-primary education*** stand out most. It seems to reduce performance gaps in reading and mathematics between first generation migrants and the native majority. Furthermore, it seems to reduce the gaps in mathematics performance between second generation migrants and the native majority.
- There are also a couple of other policies that stand out, but then only for reading or mathematics, or for first or second generation migrants. These include: ***smaller classes*** (reducing gaps in reading between first generation migrants and the native majority), ***teachers education*** (declining the gaps in mathematics between first generation migrants and the native majority), ***vocational specificity*** (increasing the gaps in mathematics between first generation migrants and the native majority), and ***educational expenditure*** (increasing the gaps in mathematics between second generation migrants and the native majority).

Below we zoom in on the separate effects for gaps in reading and gaps in mathematics performance. All outcomes of the models, including estimates and standard errors, can be found in the online appendix (www.isotis.org).

**GAPS IN READING PERFORMANCE.** Panel A summarizes the results for gaps in reading performance between first generation migrants and the native majority. As emphasized above, most policies show mixed effects.

***Pupil/teacher ratio*** might form an exception. We find positive effects in the between models, but negative effects in the within models. The between effects may be biased by unobserved between country heterogeneity, something that does not affect the within models, and therefore can be considered the better test (see also the method section). The within models show that when the pupil/teacher ratio's increase in a country the gaps in reading performance between first generation migrants and the native majority widen. In other words, when pupil/teacher ratios decline within a country (and classes probably become smaller), the performance of first generation migrants increases more than those of the native majority. Especially in grade 4 those effects are substantial. A reduction of approximate 6 students per class is already sufficient to let the disadvantage of first generation migrants decline with half a standard deviation. For example, this represents the gap between an average student and one that belongs to the top 30 percent in her country (assuming a normal distribution). At age 15 and adulthood effect sizes are smaller, but still considerable: a reduction of 6 students declines the gaps with respectively approximately one and two tenth of a standard deviation.

Furthermore, often a lot is expected from ***pre-primary education***, and the best test available (grade 4 within) indeed shows that the gaps decline when more children enroll within pre-primary education. Thus, it seems that pre-primary education increases inequality by socio-economic background, but decreases by migration background. The effect is moderate. A ten percent point increase in the coverage of pre-primary education declines the disadvantage of first generation migrants with 0.16 standard deviation (almost representing the gap between an average student and one belonging to the top 40 percent). For the start age of compulsory education, however, we find no significant effect using information from the within model using data from grade 4.

Panel B summarizes the results for gaps in reading performance between second generation migrants and the native majority. ***No effect clearly stands out***. Teacher salaries and teachers experience show the most robust effects, with four out of eight effects being negative, suggesting that gaps in reading performance between second generation migrants and the native majority increase when the salaries and experience of teachers increase. However, both for teachers experience and salaries three out of four significant effects occur in the between model which we consider least reliable because these models do not look at actual changes over time and are therefore more likely to be affected by unobserved (time-invariant) cofounders. Similarly, three out the four positive effects for teaching hours occur in the between model.

Furthermore, and in contrast to panel A (and as we discuss below from panel C and D), we find a negative effect for grade 4 pre-primary education in the within model, contesting the idea that migrants can profit most from pre-primary education. Yet the effect size is small. A 17 percent increase in the coverage of pre-primary education is needed to increase the disadvantage of second generation migrants with (only) a tenth of a standard deviation.

**GAPS IN MATHEMATICS PERFORMANCE.** Panel C summarizes the results for gaps in mathematics performance between first generation migrants and the native majority. Almost all policies show ***mixed results***.

One exception is ***teachers education*** for which all within country models show a positive effect. Together with the pseudo panel models we also consider them most reliable (see for a further explanation the method section). They may deviate from the between models that show negative effects because they are less likely to be affected by unobserved heterogeneity and look at changes over time. These within models indicate that the higher teachers become educated in a country, the more the gaps between natives and first generation migrants in mathematics decline. Effect sizes are particularly large at grade 4 in which only a seven percent point increase in the number of teachers is needed to decline the disadvantage of first generation migrants with a tenth of a standard deviation. In grade 8 and age 15, respectively, a 30 and 33 percent point is needed to reduce the disadvantage to the same extend.

Furthermore, for ***vocational specificity*** we find from grade 8 onwards, when the policy is expected to have the largest impact, negative effects in the between and pseudo panel models. In addition, we find a negative effect in the within model at grade 8. These results are thus quite consistent. They suggests that the more students enroll in upper vocational education, the worse first generation migrants perform compared to natives. The significant between-estimates are around 0.010. Considering that the coverage between countries range between almost zero and 80 percent this effect size is substantial. For example, an 80 percent point greater coverage of upper secondary vocational education increases the disadvantage of first generation migrants in mathematics performance with approximately 0.8 standard deviation. This represents the gap between an average student and one belonging to the top 20 percent. The significant within-estimates have approximately the same magnitude, while the estimates of the pseudo panel are slightly smaller.

*Table 7. Overview outcomes hybrid and pseudo panel models on gaps in student achievement sores by migration background*

Panel A: Gaps in reading performance: natives vs first generation migrants

| | Between | | | Within | | | Pseudo | |
|---|---|---|---|---|---|---|---|---|
| | G4 | A15 | Adult | G4 | A15 | Adult | A15 | Adult |
| Educational expenditure | n | n | n | | p | | | n |
| Pupil/teacher ratio | p | p | p | n | n | n | | |
| Teacher salaries | | n~ | | | p | n | n | |
| Teaching hours | p | p | | n | p | | | |
| Teacher educ. (% master d.) | n | | | p | p | | | |
| Teacher experience | n | n | | n | | | | |
| Pre-primary school enrol. | n | n | n | p | | p | n | |
| Compulsory edu.: start age | | n | n | | p | n | n | n~ |
| Compulsory edu.: end age | n | n | n | n | | p | | |
| Tracking age | p | p | | | p | | | |
| Vocational specificity | n | n | n | | | | n | |

Panel B: Gaps in reading performance: natives vs second generation migrants

| | Between | | | Within | | | Pseudo | |
|---|---|---|---|---|---|---|---|---|
| | G4 | A15 | Adult | G4 | A15 | Adult | A15 | Adult |
| Educational expenditure | n | n | n | | | | | |
| Pupil/teacher ratio | p | p | p | | n | | | |
| Teacher salaries | n | n | n | | | | n | |
| Teaching hours | p | p | p | n | | | p | |
| Teacher educ. (% master d.) | n | | | | p | | | |
| Teacher experience | n | n | n | n~ | | | | |
| Pre-primary school enrol. | n | n | n | n | p | | | |
| Compulsory edu.: start age | | n | n | | n | | n~ | |
| Compulsory edu.: end age | n | n | | n~ | | p~ | | |
| Tracking age | p | p | p~ | NT | | | | |
| Vocational specificity | n | n | n | | | | | |

Panel C: Gaps in mathematics performance: natives vs first generation migrants

| | Between | | | | Within | | | | Pseudo | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | G4 | G8 | A15 | Adult | G4 | G8 | A15 | Adult | G8 | A15 | Adult |
| Educational expenditure | p | n | n | n | | n | p | n~ | NT | | n |
| Pupil/teacher ratio | p | n | p | p | | n | n | p~ | NT | | |
| Teacher salaries | n | | n | n | n | n | p | p~ | NT | | |
| Teaching hours | n | p | p | p | n | p | p | | NT | | |
| Teacher educ. (% master d.) | n | n | | | p | p | p | p | NT | | |
| Teacher experience | n | n | n | n | p | p | | n | NT | | |
| Pre-primary school enrol. | | n | n | n | p | p | n | p | NT | n | |
| Compulsory edu.: start age | n | n | n | n | p | p | p | | NT | | n~ |
| Compulsory edu.: end age | n | n | n | n | n | p | | p | NT | | |
| Tracking age | p | p | p~ | p | p | NT | p | | NT | | |
| Vocational specificity | p | n | n | n | p | n | | | NT | n | n |

Panel D: Gaps in mathematics performance: natives vs second generation migrants

| | Between | | | | Within | | | | Pseudo | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | G4 | G8 | A15 | Adult | G4 | G8 | A15 | Adult | G8 | A15 | Adult |
| Educational expenditure | p | n | n | n | | | n~ | | NT | n | n |
| Pupil/teacher ratio | p | p | p | p | n | | n~ | | NT | | |
| Teacher salaries | | | n | n | n | n | p~ | n | NT | | |
| Teaching hours | p | p | p | p | n | p | | | NT | | |
| Teacher educ. (% master d.) | n | n | | | p | | p | | NT | | |
| Teacher experience | n | n | n | n | p | n | n | | NT | | |
| Pre-primary school enrol. | n | n | n | n | p | p | p | p~ | NT | n | |
| Compulsory edu.: start age | n | n | n | | p | p | | | NT | | n~ |
| Compulsory edu.: end age | n | n | n | | n | | p | p | NT | | |
| Tracking age | p | p | p | p~ | p | NT | | | NT | | |
| Vocational specificity | p | n | n | n | p | | | | NT | n | n |

Notes: between and within effects for grade 4 mathematics are not controlled for parental education.
Abbreviations: p = positive effect; n = negative effect; NT = not testable; blank = not significant; (P<.05); ~ borderline significant (P<.10); grey= based on less than 30 cases; G4 = grade 4; G8=grade 8; A15= age 15.

Moreover, we again find a positive effect for ***pre-primary school*** enrollment for the grade 4 within model, implying that performance gaps in mathematics between migrants and natives decline when more children attend pre-primary education. However, the effect size is only a quarter of those for reading: a 25 percent point increase in the coverage of pre-primary education is needed to decline the disadvantage of the first generation with a tenth of a standard deviation. This support the idea that pre-primary education is of greater value for the language proficiency of migrants than the mathematics performance.

Panel D, finally, summarizes the results for gaps in mathematics performance between second generation migrants and the native majority. Again there are many **mixed effects**.

Six out of the ten effects for ***educational expenditure*** are negative and (borderline) significant, especially at grade 8, age 15 and adulthood in which we find (borderline) significant effects in the between, within and pseudo panel models. Simplified, an increase in educational expenditures results in larger performance gaps in mathematics between the native majority and second generation migrants. The effects sizes are quite strong. The between models show that countries that spend one percentage point more on education the disadvantage of second generation migrants is between 0.13 (grade 8) and 0.24 (age 15) larger, while the pseudo panel models show that every point increase in education expenditure is associated with a further increase in the disadvantage of second generation migrants by 0.18 standard deviation between grade 8 and age 15 increase and with and another 0.27 standard deviation between age 15 and adulthood. When countries increase spending with one percent point the disadvantage of second generation migrants in mathematics at age 15 increase with a modest 0.03 standard deviation. Multiple explanations for these differences might be given, of which we find two most plausible. Or the between and pseudo panel models might be affected by unobserved heterogeneity, or it might be that it takes a while before the native majority figures out how to use the additional spending in their favor.

