CHAPTER 1

Does curricular tracking explain global SES gaps? an international comparison of the SES achievement gaps from 2000 to 2015

Abstract: The literature on achievement inequality has recently started to focus on the dynamics of the socio-economic achievement gap in cognitive abilities. The main findings come from research in the U.S. revealing that the 90th/10th income achievement gap has widened by about 50% in the last 30 years. This chapter aims to investigate whether there are discernible patterns in the evolution of the achievement gap from a comparative perspective. Using over 15 years of data and 32 countries from the Program for International Student Assessment (PISA), I find that there is considerable variation in the way in which the gap is evolving, with the U.S. and Germany closing at about 50% and 30% in the last 15 years while France is widening at a similar rate. I find that curricular tracking and vocational enrollment explain 40% of the variance in the achievement gap between countries and show that the relationship is conditioned by a strong interaction. Low curricular tracking is associated with a small achievement gap, whereas high levels of curricular tracking is associated with wide achievement gaps. However, once tracking is coupled with high vocational enrollment this can remedy the potential adverse effects and reduce the gap by over 1 standard deviation. I use simulations to show that switching to less curricular tracking can help decrease a country's SES gap by about 11% while switching to more tracking would increase the achievement gap by about 51% percent.

1.1 Introduction

Inequality, be it economic or of non-pecuniary causes, is a crucial determinant of the opportunities a child will be exposed to. For the last decade or so, inequality has been rising and the literature on inequality has found evidence of increasing gaps between the rich and the poor in several dimensions.

Researchers have found evidence that, for example, for the U.S., economic inequality (Piketty and Saez, 2003), inequality of opportunity (Chetty et al., 2014) and inequality of achievement (Reardon, 2011) have increased in the last 20 years. Unfortunately, most of the studies to date mainly concentrate on the U.S.. It is still scarce to find studies of inequality from a comparative perspective, let alone to provide empirical explanations for the high inter-country variability.

This chapter looks to study the cognitive achievement gap between the 90th and 10th percentile of the socio-economic distribution (SES from now on). The gap can be referred to as the achievement gap between the top and bottom SES groups. Using the Program for International Student Assessment (PISA) survey, I calculate the achievement gap for all available countries that participated in all the six available PISA waves, building a country pseudo-panel that shows a time trend of 15 years.

More concretely, I aim to study the evolution of this achievement gap and a possible explanation as to why there are such big differences between countries. The two questions of interest are: how have achievement gaps changed over time and what explains such changes. Building on findings from studies in educational institutions, I suggest the degree of curricular tracking of a country and its level of vocational enrollments as possible explanations for the high variation between countries.

The findings from this chapter suggest that there is reasonable variation in the evolution of the gap. For example, the U.S. has been closing its 90/10 gap by about 50% with a lower bound of nearly 35%. The reduction is also present in other countries such as Germany and Poland. Other countries such as France and Austria experience the contrary with persistent increases in the achievement gap. The results show that these gaps differ very much in their composition, where in some countries the top 90th percent of the SES gap are benefiting more whereas in others the bottom group is catching up at very rapid pace. Robustness checks show that these results replicate for the 80/20 and 70/30 gaps providing more strength to the validity of the findings.

I find that the degree of curricular tracking of a country seems to be a possible explanation for the changes in achievement gap; however it must be studied by also factoring in the degree of vocational enrollment. The results suggest that if a country has low levels of curricular tracking the gaps are considerably smaller than when tracking is highly predominant in the curriculum. When tracking is present, the gap is bigger by around .50/.60 standard deviations. However, once curricular tracking enters the picture, vocational enrollment can help ease the burden of inequality and reduce the achievement gap significantly. Simulations show that if every country switched to little or no curricular tracking their gaps would shrink on average by about 11%.