Furthermore, we again find a positive effect for grade 4 ***pre-primary education***, once more confirming that migrants can benefit most from pre-primary education. Yet, as was the case with first generation migrants, effect sizes are again considerably smaller than for performance in reading, needing even approximately a 33 percent increase in the coverage of pre-primary education to decline the performance gaps with only a tenth of a standard deviation. In addition, using the same model, we find a positive effect for the start age of compulsory education, suggesting that an earlier start of compulsory education declines the gaps between natives and second generation migrants.

## 5. DISCUSSION

The mere size of the dataset, the number of policies taken into account, and the use of several advanced models simultaneously which are applied the same for each policy are the absolute analytic strengths of this report. Yet, this rigorousness, standardization and the size of the dataset also came with some costs, and these have to be discussed before we turn to our conclusion and policy recommendations.

First, as a measure of socioeconomic status we relied (most often) on self-report measures from children about their parents education and the number of books at home. As discussed in greater length in the method section, there is much doubt about the usage of the number of books at home in comparative research. Reports on parental education are probably better than those about the number of books at home, and these report are considered of good enough quality for comparative research (Lien 2001; Jerrim and Micklewrigh 2014). However, in contrast to the number of books at home, parental education is not asked in every survey-year. In addition, because education is measured in different ways in different survey-year combinations, we had to compress the answer categories as good as possible, causing that middle and high educated categories sometimes consists of different sizes and other educational levels across surveys and countries. Consequently, there might be much noise in this measure, resulting in underestimation of the estimates, and fewer robust and more not significant effects. Related to this, we have only very crude information about the migration background of respondents, and for instance do not know their country of origin. As a result, we mix together different groups of immigrants, who – although we control for instance for the education of their parents – are very different. This can partly explain why we found so many mixed results for the policies that should effect the gaps between natives and migrants.

Second, for several independent variables we had only crude estimates or proxies. For example, teacher quality was assessed with the experience and education of teachers. Although these are indications of the quality of a teacher, other aspects, such as pedagogical skills, might be more important. Also for educational expenditure, pupil/teacher ratio, and the salaries of teachers we matched information about primary and secondary education to the respective ages, but these groups cover several ages, resulting in imprecise measures. Similarly, we use information about grade 8 and age 15 for the estimates of teaching hours, teachers education, and teachers experience when students were 15 years old and an adult. These imprecise measures might have especially affected the within and pseudo panel models, which rely on a comparison across time, as it increases the risk that we associated a bias or noise at one point in time with a bias or noise at another point in time.

Third, we had to make many choices about which methods we use and how we model them. Some of these choices can be debated. We did not use weights for the multilevel models with the between and within-estimates. Also we only included between- and within-estimates of the national policies but not for the effects of parental education, number of books at home, and migration background. As a result, the resulting cross-level interactions are no “true” between and within effects. In the between and within models we also only nested respondents in countries, but did not nest them in time and country-years because these models did not always converge (although we control for year). Related, we did not include random effects for parental education, number of books at home, and migration background in the hybrid models because these models did not always converge – due to small variances. This is generally accepted in the fixed effect literature because it is not needed to perfectly mimic fixed effects (Allison 2009; Giesselman and Schmidt-Catran 2018), but is often recommended in the multilevel literature (Heisig and Schaeffer 2018).

Furthermore, we decided to only control for GDP in the models and not for other policies that might be implemented together with the policies under consideration because of power issues – otherwise we had to delete too many cases. However, due to this, the estimates are still biased by omitted variables, also the within-estimates. Moreover, we think that the autoregressive pseudo panel models are most intuitive for our purpose. However, fixed effect models are more often applied for this type of data – although we found them too demanding for our data. Finally, in the autoregressive models we control for lagged effects of the dependent variables, based on estimates from earlier grades/ages. However, the period of the lag differed between the models. For example, for reading performance the models that predict performance gaps at age 15 are controlled for the performance gap at grade 4, while for mathematics they are controlled for the gaps at grade 8. Besides that it is the question which lag is best, it might explain differences between the models that use this method.

## 6. SUMMARY OF THE RESULTS

Reducing inequalities in educational achievement, like those by socioeconomic status (SES) and migration background, is not only a moral obligation to improve the lives of these children, but also shows that much potential is unused in modern societies. In this report we reviewed the literature on how successful a broad range of educational policies are in reducing these inequalities. In addition, we tested how successful these policies are in tackling inequalities in high income countries using all available large scale student assessment data (i.e. the PIRLS, TIMMS, PISA and PIAAC) and a variety of methods in which we look both at between country differences, within country changes over time, and in which we follow cohorts over their life course.

### 6.1 EDUCATIONAL INEQUALITIES BETWEEN COUNTRIES, OVER TIME AND ACROSS THE LIFE COURSE

The results show that there are particularly large inequalities by socioeconomic background. However, the magnitude of the differences varied substantially. Comparing the performance of children with low and high educated parents in mathematics and reading, performance gaps range from an exceptional low of approximately 0.3 standard deviation in Macao and Hong Kong, around 0.4 standard deviation in the Scandinavian countries, around 0.7 standard deviation in Germany, France the UK, and US, to around 1 standard deviation in Poland, Slovakia, Hungary, and Chile. The differences by the number of books at home were slightly lower. Gaps between children with few and many books at home are in Europe often larger than in other regions of the world such as the Gulf region, especially compared to the gaps by parental education. As the number of books at home is often regarded as a proxy for the cultural capital at home, it might be that parents cultural capital plays a more pronounced role in Europe than in other parts of the world.

There are also substantial differences between migrants (or their descendants) and the native majority (i.e. third generation migrants or higher). The gaps between first generation migrants and the native majority are particularly large, while the performance of second generation migrants almost always comes closer to those of the native majority. First generation migrants have in some European societies an average disadvantage of around 0.5 standard deviation in mathematics and literacy scores. Differences are especially large in North-Western European countries such as Belgium, Germany, and Austria. But in many countries, migrants also outperform the native population. This especially occurs in liberal and selective immigration countries such as the UK, Australia, Panama and the Gulf region. These countries and regions are known to attract immigrants with a greater skillset and who are determined to succeed in the educational system and labor market (Devitt 2011). While a part of the differences by migration background is explained by taking socioeconomic background into account, the overall effects of migration background provided very similar country rankings as without controls for parental education and the number of books in the household.

Across the life course inequalities by socioeconomic and migration background are already relatively large at grade 4, are stable or may even decline while children are in primary and secondary school, but increase again thereafter or remain stable at best. Both types of inequalities may tend to increase over the life course because children and young adults with a migration background and low socioeconomic background grow up, live and work in a cognitively less stimulating and beneficial environments (Griga and Hadjar 2014; Skopek and Passaretta 2018). The last 10 years are the inequalities by socioeconomic background more or less stable. First and second generation migrations, however, seem to slowly catch up with the native majority over time, especially at grade 4 and grade 8.

## 6.3 TACKLING EDUCATIONAL INEQUALITIES

In the literature there is much debate which policies are effective in tackling both types of inequalities. So far almost all policies have been questioned by their effectiveness. This is also reflected by our results. There is no national institution for which we found unambiguous positive or negative effects. Also not when we looked at inequalities by socioeconomic or migration background solely. This can explain why there are many mixed findings reported in the literature. It also shows the importance of using several methods and applying the same methods to different outcomes and age-groups. Such a broad approach helps to assess the robustness of educational policies. Most likely is that the effectiveness of educational policies is affected by additional factors, such as the quality of them or how easily they can be used by disadvantaged groups. Also the drawbacks from our data and methods may play a role (see the discussion section). Thus, we have to be very careful by pointing at one policy that should always work. Nonetheless, several policies appear to be more promising in reducing educational inequalities than others.

### 6.3.1 TACKLING INEQUALITIES BY PARENTAL BACKGROUND

To reduce socioeconomic inequalities a popular cry is to increase the input in the educational system, most directly by additional expenditure, and indirectly by reducing class sizes, increasing the salaries of teachers and their quality, and to teach more hours. In the literature especially the effectiveness of additional expenditure has been questioned (Hanushek and Wößmann 2017). Our results confirm that educational expenditure says little about the size of socioeconomic inequalities. Although we find slightly more support for the fact that educational expenditures decline socioeconomic differences, we also find many not statistically significant effects and even several effects pointing at the opposite conclusion. Also class size seemed to matter little, despite many popular cries for smaller classes and a substantial body of scientific literature that supports this idea (Angrist and Lavy 1999; Krueger 1999; Blatchford et al. 2003; Maasoumi et al. 2005; Fredriksson et al. 2013; Bosworth 2014). Several models even show that the performance of children with high educated parents increased most in countries with smaller class sizes or in which the class sizes declined. We sometimes saw the same for more teaching hours. This warns us that advantaged children might actually profit most from these policies. Although disadvantaged children might need the additional time and attention resulting from small classes most, advantaged children might be best able to get this attention and might find the additional teaching hours least stressful (Cattaneo et al. 2017). In line with this, we found substantial support that increasing the experience and education of teachers increases the performance of advantaged children with high educated parents and many books at home most, despite that the literature shows that disadvantaged children might learn most from these teachers (Aaronson et al. 2007; Desimone and Long 2010; Montt 2011). Raising the experience and educational level of teachers might increase educational inequalities because these teachers often work in schools in which advantaged students are overrepresented (Borman and Kimball 2005; Clotfelter et al. 2010; Mansfield, 2015). In addition, high educated teachers often have a similar cultural repertoire of advantaged children causing that they often value their skills more, expect more from them, and easier explain something to them (e.g. Bourdieu and Passeron 1977). Thus, when policies are implemented that increase the level of education and experience of teachers, it would help to reduce educational inequalities when it is ensured that experienced and high educated teachers work at schools that disadvantaged children attend and that these teachers adjust their classes in line with the needs of these children. Interesting is that higher salaries are more often – but not always – found to reduce socioeconomic inequalities, especially in the within and pseudo panel models that provide the best indication of a causal relationship. One reason for this contrast to the experience and educational level of teachers is that higher salaries attract a broader range of teachers – with a broader set of teaching practices and cultural affinities –, and helps the motivation of and respect for teachers (Dolton and Marcenaro-Gutierrez 2011; Chetty et al. 2014).

Inequalities in educational systems are not only determined by the sheer input, but also by structural factors, such as the coverage of pre-primary education, the compulsory age, the tracking age, and the vocational specificity of educational systems. These structures are historically grown. Although for citizens, teachers and policy makers it is often less popular to change these policies because this will require drastic changes in the educational system, to which people have to get used and which affect the position of current teachers, they may be key in explaining the quality of educational systems and the inequalities they produce. One policy that still can get huge support of citizens, but maybe less for policy makers because of its costs, is to increase the enrolment in pre-primary education. Especially high quality pre-primary education has the potential to reduce educational inequalities. Among others, early education takes much of the stress from disadvantaged parents and their children away and partly replaces the home-environment and offers high quality education instead (Raudenbush and Eschmann 2015; Kulic et al. forthcoming). Yet, children from advantaged families may learn most during pre-primary education because they make more use of it and can get most out of it, for example because they tend to go to better facilities with higher quality staff, more advanced peers, and more parent-teacher and parent-parents interaction, and because they may get most attention from teachers (Dearing, Kreider and Weiss 2008; Ribeiro et al. 2017; Cook et al. 2018). Although we do not have data on the performances of children attending early education, we could observe how differences in pre-primary school attendance at age 3 affected inequalities at grade 4 (approximately the age of 10). We found substantial evidence that inequalities are larger in systems in which more children attend pre-primary education, warning once more that it are these children that naturally get most out of education (cf. Bourdieu and Passeron 1977). In addition, we found substantial – but not decisive – evidence that an earlier start of compulsory education, that ensures access to early education for all children, also widens the gaps between children with few and many books at home and children with low, middle and high educated parents. This indicates that it is probably not only a matter of access to early education that increases inequalities, but also what care different children get. For a latter end age of compulsory education we found more mixed effects, underscoring that intervention later in life are often more limited in size (Oreopoulos 2007; Wößmann 2008; Murnane 2013; Raudenbush and Eschmann 2015).

The age of tracking age and the vocational specificity of educational systems are other characteristic aspects of educational systems. An early tracking age is often argued to help reproduce inequalities because it offers children from disadvantaged families less time to catch up with children from advantaged families and because advantaged parents are more often “far-sighted” in their choices “pushing” their children earlier to higher levels of education (Lucas 2001; Van de Werfhorst 2018). A substantial number of our findings support this finding, being in line with the majority of the findings in the literature (e.g. Ammermüller 2005; Breen and Jonsson 2005; Brunello and Checchi 2007; Van de Werfhorst and Mijs 2010; Van de Werfhorst 2018). Closely related to the age of tracking is the vocational specificity of educational systems. Even early tracking systems differ in the percentage of students that enroll in upper vocational education. Vocational education disproportionally attracts disadvantaged children and might both work as a safety net by offering education that fits the needs of these students (Shavit and Müller 1998, 2000), as well as a trap because job related skills are often learned at the costs of general skills such as literacy and mathematic skills (Hanushek et al. 2017; Kerckhoff 2001; Heisig and Olga 2015; Van de Werfhorst and Mijs, 2010; Bol and Van de Werfhorst 2013). Still much remains unknown about the consequences of a vocational system, and also our results leave it in the middle whether it disperses or reduces socioeconomic inequalities.

### 6.3.2. TACKLING INEQUALITIES BY MIGRATION BACKGROUND

Regarding children's migration background there were (even) more statistically not significant and mixed effects than with respect to parents socioeconomic status. Consequently, very few policies seem to be promising in reducing gaps in mathematics and reading performance between migrants and natives. Probably this is caused by the fact that, although we control for their socioeconomic status, we compare so many different immigrant groups, for example with respect to the countries of origin and reasons to migrate. To reduce those gaps policies are probably needed that take into account the specific characteristics of the host society and of the specific immigrant groups that are targeted. Thus, tailor-made policies are probably needed.

Migrants are often argued to potentially benefit from additional inputs in the educational system, such as larger expenditures, smaller classes, more teaching hours and better trained and more experienced teachers. These inputs are argued to help overcome the main handicap of migrants: that they are often less familiar with the peculiarities of the host societies culture and structure, and that they often are less fluent in the language spoken at school and in society. For example, smaller classes are argued to cause that migrants feel more at ease and it helps to get them extra attention, and better classified and more experienced teachers, might know better what immigrant student's need, have more patience with them, be more highly committed, and have better pedagogical skills to train them (e.g. Aaronson et al. 2007; Schneeweis 2011; Ammermueller 2013). Our within models show substantial support that smaller classes and higher educated teachers can help the performance in mathematics and reading more than those of the native majority. However, the between and pseudo panel models contest this idea, showing opposing or not significant effects. In addition, we find some support that in countries with more experienced teachers and higher salaries gaps in reading and mathematics performance between natives and migrants are larger, while for educational expenditure and teaching hours we find mixed results.

Greater inputs might not per definition benefit immigrant children most, because the native majority often makes more use of them. For example, the better educated and more experienced teachers disproportionally work at schools in which migrants are underrepresented (Borman and Kimball 2005; Clotfelter et al. 2010; Mansfield 2015). In addition, dissimilarities between teachers and children might explain part of the gaps between migrants and the native majority because similarity breeds trust and positive expectations, and higher educated and more experienced may be culturally more distant from migrants (Dee 2005). Thus, greater input may despite all good intentions result in unintended effects and might not decline and even increase the gaps between migrants and natives.

On a structural level, the most promising policy, and from which often most is expected, is perhaps early education. Due to early education children of migrants are argued to become sooner fluent in the language of the host society and to sooner get used to the cultural peculiarities of the host society. In turn, this can ease communication with other children and teachers, and facilitate further learning. This way, early education can stimulate the development of a broad set of skills, such as language and mathematics skills (e.g. Raudenbush and Eschmann 2015; Ribeiro, Zachrisson and Dearing 2017; Skopek and Passaretta 2018 Kulic et al. forthcoming). Our results partly support this theory. It appears that an increase in the number of children that enroll in pre-primary education in a country closes the gap in reading and mathematics performance between natives and migrants at our earliest observation, that is in grade 4. However, contrary to the expectations, when the compulsory age declines, which causes that more children attend pre-primary education, the performance of migrants in mathematics increases less than those of natives. Thus, some caution has to be taken into account about the success of pre-primary education.

Other structural factors of educational systems are the age of tracking and the vocational specificity. Migrants are argued to benefit from later tracking because this gives them time to catch up with natives and also to get used to the culture and educational system of the host society (Kirsten and Granato 2007; Alba and Waters 2011; Spörlein and Schlueter 2016). While often supported, research is inconclusive about its effects (Spörlein and Schleuter 2016). Our research shows that the performance of first generation migrants in reading increases more than those of natives if the tracking age increases, but not in mathematics. In addition, gaps do not become smaller between grade 4 (approximately age 10) and age 15 in systems in which students are later tracked. Thus, it is safe to assume that later tracking does not increase the gaps between migrants and natives, but it remains the questions whether it declines them. A *slightly* less promising strategy to decline gaps between migrants and the native majority is to decline the size of upper vocational education. The pseudo panel models show that in systems in which many children attend upper vocational education the performance of migrants in mathematics and reading increase less over their life course (up to the age of 15) than those of the native majority – except for reading for second generation migrants. This underscores the idea that vocational education can work as a trap for students because vocational skills are often taught at the expensive of general skills such as mathematics and language; skills that are needed for further development in their career (Bol and Forster 2017; Hanushek et al. 2017; Rözer and Bol 2019).

All in all, the most important lesson from this report is that there might be many unintended side-effects of educational policies. Educational policies often are implemented with the right intentions and have the potential to reduce inequalities, still they often do not or do the opposite. Advantaged children can make more use of these policies because of their cultural repertoire that is more appreciated in the educational system and because they learn more efficient. In addition, their parents can set up institutions that additionally support their children, and can bend these policies in their advantage, for example the use of pre-primary education or the use of better educated and more experienced teachers (e.g. Bourdieu and Passeron 1977; Lucas 2001). Therefore, policies should be designed that ensure not only the access for all, but also help disadvantaged groups to at least equally benefit from them, for instance by making teachers aware of the cultural practices and needs of these groups or by supporting the home environment and social networks of disadvantaged groups. Furthermore, this reports draws attention to structural factors, such as a later age of tracking and its vocational specificity. Despite that these policies might be more difficult to implement and may initially face a lot of criticism from citizens and teachers who have to get used to a new system, they might be crucial in tackling educational inequalities.

## 7. POLICY RECOMMMENDATIONS

- Governments implement a wide variety of policies to reduce educational inequalities and to use the potential of all students. Yet, we found that policies which are often meant to reduce inequalities had no statistically significant effects, or even showed that these policies widened educational inequalities. Advantaged groups often make more use of these policies and know better how to bend the current structures in their advantage. Consequently, policy makers should **not expect too much from general policies and better implement policies that are specifically targeted at disadvantaged groups to reduce educational inequalities**.
- Early education was one such policy for which several unintended consequences were observed, despite its popularity to decrease performances gaps between advantaged and disadvantaged children. **To ensure that early education reduces performance gaps two things appear to be crucial in the literature: assurance of equal access and the guarantee of high quality care for disadvantaged children.** High quality care goes beyond ensuring highly qualified staff for disadvantaged groups, but also that the staff is sensitive for their needs. Especially migrants are argued and shown to benefit from early education.
- Another policy which is well known to potentially have negative side-effects is the increase of teachings hours. The literature shows that **additional teaching hours are most likely to tackle inequalities when they are used to let disadvantaged children catch-up**, for example by repeating old material instead of learning something new.
- Raising the experience and education of teachers are shown to often increase the gaps between advantaged and disadvantaged children. **To make it more likely that raising the experience and education of teachers reduce performance gaps it needs to be assured that all children have equal access to them and that these teachers are sensitive for the needs of the disadvantaged.** This asks for more attention for the needs of children with a migration and low socioeconomic background in educational programs for (highly educated) teachers, and in extra training. These teachers might benefit from hands-on-tools to tackle those inequalities.
- To reduce performance gaps between migrants and the native majority governments could **reduce the number of students per class**. As argued in the literature, and largely supported by our results, this might especially increase the language performance of (first generation) migrants.
- There is ample evidence in the literature that later tracking is likely to reduce educational inequalities – something largely confirmed by our results. Hence, governments could work towards **increasing the age in which students are tracked by their abilities**.
- Disadvantaged groups disproportionally go to vocational education, while within vocational education vocational skills are often taught at the expensive of general skills, such as language and mathematics. These are crucial skills for life-long learning and flexibility. As such a great vocational specificity increases the risk of widening performance gaps and later chances in life. Thus, governments could **either reduce the size of vocational education or ensure that within vocational education general skills such as language and mathematics get sufficient attention**.
- Gaps in educational performance largely occur early in life. Yet, they also increase steeply after children leave secondary education. This suggests that not only much is to gain early in life, but also that **much is to gain later in life**. Lifelong learning might be one candidate to reduce these gaps.

## REFERENCES


Aaronson, D., Barrow, L. and Sander, W. (2007). Teachers and student achievement in the Chicago public high schools. *Journal of Labor Economics*, 25(1): 95-135.

Acemoglu, D. and Angrist, J. (2000). How Large Are Human Capital Externalities? Evidence from Compulsory Schooling Laws. *NBER Macroeconomics Annual*, 15: 9-59.