The findings have two specific implications. First, researchers must avoid 'generalizing' or 'summarizing' inequality with one single indicator across many countries (Chmielewski, 2016). There are widespread contextual differences between countries, and each one should be studied with such details in mind. Second, further research should attempt to explain why some countries have experienced particular trajectories of inequality. I explore a possible explanation: the degree of curricular tracking and vocational enrollment of the country. Tracking seems to exacerbate inequality, but if coupled with a reasonable level of vocational enrollment it can help to alleviate the negative effects of curricular tracking.

The chapter first discusses the literature on the SES achievement gap as well as on the known evidence of the effects of curricular tracking. It then introduces the research question together with the methodology. In the first part of the results section, the chapter explores how gaps have changed over time while the second part implements the modeling section. The limitation section acknowledges the main limitations of the study and the chapter concludes with an overall discussion of the results.

1.2 Literature Review

Educational inequality and its long-term impacts are topics that have been prominent in the social science literature for the past 30 years. The idea of meritocracy and intergenerational transfers has motivated, for a good part of the 20th and 21st century, much of the research on social mobility and social inequalities. When James Coleman released his famous Coleman report (Coleman et al., 1966), he helped to show that a family's socioeconomic status and a student's performance are tightly linked. Since then, the topic has been studied extensively and several authors have contested whether the relationship is an invariable social law or a product of institutional arrangements [1]. Today, we have a much stronger understanding of the relationship.

Psychologists have been investigating child development for at least half a century and they find that the early stages in a child's life course are extremely important, if not the most important, for cognitive development and defining personality traits (Duyme et al., 1999; Waldfogel, 2006). The work of James Heckman helped to bring wider empirical attention to the subject. Heckman showed that cognitive and non-cognitive inequalities are present even before a student enters school (Heckman, 2006). To explain these, and other findings, Cunha et al. (2006) hypothesize that the cognitive level of a child at time t is a direct function of the experiences at time $t-1$. While it sounds straightforward, its importance is often missed. The model implies that investments into the education of a child compliment each other. It is difficult to easily compensate for an earlier lack of investment by investments at later stages. There are specific periods of skill formation in a child's life in which investment is particularly cost-effective. As a general rule, the earlier the investment in a child's education the greater the return. When tested against the data, Cunha et al. (2006) show that their explanation is consistent with other empirical studies.

Investments in individuals also support society as a whole as it boosts economic activity, it helps the labor market improve job conditions and maintain a rapid economic growth (Hanushek and Woessmann, 2007). Despite these findings and the recommended strategies to reduce the gap in cognitive achievement, we still find that a significant relationship between parental education and future destination is present in virtually all empirical studies of social mobility and inequality (Breen and Goldthorpe, 1997; Breen and Jonsson, 2007; Waldfogel, 2006; Bradbury et al., 2015; Chetty et al., 2016).

Since Coleman et al. (1966), educational researchers have spent most of their time studying the mechanisms through which this inequality has arisen. Naturally, they want to do that in order to reverse it and help every child reach his or her fullest potential. Despite the efforts, the literature has concentrated little on the magnitude of the gaps, especially in terms of cognitive abilities. We have little comparative evidence on which countries have large or small cognitive gaps resulting from SES origins. And even more importantly, we have not assessed whether policy efforts to reduce inequality have actually had an impact in reducing the achievement gap over time.

[1]For a detailed review of the long literature on educational inequality, please see Gamoran (2001).

Despite the shortcomings, we have indirect evidence of the relationships from the vast literature on social mobility. We know that virtually in all countries, developed or developing, there is inequality of opportunity. But there is considerable variation in the magnitude of inequality. For example, in the Scandinavian countries, particularly Denmark and Sweden, inequality is low compared to other countries and individuals have greater control over their destiny in terms of class mobility (Esping-Andersen et al., 2012; Breen and Jonsson, 2007; Shavit and Blossfeld, 1993).