Achilles, C.M. (1999). *Let's put kids first, finally: Getting class size right*. Oaks: Corwin Press.

Alba R. and Waters M.C. (2011). *The next generation: immigrant youth in comparative perspective*. New York: New York University Press.

Alexander, K.L., Entwisle, D.R. and Olson, L.S. (2001). Schools, achievement, and inequality: A seasonal perspective. *Educational Evaluation and Policy Analysis*, 23(2): 171-191.

Allison, P.D. (2009). *Fixed effects regression models*. Los Angeles: Sage.

Ammermüller A. (2005). *Educational opportunities and the role of institutions*. ZEW Discuss. Pap. 05-44, Cent. Eur. Econ. Res., Mannheim.

Ammermüller A. (2013). Institutional Features of Schooling Systems and Educational Inequality: Cross-Country Evidence From PIRLS and PISA. *German Economic Review*, 14(2): 190-213.

Andon, A., Thompson, C.G. and Becker, B.J. (2014). A quantitative synthesis of the immigrant achievement gap across OECD countries. *Large-scale Assessments in Education*, 2(1): 1-20.

Angrist, J.D., and Lavy, V. (1999). Using maimonides' rule to estimate the effect of class size on scholastic achievement. *Quarterly Journal of Economics*, 114(2): 533-575.

Angrist, J.D. and Krueger, A.B. (1991). Does compulsory school attendance affect schooling and earnings? *The Quarterly Journal of Economics*, 106(4): 979-1014.

Aucejo, E.M., Romano, T.F. (2016). Assessing the effect of school days and absences on test score performance. *Economic Education Review*, 55: 70-87.

Bauer, P. and Riphahn, R.T. (2013) Institutional determinants of intergenerational education transmission: Comparing alternative mechanisms for natives and immigrants. *Labour Economics*, 25: 110-122.

Batruch, A., Autin, F., Bataillard, F. and Butera, F. (2018). School Selection and the Social Class Divide: How Tracking Contributes to the Reproduction of Inequalities. *Personality and Social Psychology Bulletin*, 45(3): 477-490.

Banerjee, A.V., Cole, S., Duflo, E., and Linden, L. (2007). Remedying education: evidence from two randomized experiments in India. *Quarterly Journal of Economics*, 122(3): 1235-1264.

Becker, G. S. (1994). *Human Capital Chicago.* IL: University of Chicago Press.

Becker, B. Klein, O. and Biedinger, N. (2013). The Development of Cognitive, Language, and Cultural Skills From Age 3 to 6: A Comparison Between Children of Turkish Origin and Children of Native-Born German Parents and the Role of Immigrant Parents' Acculturation to the Receiving Society. *American Educational Research Journal*, 50(3): 616 -649.

Bellei, C. (2009). Does lengthening the school day increase students' academic achievement? Results from a natural experiment in Chile. *Economic Education Review*, 28 (5): 629-640.

Berlinski, S. Galiani, S. and Gertler, P. (2009). The effect of pre-primary education on primary school performance. *Journal of Public Economics,* 93: 219 -234.

Biavaschi, C. et al. (2013). *Youth unemployment and vocational training.* Background paper for the World Development Report 2013. Technical report, World Bank.

Biedinger, N, Becker, B. and Rohling, I. (2008). Early ethnic educational inequality: The influence of duration of preschool attendance and social composition. *European Sociological Review*, 24(2): 243-256.

Blatchford, P., Bassett, P., Goldstein, H. and Martin, C. (2003). Are class size differences related to pupils' educational progress and classroom processes? Findings from the institute of education class size study of children aged 5 -7 years. *British Educational Research Journal* 29(5): 709-30.

Blatchford, P., Russell, A., Bassett, P., Brown, P. and Martin, C. (2007). The effect of class size on the teaching of pupils aged 7 -11 years. *School Effectiveness and School Improvement*, 18: 147-172.

Blatchford, P., Bassett, P. and Brown, P. (2011). Examining the effect of class size on classroom engagement and teacher-pupil interaction: Differences in relation to pupil prior attainment and primary vs. secondary schools. *Learning and Instruction,* 21: 715-730.

Boarini, R., Lüdemann, E. (2009). The role of teacher compensation and selected accountability policies for learning outcomes: an empirical analysis for OECD countries. *OECD Journal: Economic Studies Volume*, 211-230.

Bobak, M., Jha, P., Nguyen, S. and Jarvis, M. (2000). Poverty and smoking. In Jha, P. and Chaloupka, F. J. (Eds), *Tobacco Control in Developing Countries*. Oxford: Oxford University press.

Bol, T. and Van de Werfhorst, H.G. (2013). Educational systems and the trade-off between Labor market allocation and equality of educational opportunity. *Comparative Education Review* 57(2): 285-308.

Bol, T., Witschge, J., Van de Werfhorst, H.G. and Dronkers, J. (2014). Curricular tracking and central examinations: Counterbalancing the impact of social background on student achievement in 36 Countries. *Social Forces*, (1): 1-48.

Borgna C. and Contini D. (2014) Migrant achievement penalties in Western Europe: Do educational systems matter? *European Sociological Review*, 30(5): 670-683.

Borman, G.D. and Kimball, S.M. (2005). Teacher quality and educational equality: Do teachers with higher standards based evaluation ratings close student achievement gaps? *The Elementary School Journal*, 106(1): 3-20.

Bosworth, R. (2104). Class size, class composition, and the distribution of student achievement. *Education economics*, 22(2): 141-165.

Boudon R. (1974). *Education, Opportunity, and Social Inequality*. New York: Wiley.

Bourdieu, P. and Passeron, J. (1977). *Reproduction in education, society and culture*. London: Sage.

Bourdieu, P. (1984). *Distinction: A Social Critique of the Judgement of Taste*. London: Routledge.

Braga, M., Checchi, C. and Meschi. E. (2013). Educational policies in a longrun perspective. *Economic Policy*, 28(73): 45-100.

Bradbury, B., Corak, M., Waldfogel, J. and Washbrook, E. (2015). *Too many children left behind. The U.S. achievement gap in comparative perspective.* New York: Russell Sage Foundation.

Breen, R. (2005). Explaining cross-national variation in youth unemployment: market and institutional factors. *European Sociological Review,* 21: 125-34.

Breen, R. and Goldthorpe, J.H. (1997). Explaining Educational Differentials Towards a Formal Rational Action Theory. *Rationality and Society*, 9(3): 275-305.

Breen, R. and Jonsson, J.O. (2005). Inequality of Opportunity in Comparative Perspective: Recent Research on Educational Attainment and Social Mobility. *Annual Review of Sociology*, *31*(1): 223-243.

Brinbaum, Y. and Cebolla-Boado, H. (2007). The school careers of ethnic minority youth in France success or disillusion? *Ethnicities*, 7: 445-474.

Brito, N.H., Fifer, W.P. Myers, M.M. Eiott, A.M. and Noble, K.G. (2016). Associations among family socioeconomic status, EEG power at birth, and cognitive skills during infancy. *Developmental Cognitive Neuroscience*, 19: 144-151.

Brown, G., and Micklewright, J., (2004). *Using international surveys of achievement and literacy: A view from the outside*. Montreal (Canada): UNESCO Institute for Statistics.

Brown, G., Micklewright, J., Schnepf, S.V. and Waldmann, R. (2007). International surveys of educational achievement: how robust are the findings? *Journals of the Royal Statistics Society: Statistics A*, 170(3): 623-646.

Brunello, G. and Checchi, D. (2007). Does school tracking affect equality of opportunity? New international evidence. *Economic Policy*, 22: 781-861.

Buchmann, C. and Park, H. (2009) 'Stratification and the Formation of Expectations in Highly Differentiated Educational Systems', *Research in Social Stratification and Mobility* 27(4): 245-67.

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.

Camilli, G., Vargas, S., Ryan, S. and Barnett, W.S. (2010). Meta-analysis of the effects of early education interventions on cognitive and social development. *Teachers College Record*, 112: 579-620.

Carlsson, M., Dahl, G.B., Öckert, B. and Rooth, D.O. (2015). The effect of schooling on cognitive skills.

Review of Economics Statistics, 97(3): 533-547.
Caro, D.H., Mcdonald. J.T. and Willms, J.D. (2009). Socio-Economic Status and Academic Achievement Trajectories from Childhood to Adolescence. *Canadian Journal of Education,* 32(3): 558-90.
Carneiro, P. and Heckman, J. (2003). *Human capital policy*. In: Heckman, J., Krueger, A. (Eds.), Inequality in America: What role for human capital policies? MIT Press
Cattaneo, M.A., Oggenfuss, C. and Wolter, S.C. (2017). The more, the better? The impact of instructional time on student performance. *Education Economics,* 25(5): 433-445.
Cebolla-Boado, H., Radl, J. and Salazar, L. (2017). Preschool education as the great equalizer? A cross-country study into the sources of inequality in reading competence. *Acta Sociologica*, 60: 41-60.
Cheadle, Jacob E. (2008). Educational Investment, Family Context, and Children's Math and Reading Growth from Kindergarten Through the Third Grade. *Sociology of Education* 81(1): 1-31.
Checchi, D. and Van de Werfhorst, H.G. (2017). Policies, skills and earnings: how educational inequality affects earnings inequality. *Socio-economic Review*, 0(0): 1-24.
Chetty, R., Friedman, J.N. and Rockoff, J.E. (2014). Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood. *American Economic Review*, 104: 2633-2679.
Chmielewski, A.K. and Reardon, S.F. (2016). Patterns of cross-national variation in the association between income and academic achievement. *AERA Open*, 2(3): 1-27
Chmielewski, A.K. (2019). The Global Increase in the Socioeconomic Achievement Gap, 1964-2015. *American Sociological Review*, 84(3) 517–544.
Clotfelter, C.T., Glennie, E., Ladd, H. F. and 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(5-6): 1352-1370.
Clotfelter, C.T., Glennie, E., Ladd, H.F. and Vigdor, J.L. (2010). Teacher credentials and teacher achievement in high school. A cross-subject analysis with student fixed-effects. *The journal of human resources*, 45(3): 655-681.
Cobb-Clark, D. A., Sinning, M. and Stillman, S. (2012). Migrant youths' educational achievement: The role of institutions. *ANNALS of the American Academy of Political and Social Science*, 643(1): 18-45.
Coleman, J. (1996). Equality of educational opportunity. United States. Office of Education.
Conger, R.B. and Donnelan, M.B. (2007). An interactionist perspective on the socioeconomic context of human development. Annual Review Psychology, 58: 175-199.
Cook, K.D., Dearing, E. and Zachrisson, H.D. (2018). Is Parent -Teacher Cooperation in the First Year of School Associated with Children's Academic Skills and Behavioral Functioning? *International Journal of Early Childhood*, 50: 211-226.
Cooper, H., Nye, B., Charlton, K. et al. (1996). The e*ff*ects of summer vacation on achievement test scores: A narrative and meta-analytic review. *Review of Educational Research*, 66(3): 227-268.
Cornelissen, T., Dustmann, C., Raute, A. and Schönberg, U. (2018). Who benefits from universal child care? Estimating marginal returns to early child care attendance. *Journal of Political Economy,* 126(6): 2356–2409.
Cortes, K.E., Goodman, J.S. and Nomi, T. (2015). Intensive math instruction and educational attainment: long-run impacts of double-dose algebra. *Journal of Human Resources*, 50 (1): 108-158.
Crosnoe, R. (2007). Early child care and the school readiness of children from Mexican immigrant families. *International Migration Review*, 41(1): 152-181.
Crozier, G. and Davies, J. (2007). Hard to reach parents or hard to reach schools? A discussion of home-school relations, with particular reference to Bangladeshi and Pakistani parents. *British Educational Research Journal*, 33: 295-313.
Dämmrich, J. and Triventi M. (2016). From primary school to young adulthood: a cross-national analysis of cognitive competencies and related social inequalities. In: *Model of secondary education and social inequality: an international comparison* (eds. Blosffeld H.P., Bucholz, S., Skopek, J. and Triventi M.). Cheltenham UK / Northampton USA: Edward Elgar Publishing.
Dearing, E., Kreider, H. and Weiss, H. B. (2008). Increased family involvement in school predicts improved child -teacher relationships and feelings about school for low-income children. *Marriage*

*& Family Review, 43*(3-4): 226-254.
Dearing, E., Berry, D., amd Zaslow, M. (2006). Poverty During Early Childhood. In K. McCartney & D. Phillips (Eds.), *Blackwell handbooks of developmental psychology. Blackwell handbook of early childhood development* (pp. 399-423).
Deaton, A. (1985). Panel data from time series of cross-sections. *Journal of Econometrics*, 30(1-2): 109-126.
Dee, T.S. (2005). A teacher like me: Does race, ethnicity, or gender matter? *American Economic Review, 95*(2): 158-165.
De Feyter, J.J. and Winsler, A. (2009). The early developmental competencies and school readiness of low-income, immigrant children: Influences of generation, race/ethnicity, and national origins. *Early Childhood Research Quarterly*, 24(4): 411-431.
De Graaf, N.D., De Graaf, P.M. and Kraaykamp, G. (2000). Parental Cultural Capital and Educational Attainment in The Netherlands: A Refinement of the Cultural Capital Perspective. *Sociology of Education*, 73(2): 92-111.
De Graaf, P.M. (1986). The Impact of Financial and Cultural Resources on Educational Attainment in The Netherlands. *Sociology of Education*, 59(4): 237-246.
De Leeuw, J. and Kreft, I. (1986). *Journal of Educational Statistics*, 11(1), 57-85.
Desimone, L.M. and Long, D. (2010). Teacher Effects and the Achievement Gap: Do Teacher and Teaching Quality Influence the Achievement Gap Between Black and White and High- and Low-SES Students in the Early Grades? *Teachers College Record*, 112(12): 3024-3073.
Devereux, P.J. (2007) Small-sample bias in synthetic cohort models of labor supply. *Journal of Applied Econometrics*, 22: 839-848.
Dewey, J., Husted, T.A. and Kenny, L.W. (2000). The ineffectiveness of school inputs: a product of misspecification? *Economics of Education Review*, 19: 27-45.
Dicks, A. and Lancee, B. (2018). Double disadvantage in school? Children of immigrants and the relative age effect: A regression discontinuity design based on the month of birth. *European Sociological Review,* 34(3): 319-333.
DiStasio, V. (2017). 'Diversion or safety net?' Institutions and public opinion on vocational education and training. *Journal of European Social Policy*, 27(4): 360-372.
Dolton, P. and Marcenaro-Gutierrez, O.D. (2011). If you pay peanuts do you get monkeys? A cross-country analysis of teacher pay and pupil performance. *Economic Policy*, *26*(65), 5-55.
Domina, T. (2005). Leveling the home advantage: Assessing the effectiveness of parental involvement in elementary school. *Sociology of Education*, 78: 233-249.
Downey, D.B., Von Hippel, P.T. and Broh, B.A. (2004). Are schools the great equalizer? Cognitive inequality during the summer months and the school year. *American Sociological Review,* 69(5): 613-35
Dronkers, J., Van der Velden, R. and Dunne, A. (2011). The effects of educational systems, school-composition, track-level, parental background and immigrants' origins on the achievement of 15-years old native and immigrant students. A reanalysis of PISA 2006. *ROA Research Memorandum Series*.
Duncan G.J. and Magnuson K. (2013). *The long reach of early childhood poverty*. In Economic Stress, Human Capital, and Families in Asia, ed. W-JJ Yeung, MT Yap, pp. 57 -70. Dordrecht, The Neth.: Springer.
Dustmann, C. (2004). Parental background, secondary school track choice, and wages. *Oxford Economic Papers*, 56(2): 209-230.
Dustmann, C., Frattini, T. and Lanzara, G. (2012). Educational achievement of second-generation immigrants: an international comparison. *Economic Policy*, 27(69): 143-185.
Esping-Andersen G., Garfinkel, I., Han, W., Magnuson, K., Wagner, S. and Waldfogel, J. (2012). Child care and school performance in Denmark and the United States. *Children and Youth Services Review*, 34(3): 576-589.
Evans, G. W. and Kim, P. (2013). Childhood poverty, chronic stress, selfregulation, and coping. *Child Development Perspectives*, 7(1): 43-48.
Ewijk, R. and Sleegers, P. (2010). The effect of peer socioeconomic status on student achievement: A

meta-analysis. *Educational Research Review*, 5: 134-150.
Fan, X. and Chen, M. (2001). Parental involvement and student's academic achievement: A meta-analysis. *Educational Psychology Review,* 13(1):1-22.
Farbman, D. and Kaplan, C. (2005). *Time for a change: The promise of extended time schools for promoting student achievement*. Massachusetts 2020 educational opportunity.
Farkas, G. and Beron, K. (2004). The detailed age trajectory of oral vocabulary knowledge: differences by class and race. *Social Science Research,* 33(3): 464-97.
Fekjaer, S.N. (2007). New differences, old explanations: can educational differences between ethnic groups in Norway be explained by social back ground? Ethnicities, 7(3): 367-389.
Feinstein, L. (2003). Inequality in the early cognitive development of British children in the 1970 cohort. *Economica*, 70(1), 73-97.
Felfe, C., Nollenberger, N. and Rodríguez-Planas, N. (2015). Can't buy mommy's love? Universal childcare and children's long-term cognitive development. *Journal of Population Economic*s, 28: 393-422.
Feng, L. and Sass, T.R. (2018). The impact of incentives to recruit and retain teachers in "hard-to-staff" subjects. *Journal of Policy Analysis and Management*, 37(1): 112-135.
Fernald, A., Marchman, V.A. and Weisleder, A. (2013), SES differences in language processing skill and vocabulary are evident at 18 months. *Developmental science*, 16(2): 234 -248.
Figazzolo, L. (2007). *Impacts of PISA 2006 on the education policy debate.* Brussels: Education International.
Flores, A. (2007). Examining disparities in mathematics education: achievement gap or opportunity gap? *The High School Journal*, 91(1): 29-42.
Forster, A. and Bol, T. (2018). Vocational education and employment over the life course using a new measure of occupational specificity. *Social Science Research*, 70: 176-197.
Fredriksson, P., Öckert, B. and Oosterbeek, H. (2013). Long-term effects of class size. *Quarterly Journal of Economics*, *128*: 249-285.
Gal, I. and Tout, D. (2014). *Comparison of PIAAC and PISA frameworks for numeracy and mathematical literacy*. OECD Education Working Papers, no, 102. Paris: OECD publishing.
Georgieff, M.K. (2007). Nutrition and the developing brain: nutrient priorities and measurement. *American Journal of Clinical Nutrition*, 85: 614S-20S.
Geven, S. (2019). The impact of school tracking on school misconduct: variations by migration background in England, The Netherlands, and Sweden. *Journal of Ethnic and Migration Studies*, DOI: 10.1080/1369183X.2019.1600402.
Giesselmann, M. and Schmidt-Catran, A. (2018). Getting the within estimator of cross-level interactions in multilevel models with pooled cross-sections: Why country dummies (sometimes) do not do the job. *Sociological Methodology*. https://doi.org/10.1177/0081175018809150
Glazerman, S. and Seifullah, A. (2012). *An evaluation of the Chicago teacher advancement program (Chicago TAP) after four years.* Princeton, NJ: Mathematica Policy Research.
Goodman, A. and Sianesi, B. (2005). Early education and children's outcomes: how long do the impacts last? *Fiscal Studies*, 26(4): 513 -548.
Greenwald, R., Hedges, L.V. and Laine, D. (1996). The effect of schooling resources on student achievements. *Review of educational research,* 66(3): 361-396.
Griga, D. and Hadjar, A. (2014). Migrant Background and Higher Education Participation in Europe: The Effect of the Educational Systems. *European Sociological Review*, *30*(3): 275-286.
Guryan, J. (2001). *Does Money Matter? Regression-Discontinuity Estimates from Education Finance Reform in Massachusetts*. NBER Working Paper 8269, 2001
Gustafsson, J. (2003). What do we know about effects of school resources on educational results? *Swedisch Economic Policy Review*, 10: 77-110.
Hackman, D.A. and Farah, M.J. (2009) Socioeconomic status and the developing brain. *Trends in Cognitive Science*, 13: 65-73.
Hallinan, M.T. (1994). Tracking: From theory to practice. *Sociology of Education,* 67(2): 79-84.
Hanushek, E.A. (1981). Throwing money at schools. *Journal of Policy Analysis and Management*, 1: 19-41.