From Denmark, researchers have learned a great deal about how to improve social mobility. But that was possible by first learning that Denmark is by far one of the most mobile countries compared to other European countries; this is the case specially for families of low-SES origin (Bjorklund and Jantti, 2009; Jaeger and Holm, 2007). The important finding came when research discovered the main reason behind their social escalator: its educational system. For example, research by Esping-Andersen et al. (2012) and Bauchmuller et al. (2014), shows that the Danish education system has important and longstanding impact on improving opportunities. The education system is completely subsidized for all children, otherwise giving opportunities to families who would not have been able to pay. In addition to this, Denmark is recognized as a world-leader in terms of public support for its child-care system as it spends around 2% of GDP, and has among the highest enrollment rates for children under 6 years old (Esping-Andersen et al., 2012). Moreover, it separates students into different curricular tracks late, at 16 years of age, compared to other European countries, something which has been linked to less educational inequality (Hanushek et al., 2006). These two traits make the Danish system effective in promoting equality relative to the other European countries. The early schooling experience attempts to put children on the same level and the process is not stopped by different curricular tracks as it starts at around age 16 when cognitive abilities are less malleable (Kautz et al., 2014). In short, the importance of this finding is that we should first study the presence and size of the effect and then proceed to find the causes behind it.

The first attempts to study the evolution of the achievement gap has found that the gap in cognitive abilities between high-SES and low-SES children has been widening over the years (Reardon, 2011). The literature on the topic has mainly concentrated on studying the case of the U.S. but other international evidence is emerging with a similar landscape. The U.S. is usually the case of study as it is the only country where cognitive testing is present from as early as 1940. Using this information, Reardon (2011) was the first to investigate the evolution of the cognitive gap and the results were surprising. Not only has the cognitive gap between the 90th income percentile and the 10th income percentile grown over time, but it has grown faster and to be wider than the highly contested white-black gap (Magnuson and Waldfogel, 2008). The gaps have actually reversed and we find that the income achievement gap is nearly twice as large as the black-white achievement gap (quite the opposite to 20 years back).

Reardon (2011) finds that the increase in the gaps has occurred predominantly from the 1970's until the 2000's. In fact, the hard numbers suggest that the gap widened by 40-50%. The author also estimates the rate of change using data as early as 1940 and finds an even higher increase of 75%. Given that the studies before 1970 are less reliable in terms of comparability and sampling design, the author computes all results for before/after 1970. To provide a definitive answer to the size of the gap, Reardon (2011) concludes that the 90/10 income gap in the U.S. has a standard deviation (SD from now on) of 1.25 in tests scores for the year 2001. Using longitudinal data, Bradbury et al. (2015) find similar results. Their empirical analysis suggests that for 14 year old Americans, there is a SD of above 1, but lower than 1.25. Surprisingly, Duncan and Magnuson (2011) find results similar to the previous studies and confirm a gap with a SD of 1.25-1.50. One important drawback of some of these studies is that they fail to present the uncertainty of these estimates. Not necessarily to gauge their statistical significance, but to simply assess how much we can trust their accuracy. It could be that the gap is at 1.25 standard deviations, varying up to 1.75 and down to 0.80. Right now, we do not know the upper/lower bound of this estimate, making it difficult to compare when new evidence arises.

Interestingly, the widening of the achievement gap has been paralleled by a growth of income inequality, which may be telling . Reardon (2011) offers several possible links, with the most reasonable being that family investment patterns have changed so that high income families now invest more resources on their children. The explanation lies in the fact that increasing income became more strongly correlated with other positive family traits related to time allocation and welfare services.

In a follow-up study, Reardon and Portilla (2015) uncovered a reversal of the trend. The follow-up study concentrated solely on kindergarten children in the U.S. for the years 1998, 2006 and 2010. They found that the 90th/10th income gap in readiness closed modestly. Furthermore, using data from fall and spring in the same kindergarten year, they calculated that the gap narrowed at a rate of 0.01 and 0.008 SD per year for mathematics and literacy between 1998 and 2010. They also calculated the same changes for a number of personality traits such as self-control and externalizing behavior and found similar results. In contrast, Reardon (2011) finds that in a 30-year span the gap was systematically increasing at a rate of 0.02, something reasonably close to the previous estimates. Their results not only hold for the income achievement gap, but they also found a decline in the white-hispanic gap (although not for the white-black gap). It should be noted that perhaps the reversal of the trend in Reardon (2011) would had been evident if data were available for years after 2000, the time-point from which Reardon and Portilla (2015) find the reversal.