Hanushek, E.A. (1986). The economics of schooling: production and efficiency in public schools. *Journal of Economic Literature*, 24: 1141-77.
Hanushek, E.A. (1997). Assessing the effects of school resources on student performance: an update. *Educational Evaluation and Policy Analysis*, 19(2): S141-64.
Hanushek, E.A. (1977). Efficient estimators for regressing regression coefficients. *The American Statistician*, 28(2), 66-67.
Hanushek, E.A. (2003). The failure of input-based schooling policies. *Economics Journal*, 113(485): F64 -F98.
Hanushek, E.A. (2006). School resources. In Eric A. Hanushek, & Finis Welch (Eds.), *Handbook of the economics of education* (pp. 865-908). Amsterdam: Elsevier.
Hanushek, E.A., Kain, J.F., Rivkin, S.G. (1999). *Do Higher Salaries Buy Better Teachers*? NBER Working Paper #7082.
Hanushek, E.A., Kain, J.F., Markman, J.M. and Rivkin, S.G. (2003). Does peer ability a*ff*ect student achievement? *Journal of Applied Econometrics*, 18: 527-544.
Hanushek, R.A., Piopiunik, M. and Wiederhold, S. (2018). The value of smarter teachers: International evidence on teacher cognitive skills and student performance. *Journal of Human Resources*. doi: 10.3368/jhr.55.1.0317.8619R1
Hanushek, E.A. and Rivkin, S. (2005). Teachers, schools and academic achievement. *Econometrica* 73(2): 417-58.
Hanushek, E.A., and Wössmann, L. (2006). Does Educational Tracking Affect Performance and Inequality? Differences-in-Differences Evidence Across Countries. *The Economic Journal*, *116*(1984): 63-76.
Hanushek, E.A., and Wössmann, L. (2011). The economics of international differences in educational achievement. CESifo Working Paper, No. 3037.
Hanushek, E.A., and Wössmann, L. (2012). Do better schools lead to more growth? Cognitive skills, economic outcomes, and causation. *Journal of economic growth*, 17(4): 267-321.
Hanushek, E.A., Schwerdt, G. Wößmann, L. and Zhang, L. (2017). General education, vocational education, and labor-market outcomes over the life-cycle. *The Journal of Human Resources,* 52(1): 48-87.
Harn, B.A., Thompson, S.L. and Roberts, G. (2008). Intensifying instruction: Does additional instructional time make a difference for the most at-risk first graders? *Journal of Learning Disabilities*, 41(2),:115-125.
Harris, D.N., Sass, T.R. (2011). Teacher training, teacher quality and student achievement. *Journal of Public Economics*, 95(7): 798 -812.
Hart, B. and Risley, T. R. (2003). The Early Catastrophe. The 30 Million Word Gap. *American Educator*, *27*(1), 4-9.
Heath, A.F. and Brinbaum, Y. (2007). Guest editorial explaining ethnic inequalities in educational attainment. *Ethnicities*, 7: 291-304.
Heath, A.F., Rothon, C. and Kilpi, E. (2008). The second generation in western Europe: Education, unemployment, and occupational attainment. *Annual Review of Sociology*, 34: 211-235.
Heath, A. and Brinbaum, Y. (2014). The comparative study of ethnic inequalities in Educational careers. In A. Heath, Y. Brinbaum (Eds.), Unequal attainments: Ethnic educational inequalities in ten western countries, (p. 245). Oxford: British Academy.
Heath, A. and Brinbaum, Y. eds. (2014b). *Unequal Attainments: Ethnic Educational Inequalities in Ten Western Countries*. Oxford: OUP/British Academy | Proceedings of the British Academy 196.
Heckman JJ. (2000). Policies to foster human capital. *Research in Economics,* 54(1): 3-56.
Heckman, J.J. (2006). Skill formation and the economics of investing in disadvantaged children. *Science*, 30(312): 1900-1902.
Heckman, J.J. (2009).
Hendricks, M. (2014). Does it pay to pay teachers more? Evidence from Texas. *Journal of Public Economics*, 109: 50-63.
Heisig, J.P. and Schaeffer, M. (2017). The costs of simplicity: Why multilevel models may benefit from accounting for cross-cluster differences in the effects of controls. *American Sociological Review*,

82(4): 796-827.
Heisig, J.P. and Solga, H. (2015). Secondary education systems and the general skills of less and intermediate-educated adults: A comparison of 18 Countries. *Sociology of Education*, 88(3): 202 -225.
Heubener, M., Kuger, S. and Marcus, J. (2017). Increased instruction hours and the widening gap in student performance. *Labor Economics*, 47: 15-34.
Hillmert, S. and Jacob, M. (2010). Selections and social selectivity on the academic track: A life-course analysis of educational attainment in Germany. *Research in Social Stratifcation and Mobility, 28*(1), 59 -76.
Hsin, A. and Felfe, C. (2014). When does time matter? Maternal employment, children's time with parents, and child development. Demography, 51(5), 1867-1894.
Horn D. (2009). Age of selection counts: a cross-country analysis of educational institutions. *Education Research Evaluation,* 15: 343-66.
Horvat, E.M., Weininger, E.B. and Lareau, A. (2003). From social ties to social capital: Class differences in the relations between schools and parent networks. *American Educational Research Journal*, 40(2): 319-351.
Hoxby, C. (2000). *Peer effects in the classroom: Learning from gender and race variation.* NBER Working Paper No. 7867
Ichou, M. and van Zanten, A. (2014). *France*. In Stevens, P. A. J. and Dworkin, A. G. (Eds.), The Palgrave Handbook of Race and Ethnic Inequalities in Education. Basingstoke: Palgrave Macmillan, pp 328 -365.
Jackson, M., Jonsson, J.O. and Rudolphi, F. (2012). Ethnic inequality in Choice driven education systems: A longitudinal study of performance and choice in England and Sweden. *Sociology of Education*, 85(2): 158-178.
Jackson, C.K., Johnson, R.C. and Persico, C. (2016). The effects of school spending on educational and economic outcomes: Evidence from school finance reforms. *The Quarterly Journal of Economics*, 131(1): 157-218.
Jackson, C.K., Wigger, C. and Xiong, H. (2018). Do school spending cuts matter? Evidence from the great recession. NBER working papers: no. 24203.
Jakubowski, M. (2010). *Institutional tracking and achievement growth: Exploring difference-in-differences approach to PIRLS, TIMSS, and PISA data*. In J. Dronkers (Ed.), Quality and inequality of education (pp. 41 -81). Dordrecht, Netherlands: Springer Science and Business Media.
Jaime-Castillo, A.M. (2015). Modeling multilevel data. The estimated dependent *variable approach.* Spanish Stata users group meeting IE Business School, Madrid.
Jerrim, J. and Choi, A. (2013). The mathematics skills of school children: how does England compare to the high performing East Asian jurisdictions? *Journal of Education Policy*, 29(3): 349-376.
Jerrim, J. and Micklewright, J. (2014). Socio-economic Gradients in Children’s Cognitive Skills: Are Cross-Country Comparisons Robust to Who Reports Family Background? *European Sociological Review*, 30(6): 766-781.
Jez, S.J. and R.W. Wasmer (2015). The impact of learning time on academic achievement. *Education and Urban Society*, 47(3): 284-306.
Jonsson, J.O. amd Rudolph,i F. (2011). Weak performance - strong determination: School achievement and educational choice among children of immigrants in Sweden. *European Sociological Review*, 27(4): 487-508.
Kamerman, S.B., Neuman, M., Waldfogel, J. and Brooks-Gunn, J. (2003). *Social policies, family types and child outcomes in selected OECD countries* (OECD Social, Employment and Migration Working Paper 6). Paris: OECD.
Kane, T.J., Rockoff, J.E. and Staiger, D.O. (2008). What does certification tell us about teacher effectiveness? Evidence from New York City. *Economics of Education Review*, 27: 615-631.
Kao, G. and Tienda, M. (1995). Optimism and achievement: The educational performance of immigrant youth*. Social Science Quarterly*, 76(1): 1-19.
Kao G. and Thompson J. (2003). Racial and ethnic stratification in educational achievement and

attainment. *Annual Review of Sociology*, 29: 417-442.
Kerckhoff, A.C. (2001). Education and social stratification processes in comparative perspective. *Sociology of Education*, 74(Extra issue): 3-18.
Kirsten, C. and Granato, N. (2007). The educational attainment of the second generation in Germany: Social origins and ethnic inequality. *Ethnicities*, 7(3): 343-366.
Krueger, A.B. (1999). Experimental estimates of education production functions. *Quarterly Journal of Economics* 114(2): 497-532.
Kulic, N., Skopek, J., Triventi, M, and Blossfeld, H. (Forthcoming). Social background and children's cognitive skills. The role of early childhood education and care in a cross-national perspective, *Annual Review of Sociology*, forthcoming.
Lafortune, J., Rothstein, J. and Schanzenbach, D.W. (2018). School finance reform and the distribution of student achievement. *American Economic Journal*: Applied Economics, 10(2): 1 -26.
Lahaie, C. (2008). School readiness of children of immigrants: Does parental involvement play a role? Social Science Quarterly, 89(3), 684-705.
Lange, S. and Werder, M. (2017). Tracking and the intergenerational transmission of education: Evidence from a natural experiment. *Economics of Education Review,* 61: 59-78.
Lankford, H., Loeb, S. and Wyckoff, J. (2002). Teacher sorting and the plight of urban schools. A descriptive analysis. *Educational Evaluation and Policy Analysis*, 24: 37-62.
Lavy, V. (2015). Do differences in schools' instruction time explain international achievement gaps? Evidence from developed and developing countries. *Economic Journal*, 125(588): F397-F424.
Layte, R. and Whelan, C.T. (2009). Explaining Social Class Inequalities in Smoking: The Role of Education, Self-Efficacy, and Deprivation. *European Sociological Review*, 25(4), 399-410.
Le Donné, N. (2014). European variations in socioeconomic inequalities in students' cognitive achievement: The role of educational policies. *European Sociological Review, 30(3): 329-343.*
Lee, J.-W., Barro, R.J. (2001). Schooling quality in a cross-section of countries. *Economica*, 68(272): 465-488.
Lee, V E. and Burkam, D.T. (2002). *Inequality at the Starting Gate: Social Background Differences in Achievement as Children Begin School.* Washington, D.C.: Economic Policy Institute.
Leigh, A. (*2012).* Teacher pay and teacher aptitude. *Economics of Education Review,* 31: 41-53.
Lennon, M.L.M. and Tamassia, C. (2013). The development of the PIAAC cognitive instruments. In OECD (ed.) *Technical report of the survey on of adult skills (PIAAC)*. Paris: OECD publishing.
Leuven, E., Lindahl, M., Oosterbeek, H. and Webbink, D. (2004). *New evidence on the effect of time in school on early achievement* (Scholar Working Paper 47/04). Amsterdam: Research Institute Scholar.
Leuven, E., Lindahl, M., Oosterbeek, H. and Webbink, D. (2010). Expanding schooling opportunities for 4-year-olds. *Economic Education Review,* 29: 319-28.
Levels, M. and Dronkers, J. (2008). Educational performance of native and immigrant children from various countries of origin. *Ethnic and Racial Studies*, *31*(8): 1404-1425.
Levels, M., Dronkers, J. and Kraaykamp, G. (2008). Immigrant children's educational achievement in Western countries: origin, destination, and community effects on mathematical performance. *American Sociological Review*, 73: 835-853.
Levine, D., et al. (2018). Evaluating socioeconomic gaps in preschoolers' vocabulary, syntax and language process skills with the Quick Interactive Language Screener (QUILS). *Early Childhood Research Quarterly*, in press.
Lewis, J.B. and Linzer, D.A. (2005). Estimating regression models in which the dependent variable Is based on estimates. *Political Analysis*, 13(4): 345-364.
Lien, N., Friestad, C. and Klepp, K. (2001). Adolescents' proxy reports of parents' socioeconomic status: how valid are they? *Journal of Epidemiology and Community Health*, 55: 731-737
List, J.A., Samek, A. and Suskind, D.L. (2018). Combining behavioral economics and field experiments to reimagine early childhood education. *Behavioural Public Policy*, 2(1): 1-21.
Loeb, S. and Page, M.E. (2000). Examining the link between teacher wages and student outcomes: the importance of alternative labor market opportunities and non-pecuniary variation. *The Review of Economics and Statistics*, 82: 393-408.

Lucas, S.R. (2001). Effectively Maintained Inequality: Education Transitions, Track Mobility, and Social Background Effects. *American Journal of Sociology,* 106(6): 1642-90.
Lüdemann, E. and Schwerdt, G. (2013). Migration background and educational tracking Is there a double disadvantage for second-generation immigrants? *Journal of Population Economics*, 26(2): 455-481.
Lutz, A. (2007). Barriers to high-school completion among immigrant and later-generation Latinos in the USA. Language, ethnicity and socioeconomic status. *Ethnicities*, 7: 323-34.
Maasoumi, E., Millimet, D.L. and Rangaprasad, V. (2005). Class size and educational policy: Who benefits from smaller classes? *Econometric Reviews,* 24(4): 333-68.
Magnuson, K. A., Lahaie, C. and Waldfogel, J. (2006). Preschool and school readiness of children of immigrants. *Social Science Quarterly*, 87(5): 1241-1262.
Mandel, P. and Suessmuth, B. (2011). *Total instructional time exposure and student achievement: An extreme bounds analysis based on German state-level variation*. CESifo Working Paper Series. CESifo Group Munich.
Manning, A. and Pischke, J. (2006). *Comprehensive versus Selective Schooling in England and Wales: What Do We Know?* IZA Discussion Paper 2072. Bonn: Institute for the Study of Labor.
Mansfield, R.K (2015). Teacher quality and student inequality. *Journal of Labor Economics*, 33(3): 751-788.
Mare, R.D. (2014). Multigenerational Aspects of Social Stratification: Issues for Further Research. *Research in Social Stratification and Mobility*, 35: 121-128.
Marks, G.N. (2005). Cross-national differences and accounting for social class inequalities in education. *International Sociology*, 20(4): 483 -505.
Marks, G.N., Cresswell, J. and Ainley, J. (2006). Explaining socioeconomic inequalities in student achievement: The role of home and school factors. *Educational Research and Evaluation*, 12(2): 105-128.
McLoyd, V.C. (1998). Socioeconomic disadvantage and child development. *American Psychologist,* 53(2): 185-204.
Montt, G. (2011). Cross-national differences in educational achievement inequality. *Sociology of education*, 84(1): 49-68.
Morgan, P.L., Farkas, G. and Hibel, J. (2008). Matthew effects for whom? *Learning Disability Quarterly,* 31(4): 187-98.
Mortimore, P. (2009). *Alternative models for analysing and representing countries performance in PISA.* Brussels: Education International Research Institute.
Müller, W. and Gangl, M. (2003). *Transitions from education to work in Europe: the integration of youth into EU labor markets*. Oxford University Press Oxford.
Murnane, R.J. (2013). US high school graduation rates: Patterns and explanations. *Journal of Economic Literature*, 51: 370-422.
Nelson, C.A., Zeanah, C.H., Fox, N.A., Marshall, P.J, Smyke, A.T., and Guthrie, D. (2007). Cognitive recovery in socially deprived young children: The Bucharest Early Intervention Project. *Science*, 318(5858): 1937-1940.
Noble, K. G., Houston, S. M., Kan, E. and Sowell, E.R. (2012). Neural correlates of socioeconomic status in the developing human brain. *Developmental Science*, 15(4): 516-527.
Noble, K.G. et al. (2015). Family income, parental education and brain structure in children and adolescents. *Nature Neuroscience,* 18: 773-780.
Nores, M. and Barnett, W.S. (2010). Bene*fi*ts of early childhood interventions across the world: (Under) Investing in the very young. *Economics of Education Review*, 29: 27 -282.
Nye, B., Konstantopoulos, S. and Hedges, L.V. (2004). How large are teacher effects? Educational *Evaluation and Policy Analysis*, 26(3): 237-257.
Oakes, J. (1985). *Keeping Track*. New Haven, CT: Yale University Press.
Oreopoulos, P. (2007). Do dropouts drop out too soon? Wealth, health, and happiness from compulsory schooling. *Journal of Public Economics*, 91: 2213 -2229.
OECD (1998), Education at a Glance 1998: OECD Indicators, OECD Publishing, Paris.
OECD (2000), Education at a Glance 2000: OECD Indicators, OECD Publishing, Paris.

OECD (2001), Education at a Glance 2001: OECD Indicators, OECD Publishing, Paris.
OECD (2002), Education at a Glance 2002: OECD Indicators, OECD Publishing, Paris.
OECD (2003), Education at a Glance 2003: OECD Indicators, OECD Publishing, Paris.
OECD (2004), Education at a Glance 2004: OECD Indicators, OECD Publishing, Paris.
OECD (2005), Education at a Glance 2005: OECD Indicators, OECD Publishing, Paris.
OECD (2006), Education at a Glance 2006: OECD Indicators, OECD Publishing, Paris.
OECD (2007), Education at a Glance 2007: OECD Indicators, OECD Publishing, Paris.
OECD (2008), Education at a Glance 2008: OECD Indicators, OECD Publishing, Paris.
OECD (2009), Education at a Glance 2009: OECD Indicators, OECD Publishing, Paris.
OECD (2010), Education at a Glance 2010: OECD Indicators, OECD Publishing, Paris.
OECD (2011), Education at a Glance 2011: OECD Indicators, OECD Publishing, Paris.
OECD (2012), Education at a Glance 2012: OECD Indicators, OECD Publishing, Paris.
OECD (2013), Education at a Glance 2013: OECD Indicators, OECD Publishing, Paris.
OECD (2014), Education at a Glance 2014: OECD Indicators, OECD Publishing, Paris.
OECD (2015), Education at a Glance 2015: OECD Indicators, OECD Publishing, Paris.
OECD (2016), Education at a Glance 2016: OECD Indicators, OECD Publishing, Paris.
OECD (2017), Education at a Glance 2017: OECD Indicators, OECD Publishing, Paris.
OECD (2018), Education at a Glance 2018: OECD Indicators, OECD Publishing, Paris.
OECD (2010). *Learning for Jobs*. Paris: OECD.
OECD (2010). *Learning for Jobs*. Paris: Author.
OECD (2015). *Immigrant Students at School: Easing the Journals towards Integration (OECD Reviews of Migrant Education).* Paris: OECD.
OECD (2016). *Student learning time: A literature review.* OECD Education Working Papers, 127, OECD Publishing, Paris.
Österman, M. (2017). Tracking detracking reforms. Political explanations of institutional tracking in education. Downloaded from: diva-portal.org.
Papay, J.P. and Kraft, M.A., (2011). *Productivity returns to experience in the teacher labor market: methodological challenges and new evidence on long-term career growth*. Working Paper.
Papke, L E. (2005). The effects of spending on test pass rates: Evidence from Michigan. *Journal of Public Economics*, 89: 821-839.
Passaretta, G., and Skopek, J. (Eds.) (2018a). *Roots and Development of Achievement Gaps. A Longitudinal Assessment in Selected European Countries*. ISOTIS Report (D 1.3), Trinity College Dublin.
Passaretta G. and Skopek J. (2018b). From birth to the end of compulsory school – Social and migration-related inequality in a stratified education system. In: G. Passaretta and J. Skopek (eds.), Roots and Development of Achievement Gaps. A Longitudinal Assessment in Selected European Countries, pp. 18-49, ISOTIS Report (D 1.3), Trinity College Dublin.
Patall, E.A., Cooper, H., and Allen, A.B. (2010). Extending the school day or school year: a systematic review of research (1985 -2009). *Review Educational Research*, 80: 401-436.
Petitclerc, A., Côté, S., Doyle, O. Burchinal, M., Herba, C., Zachrisson, H.D., Boivin, M., Tremblay, R.E., Tiemeier, H., Jaddoe, V. and Raat, H. (2017). Who uses early childhood education and care services? Comparing socioeconomic selection across fve western policy contexts. *International Journal of Child Care and Education*, 11(3): 1-24.
Pfeffer, F.T. (2008). Persistent inequality in educational attainment and its institutional context. *European Sociological Review*, 24(5): 543-565.
Phalet, K., Deboosere, P. and Bastiaenssen, V. (2007). Old and new inequalities in educational attainment: ethnic minorities in the Belgian Census 1991-2001. *Ethnicities*, 7(3): 390-415.
Pischke, J. and Wachter, T. (2008). Zero returns to compulsory schooling in Germany: evidence and interpretation. *The Review of Economics and Statistics*, 90(3): 592-598.
Portes, A. and Rumbaut, R. G. (Eds.), (2001). *Legacies: The Story of the Immigrant Second Generation.* Berkeley, CA: University of California.
Potter, D. and Roksa, J. (2013). Accumulating Advantages over Time: Family Experiences and Social