The reasons why the authors find a reversal in the trend could be numerous and should be studied closely. They incorporate a number of country-level indicators to explain this change and suggest that the reversal is likely due to the high increase of preschool enrollment. They build on their previous argument by suggesting that in this same period (1998 - 2010) the income achievement gap in early schooling enrollment decreased substantially. Their conclusions, although suggestive, are speculative and have no empirical support which is why this is still an open question.

Motivated by these recent results, other authors have taken this analysis to an international context in order to discover between-country trends. The work of Bradbury et al. (2015) employs a unique comparative analysis of the achievement gap between Australia, United Kingdom, United States and Canada. Their research design is distinctive in that they use longitudinal data from children as early as age 2 and study the evolution of the achievement gap up until age 14 [2]. The core finding of their study is that the American achievement gap is much wider than the gaps in Australia and Canada. They find that once the achievement gap is present in early school entry, it does not seem to narrow or widen much over the life course. In fact, they estimate that the quality of early childhood education can only explain about 30-40% of the high school SES gap. This suggests that once the achievement gap is present before entering school, it carries a social-scar effect [3]. One exception is the UK, which they found to be a country that helps close the gap in early primary years. This can likely be due to the comprehensive schooling and also the public support by the welfare state in dimensions like health and income support.

[2]To the best of my knowledge this is not only the first study that uses panel data to study achievement gaps, but to also do it between countries

[3]However, schooling could be preventing the gap from widening even more, and rigorous Randomized Controlled Trials (RCT) show that high quality schooling can indeed help ease the gap, in some instances even close it (Campbell et al., 2002).

One limitation of their study is that they concentrate on countries which have a particular educational structure, namely the fact that there is little formal stratification in terms of curricula. The four countries have no major jump in school selection, the opposite to the average European country. A thorough review by Van de Werfhorst and Mijs (2010) sheds some light on the subject. First and foremost, they gather substantive evidence showing that countries which have a highly tracked curriculum tend to have high levels of inequality, measured in terms of achievement gaps. Hanushek et al. (2006) use the Progress in International Reading Literacy Study (PIRLS), the Trends in International Mathematics and Science Study (TIMSS) and the Programme for International Student Assessment (PISA) surveys to gauge whether highly tracked countries do indeed increase inequality after students pass the age at first selection of curricular tracking. The results suggest that early selection increases educational inequality. While less clear, there is also a tendency for early tracking to reduce mean performance. Micklewright and Schnepf (2007) using PISA but a different empirical strategy find that countries which have a high level of curricular tracking, are distinctively unequal in the difference between the top 95th and bottom 5th performers. In fact, the difference in test scores between these two groups is about 10 times higher than the average annual gain of a year of schooling.

Another limitation of the work of Bradbury et al. (2015) is that their analysis is based on four surveys that have significant differences in terms of questions, sampling and populations and cannot be easily compared. They manage to harmonize the four surveys into a comparable format and their findings do seem to be reliable. But we should be careful at interpreting these findings causally without considering that the four surveys carry great deal of differences in terms of measurement and survey questions. For this reason, we should also pay particular attention to studies such as Chmielewski and Reardon (2016) and Chmielewski (2016) who have attempted to compare gaps between countries, and to evaluate whether there is a general increase in the gap using comparable surveys. However, it should be noted that these studies tackle a completely different question from the above, namely to study cross-sectional differences between many countries, instead of over-time analysis of student gaps. Nonetheless, they do provide support for the overall finding that the achievement gap is widening over time.

Chmielewski and Reardon (2016), again using PIRLS, TIMSS and PISA, assess whether there are patterns of cross-national variation in the achievement gap. In other words, does the achievement gap differ between countries? Their work suggests that there is considerable variation in the achievement gap between top and bottom earning families across many developed countries. In comparison to the literature on achievement gaps, they find that the U.S. has a gap of 1.20 SD in 2001 which increase to around 1.30 in the year 2006 while Germany has a decreasing gap from around 1.25 to 1 SD in the same year-span. However, these numbers vary a lot and carry a great deal of overlapping uncertainty.

They go even further and link this achievement gap to several country-level indicators related to income inequality, school differentiation and central exams, among other things. These correlations are suggestive as explanatory mechanisms but they are cautious in drawing causality.

One interesting question that is still missing from the literature is how these country gaps have evolved over time. With their data, Chmielewski and Reardon (2016) have only 3 countries which are present in all waves and also have very few waves as their question of interest (income categories) was only asked at three time points. The results are more about between country gaps rather than the magnitude and evolution of the gaps.

Chmielewski (2016), building on the work of Chmielewski and Reardon (2016) and Reardon and Portilla (2015) pooled together all the previously mentioned data, together with over 10 more studies ranging from 1964 to 2015 in order to discover differences between countries. With over 50 years of data, and over 100 countries, the author finds that there seems to be a general pattern of increasing achievement gap. However, once she disentangles the relationship by country, she finds a sizable amount of heterogeneity, with some countries experiencing a narrowing of the achievement gap, others no change at all, while others record a steady increase. This is revealing as it does not really pay off to look at a general average once each country has their own distinctive gap and evolution. This highlights the notion that the increasing achievement gap is clearly not universal and should be studied in context.

One limitation of the study of Chmielewski (2016) (as well as Reardon (2011)) is that they adjust for the age of each child in all studies. Although for their modeling purposes this is the correct thing to do [4], these modeling strategies are masking age-specific achievement gaps by controlling for age. We clearly see in Reardon and Portilla (2015) that there are age-specific gaps, and they do change at a fast pace in little time.

In fact, the evolution of high/low SES gaps for preschool children might be much less

[4]The differences in achievement could simply be due to changes in cognitive abilities across the lifetime. However, as we have noted before, Bradbury et al. (2015) find that the achievement gap is stable across the life time

marked than the same gap for high school children. The explanation, although debated, has been gaining much support in recent years. For countries with high levels of curricular differentiation the transition from early schooling into the tracking system has been found to increase inequality of learning (Hanushek et al., 2006). Moreover, the vast sociological literature on educational transitions systematically finds that early selection tends to foster between-track inequality rather than erode their differences by tackling their specific needs (Van de Werfhorst and Mijs, 2010). Based on this, it is difficult to simply assume that the achievement gap has been neither constant across cohorts nor the same between ages, as tracking/no tracking might exacerbate the achievement gap.

1.3 Research questions and hypothesis

The overarching aim of this chapter is to study the evolution of the high/low SES achievement gap in the past 15 years for all PISA participating countries and propose a likely explanation for the evolution.

I develop the sub-questions and their corresponding hypothesis separately in more detail.

1. The seminal work of Reardon (2011) suggests that achievement gaps change, and they do so much quicker than we thought after recording a SES gap increase of about 40% in 30 years. Reardon and Portilla (2015) stress that they also found a significant decrease in only 15 years of data, showing how important it is to study the changes in the achievement gap. First, I will concentrate on the evolution of the gap only for 15-year olds. As we have seen before, there are reasons to consider specific age-groups when estimating achievement gaps as there might be differences between age groups. This will serve as a comparison to the single year-country snapshot of Chmielewski and Reardon (2016) and the evolution of the kindergarten gap in Reardon and Portilla (2015). Secondly, I will compare the percentage change at which the gap widened/narrowed from the first to the last year available. This will give a general idea of the overall change over time. [5]. I posit no specific hypothesis for this question given that it is purely explorative.