Class Inequality in Academic Achievement. *Social Science Research,* 42(4):1018-32.
Prado, E.L. and Dewey, K.G. (2014). Nutrition and brain development in early life. *Nutrition Reviews*, special issue.
Preston, S.H. and Campbell, C. (1993). Differential Fertility and the Distribution of Traits: The Case of IQ. *American Journal of Sociology*, 98 (5): 997-1019.
Raudenbush, S.W. and Eschmann, R.D. (2015). Does Schooling Increase or Reduce Social Inequality? *Annual Review of Sociology*, 41: 443-470.
Raftery, A.E. and Hout, M. (1993). Maximally Maintained Inequality: Expansion, Reform, and Opportunity in Irish Education, 1921-75. *Sociology of Education,* 66(1): 41-62.
Raver, C.C, Blair, C. and Willoughby, M. (2013). Poverty as a predictor of 4-yearolds' executive function: New perspectives on models of differential susceptibility. *Developmental psychology*, 49(2): 292-304.
Reardon, S.F. (2011). The Widening Academic Achievement Gap Between the Rich and the Poor: New Evidence and Possible Explanations. Pp. 91-116 in *Whither Opportunity? Rising Inequality, Schools, and Children's Life Chances*, edited by Duncan, G.J. and Murnane, R.J. New York: Russell Sage Foundation.
Reardon, S. F. and Galindo, C. (2009). The Hispanic-White achievement gap in math and reading in the elementary grades. *American Educational Research Journal*, 46(3): 853-891.
Ribeiro, L.A., Zachrisson, H.D., Dearing, E. (2017). Peer effects on the development of language skills in Norwegian childcare centers. *Early Childhood Research Quarterly*, 41: 1-12.
Rikman, D.S., Wong, H. and Winters, J.V. (2017). Relative teacher salaries and the decision to teach. *Contemporary Economic Policy*, 35(3): 542-550.
Riphahn, R. T. (2003). Cohort effects in the educational attainment of second generation immigrants in Germany: an analysis of census data. *Journal of Population Economics*, 16: 711-733.
Rivkin, S.G. and Schiman, J.C. (2015). Instruction time, classroom quality, and academic achievement. *Economic Journal*, 125 (588), F425-F448.
Rivkin, S.G., Hanushek, E.A. and Kain, J.F. (2005). Teachers, schools and academic achievement. *Econometrica*, 73(2): 417-458.
Ruhose, J. and Schwerdt, G. (2016). Does early educational tracking increase migrant-native achievement gaps? Differences-in-differences evidence across countries. *Economics of Education Research*, 52: 134-154.
Rumberger, R.W. and Thomas, S.L. (2000). The distribution of dropout and turnover rates among urban and suburban high schools. *Sociology of Education*, 73(1): 39-67.
Rumberger, R.W. and Tran, L. (2006). *Preschool participation and the cognitive and social development of language minority students*. Los Angeles, CA: University of California.
Rockoff, J.E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *American Economic Review*, 94(2): 247 -52.
Roy, J. (2011). Impact of School Finance Reform on Resource Equalization and Academic Performance: Evidence from Michigan. *Education Finance and Policy*, 6: 137-167.
Rözer, J.J and Bol, T. (2019). Labor market effects of general and vocational education over the life-cycle and across time: accounting for age, period and cohort effects. *European Sociological Review, in press.*
Rözer, J.J. and Van de Werfhorst, H.G. (2017). *Inequalities in educational opportunities by socioeconomic and migration background: A comparative assessment across European societies.* ISOTIS report D1.2.
Salikutluk, A. (2016). Why do immigrant students aim high? Explaining the aspiration -achievement paradox of immigrants in Germany. *European Sociological Review*, 32 (5): 581-592.
Scafidi, B., Sjoquist, D. and Stinebrickner, T. (2007). Race, poverty, and teacher mobility. Economics of Education Review, 26: 145-159.
Schafer, W. and Olexa, C. (1971). *Tracking and Opportunity: The Locking-out Process and Beyond*. Scranton, PA: Chandler.
Scheerens, J. (2014). *Effectiveness of Time Investments in Education. Insights from a Review and Meta-Analysis*. Springer Briefs in Education.

Schneeweis, N. (2011) Educational institutions and the integration of migrants. Journal of *Population Economics*, 24(4): 1281-1308.

Schneider, B. and Coleman, J. S. (1993). *Parents, their Children, and Schools*. Boulder, CO: Westview.

Schröder, M. (2016). How income inequality influences life satisfaction: Hybrid effects evidence from the German SOEP. *European Sociological Review,* 32(2): 307-320.

Schubert, S. and Becker, R. (2010). Social inequality of reading literacy A longitudinal analysis with cross-sectional data of PIRLS 2001 and PISA 2000 utilizing the pair wise matching procedure. *Research in Social Stratification and Mobility*, 28: 109–133.

Schunck, R. (2013). Within and between estimates in random-effects models: advantages and drawbacks of correlated random effects and hybrid models. *Stata Journal*, 13: 65-76.

Schütz, G., Ursprung, H. W. and Wößmann, L. (2008). Education policy and equality of opportunity. *Kyklos*, 61(2): 279-308.

Shavit Y, Müller W. (1998). *From School to Work. A Comparative Study of Educational Qualifications and Occupational Destinations*. Oxford: Clarendon.

Shavit, Y. and Müller W. (2000) Vocational Secondary Education. Where Diversion and Where Safety Net? *European Societies*, 2(1): 29-50.

Shen, T. and Konstantopoulos, S. (2017). Class size effects on reading achievement in Europe: Evidence from PIRLS. *Studies in Educational Evaluation,* 53: 98-114.

Sirin, S.R. (2005). Socioeconomic Status and Academic Achievement: A Meta-Analytic Review of Research. *Review of Educational Research,* 75(3): 417-53.

Skopek, J. and Passaretta, G. (2018). *The Social Stratification of Skills from Infancy to Adolescence. Evidence from an Accelerated Longitudinal Design*. ISOTIS Working Paper.

Skopek J. and Passaretta G. (2018b). The evolution of social and ethnic inequalities in cognitive achievement from preschool to secondary schooling in the UK. In: G. Passaretta and J. Skopek (eds.), Roots and Development of Achievement Gaps. A Longitudinal Assessment in Selected European Countries, pp. 108-157, ISOTIS Report (D 1.3), Trinity College Dublin.

Smith, M.L. and Glass, G.V. (1980). Meta-analysis of research on class size and its relationship to attitudes and instruction. *American Educational Research Journal,* 17(4): 419-433.

Soskice, David. (1994). Reconciling Markets and Institutions: The German Apprenticeship System. Pp. 25 -60 in *Training and the Private Sector*, edited by Lisa M. Lynch. Chicago: University of Chicago Press.

Sowell, E.R. et al. (2003). Mapping cortical change across the human life span. *Nature Neuroscience*, 6: 309-315.

Spörlein, C., and Schlueter, H. (2016). How education systems shape cross-national ethnic inequality in math competence scores: Moving beyond mean differences. *Plos One*, 1-21.

Steele, J., Murnane, R. and Willett, J. (2009). Do financial incentives help low-performing schools attract and keep academically talented teachers? Evidence from California. *Journal of Policy Analysis and Management*, 29: 451-478.

Stephens, M., and Yang, D. (2014). Compulsory education and the benefits of schooling. *American Economic Review*, 104(6): 1777-1792.

Stören, L. A. and Heiland, H. (2009). Ethnicity differences in the completion rates of upper secondary education: how do the effects of gender and social background variables interplay? *European Sociological Review*, 26(5): 585-601.

Teltemann, J, and Schunk, R. (2016). 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(6): 401-424.

Tessaring, M. and Wannan, J. (2010). *Vocational education and training: Key to the future*. Luxembourg: European Commission.

Tillman, K. H., Guo, G. and Harris, K. M. (2006). Grade retention among immigrant children. *Social Science Research*, 35: 129-156.

Timmermans, A.C., De Boer, H. Amsing, H.T.A. and Van der Werf, M.P.C. (2018). Track Recommendation Bias: Gender, Migration Background and SES Bias over a 20-Year Period in the Dutch Context. *British Educational Research Journal*, 44(5): 847-74.

Trautwein, U., Lüdtke, U. Marsh, H.W., Köller, O. and Baumert, J. (2006). Tracking, grading, and student motivation: using group composition and status to predict self-concept and interest in ninth-grade mathematics. *Journal of Educational Psychology*, 98(4): 788-806.

Tucker-Drob, E.M. and Harden, K.P. (2012). Learning motivation mediates gene-by-socioeconomic status interaction on mathematics achievement in early childhood. *Learning and Individual Differences*, 22(1), 37-45.

UNESCO (2012). *World Data on Education. Seventh edition.*

UNESCO (2007). *World Data on Education. Sixth edition.*

Urquiola, M. (2006). Identifying class size effects in developing countries: Evidence from rural Bolivia. *The Review of Economics and Statistics*, 88(1): 171-177.

Van Houtte, M. and Stevens, P.A.J. (2008). Sense of futility: The missing link between track position and self-reported school misconduct. *Youth & Society,* 40(2): 245-64.

Van de Werfhorst H.G. (2014). Changing societies and four tasks of schooling: Challenges for strongly differentiated educational systems. *International Review of Education*, 60(1): 123-144.

Van de Werfhorst, H.G. (2018). Early Tracking and Socioeconomic Inequality in Academic Achievement: Studying Reforms in Nine Countries. *Research in Social Stratification and Mobility*, 58:22–32.

Van de Werfhorst H.G. (2018). Early tracking and socioeconomic inequality in academic achievement: Studying reforms in nine countries. *Research in Social Stratification and Mobility*, 58: 22-32.

Van de Werfhorst, H.G. (2019). Early Tracking and Social Inequality in Educational Attainment: Educational Reforms in 21 European Countries. *American Journal of Education*, forthcoming.

Van de Werfhorst, H.G. and Heath, A. (2018). Selectivity of migration and the educational disadvantages of second-generation immigrants in ten host societies. *European Journal of Population*, in press.

Van de Werfhorst, H.G. and Hofstede, S. (2007). Cultural capital or relative risk aversion? Two mechanisms for educational inequality compared. The British Journal of Sociology, 58: 391-415.

Van de Werfhorst, H.G. and Van Tubergen, F. (2007). Ethnicity, schooling, and merit in the Netherlands. *Ethnicities*, 7: 416-444.

Van de Werfhorst, H.G. and Mijs, J. J. B. (2010). Achievement Inequality and the Institutional Structure of Educational Systems: A Comparative Perspective. *Annual Review of Sociology*, *36*(1): 407-428.

Van Huizen, T. and Plantenga, J. (forthcoming). Universal child care and children's outcomes: A meta-analysis of evidence from natural experiments. *Forthcomming*.

Van Huizen, T. (2018). The Evolution of Achievement Gaps from Early Childhood to Adolescence in the Netherlands. In: G. Passaretta and J. Skopek (eds.), *Roots and Development of Achievement Gaps. A Longitudinal Assessment in Selected European Countries, pp. 50-87,* ISOTIS Report (D 1.3), Trinity College Dublin.

Waldinger, F. (2006). *Does Tracking affect the importance of family background on students' test scores?* Mimeo, London School of Economics, London, UK.

Warren, J.R. (1996). Educational inequality among white and Mexican-origin adolescents in the American South-West. *Sociology of Education*, 69: 142-158.

White, K.R. (1982). The Relation between Socioeconomic Status and Academic Achievement. *Psychological Bulletin,* 91(3): 461-81.

Wößmann, L. (2016). The importance of school systems: Evidence from international differences in student achievement. Journal of Economic Perspectives, 30(3): 3-32.

Wößmann, L., and West, M. (2006). Class-size effects in school systems around the world: Evidence from between-Grade variation in TIMSS. European Economic Review, 50(3): 695-736.

Wößmann L (2009). Efficiency and equity of European education and training policies. *International Tax Public Finance*, 15: 199-230.

Wößmann L (2009). International evidence on school tracking: A review. *CESifo DICE Report* 7(1): 26-34.

Zachrisson, H.D., Janson H., and Nærde, A. (2013). Predicting early center care utilization in a context of universal access. Early Childhood Research Quarterly, *28: 74-82*.

Zachrisson, H.D. and Dearing, E. (2015). Family income dynamics, early childhood education and care, and early child behavior problems in Norway. *Child Development*, 86(2): 425-440.

ISOTIS

INCLUSIVE EDUCATION AND SOCIAL SUPPORT
TO TACKLE INEQUALITIES IN SOCIETY

This project has received funding from the
European Union's Horizon 2020
research and innovation programme
under grant agreement No. 727069.