    - Research question 1:

[5] Although no study has performed this age-specific achievement gap for comparable tests over such a long time. The results will serve as comparison for other studies that use age-specific groups, such as 4th graders using PIRLS.

(a) How is the achievement gap changing between countries?

2. The literature has concentrated narrowly on whether the gap is increasing because the top performers are getting ahead, because the bottom performers are falling behind or because both are changing at the same time. The work of Bradbury et al. (2015) is the only study that pays attention to the source of the achievement gap that I am aware of. The findings are heterogeneous for the four countries in the study but the overall evidence shows that as children grow older, top and bottom SES groups seem to grow apart at a similar rate. These findings are very important for understanding inequality through out the life course of individuals but this still does not answer whether specific age groups have gaps that change over time. Moreover, their analysis is limited to four countries that have very little to no institutionalized curricular tracking. This type of question is important because it highlights whether there are specific policies that might prevent gaps from closing over time.

   - Research question 2:
     - Is the gap originating as the top is gaining ground, the bottom is falling behind or because of a dynamic interaction between the two?
   - Hypothesis 1:
     - The theoretical argument in favor of curricular tracking posits that in countries where there is a high degree of tracking we should expect the top and bottom to be evolving at a similar rate given that curricular tracking is thought to maximize the learning experience of both groups. This is exposed in the review of Van de Werfhorst and Mijs (2010). However, the results from Hanushek et al. (2006) suggest otherwise as tracking seems to be associated with greater inequality between SES groups. That being said, I expect to find that countries with highly tracked curriculums have growing achievement gaps as children from low-SES groups are at a greater disadvantage. Conversely, countries with low curricular tracking are expected to be associated with narrowing gaps as both SES groups are thought to be less segregated and thus equalizing their learning experiences.

3. The work of Chmielewski (2016) shows that there are differences between countries in their overall levels of achievement gaps. This is different from the previous

hypothesis because this chapter concentrates only on age-specific gaps rather than age-adjusted gaps. Given that most countries available in PISA participated in all six waves, this question attempts to find a possible explanation for why there are such stark differences in the SES achievement gap between countries. Chmielewski and Reardon (2016) perform a similar analysis but they concentrate only on *income* achievement gaps rather than on a more broad SES index. Moreover, they only perform their analysis on a handful of countries which limits the amount of between-country variability. This question tests whether several dimensions of tracking and vocational enrollment explain the differences in achievement gaps between countries. This is discussed in detail in Van de Werfhorst and Mijs (2010) and Hanushek et al. (2006) but not empirically tested before in an international context.

- Research question 3:
  - Does curricular tracking and vocational enrollment explain differences between countries in achievement gap?
- Hypothesis 2:
  - According to the review by Van de Werfhorst and Mijs (2010), the tracking setup of a country should play an important role in explaining the marked differences in achievement gaps between countries seen in Chmielewski (2016). I expect to find that a reasonable percentage of the variation between countries is explained by the fact that some countries have highly institutionalized tracking setups, while other countries have more flexible tracks.

4. The reasons why the gap is changing are still speculative. Some researchers have pointed out to the share of preschool enrollment as a possible mechanism (Reardon and Portilla, 2015) while others have tested the degree of economic inequality within a country (Chmielewski and Reardon, 2016). This chapter explores a different mechanism thanks to the work of Van de Werfhorst and Mijs (2010). This questions looks to uncover whether the curricular tracking setup of a country is also a possible explanation for the *evolution* of the achievement gap. This has theoretical and empirical justifications given that tracking might exacerbate achievement gaps once it is implemented (Van de Werfhorst and Mijs, 2010). This is because tracking can have dynamic effects given that the quality of instruction in different tracks

can decrease/increase over time even if tracking indicators stay the same. However, research has not explored whether curricular tracking is operating in combination with other institutional features. In this question I look to explore whether tracking, together with the degree of vocational enrollment in a country are valid explanations for the changing achievement gap within a country.

- Research question 4:
  - Are curricular tracking and vocational enrollment related to the evolution of cognitive achievement gap?
- Hypothesis 3:

  Building on the previous hypothesis, curricular tracking should play an important role on the evolution of the achievement gap. However, the relationship is not straight forward because tracking hardly changes over a period of 15 years. Curricular tracking might exacerbate gaps through mechanisms which are not seen through traditional tracking indicators. For instance, if the quality of instruction in the lower tracks worsen over time. I hypothesize that the degree of tracking of an educational system is tightly related to changes in the gap, and the more curricular tracking, the more inequality. Moreover, the more vocational curricular tracking, the less inequality considering that it gives short term returns in terms of labor market opportunities.

1.4 Methods

1.4.1 Data

To investigate the above mentioned questions, I will use the Programme for International Student Assessment (PISA). PISA is a survey carried out every three years that aims to evaluate education systems by testing the skills and knowledge of 15-year-old students. Currently, PISA has six waves starting in 2000 up until 2015, where recently, over half a million students were tested in mathematics, literacy and science in over 70 developed/developing countries.

PISA collects data through a two-stage stratified sampling design. With the help of governments, PISA randomly chooses 150 schools in each country, where they then randomly pick thirty 15 year olds to undertake the two hour tests. The sample size for each of the waves are 127,388 for PISA 2000, 276,165 for PISA 2003, 398,750 for PISA 2006, 515,958 for PISA 2009, 480,174 for PISA 2012, and 519,334 for PISA 2015. Together with the subject tests, PISA collects personal information from students, their families and their school environment (including teacher surveys), that serves as relevant background information that can be matched to the students performance. With the recent inclusion of PISA 2015, these six waves make up a time-series analysis of 15 years, enough to visualize changes in the structure of an educational system. None of the studies cited so far has used the last PISA wave, which was recently released in December 2016. This chapter takes advantage of these six waves to build a country pseudo-panel where each country-wave will have its corresponding achievement gap, making it possible to study changes in nearly 15 years for 32 countries.

To identify a family's SES, PISA collects several variables that measure different dimensions. Classically, they ask student's their parent's educational level. Scholars have considered this to be a reliable recall given that we expect fifteen year olds to know their parent's level of education (Reardon, 2011). This question has been asked in every wave and holds a somewhat similar coding across time, although the first two waves have small differences. In spite of this, a serious limitation is the fact that parent's education is measured using the ISCED classification, something that has changed over time. For example, until PISA 2009, the preferred framework was ISCED 1997, whereas the next wave switched to the newly developed ISCED 2011 classification. Both these classification schemes have equivalent look-up tables, but this requires a detailed inspection of the codings.

Another social background variable is the International Socio-Economic Index of Occupational Status (ISEI). This variable attempts to capture the social status of the family, without asking for income information. This index variable was developed by Ganzeboom and Treiman (1996) and later refined by Ganzeboom (2010) and it attempts to measure occupational status using a continuous measure. The indicator is a reliable alternative to the classical Erikson-Goldthorpe-Portocarero classification (Erikson et al., 1979). It has been scaled for comparability between waves and some authors have used it for inequality studies, finding expected results to be consistent with the social mobility literature (Chmielewski, 2016). PISA also includes a plethora of indicators on family wealth, home educational resources, the number of books in the home, among many other material resources in the household.

Yet one of the most relevant variables for this study is a composite SES index created by the PISA team. The index of economic, social and cultural status (ESCS) was created on the basis of the following variables: the International Socio-Economic Index of Occupational Status (ISEI), the highest level of education of the student's parents, the PISA index of family wealth (which measures the material wealth of the family), the PISA index of home educational resources; and the PISA index of possessions related to "classical" culture in the family home (mainly about books in the household) (OECD, 2002). The variable, aside from capturing all relevant dimensions of SES, such as education, occupation, and material resources, takes care of transforming all mentioned variables into comparable metrics across waves.

The ESCS index was derived from a principal component analysis of standardized variables, taking the factor scores for the first principal component as measures of the PISA index of economic, social and cultural status. All countries and economies (both OECD and partner countries/economies) were assigned the same weight in the principal component analysis, while in previous cycles, the principal component analysis was based only on OECD countries. However, for the purpose of reporting, the ESCS scale has been transformed with zero being the score of an average OECD student and one being the standard deviation across equally weighted OECD countries (OECD, 2016).

To the best of my knowledge this is the first piece of research that uses the newly-released ESCS index (OECD, 2016), which was rescaled so that all ESCS indexes are suitable for over-time analysis [6]. In other words, the ESCS index does not need any transformation or coding updates as it is ready for comparison over time.

Aside from SES, the other relevant variables are test scores for mathematics and literacy. PISA does not provide a single test result for each respondent. Instead, it provides a *series* of 'plausible values' that the child could actually score. As explained in the PISA manual (OECD, 2012), these are imputed values that resemble individual test scores and have approximately the same distribution as the latent trait being measured (the true distribution of the possible scores a student can achieve).

A more intuitive explanation is this: suppose we have μ_i, the average student test score in mathematics for student i. Instead of estimating μ_i alone, plausible values estimate a distribution of possible μ's for student i, together with the likelihood of each μ_i based on

[6] These rescaled indexes can be found at http://www.oecd.org/pisa/data/2015database/ under *Rescaled indices for Trend Analyses.*

the respondents answers on the test. This is defined as the posterior distributions of μ's for student i. The reason why PISA uses this procedure is because estimating a single number μ_i is plagued with measurement error, among other types of bias (see Wu, 2005). The number of plausible values for PISA waves are usually five (although ten for PISA 2015) random draws from this distribution. In practice, each student has 5 scores for each test, which resembles their distribution. Those values are continuous, ranging from 0 to 500, with a mean of 250.

1.4.2 Coding and methodology

The aim of this chapter is to identify, disaggregate and explain country trends in the achievement gap for several countries. To represent the SES gap, most of the literature on achievement gaps has concentrated on indicators such as parental education, parental occupational status, income achievement gaps and actual SES achievement gaps (Fryer and Levitt, 2004; Hanushek et al., 2006; Saw, 2016; Bradbury et al., 2015; Byun and Kim, 2010). The actual calculation of the achievement gap varies substantially and different strategies have been implemented. For example, Micklewright and Schnepf (2007) calculate the difference in achievement by crudely subtracting the difference between the 95th and 5th percentile of the mathematics distribution. Although in principle you should be able to capture some type of SES effect like this, theoretically, it should be much more accurate to difference out the mean score of, for example, parental education or some other SES proxy. Saw (2016), for instance, used parental education as a proxy of SES, whereas Byun and Kim (2010) use a similar SES index as the one used in this chapter, but created by them.

Reardon and Portilla (2015), Chmielewski and Reardon (2016) and Chmielewski (2016) used a different method developed by Reardon (2011), which I partially adopt in this chapter. SES achievement gaps are measured as the difference in standardized achievement between the 90th and 10th percentiles of the chosen SES variable. The rule of thumb to choose the 90th, 50th and 10th percentile is arbitrary, as others have used, for example, the 95th, 50th and 5th (Micklewright and Schnepf, 2007). I use the conventional 90/10 cutoffs in the literature following the standard set by Reardon (2011).

For each country in each wave, SES disparities in achievement are measured as the gap in standardized achievement between the 90th and 10th percentiles of each country's distribution of each SES variable, following the method for income achievement gaps by

Reardon (2011). The *original* strategy of Reardon (2011) is as follows: first, achievement is standardized (see below for a statistical explanation of the standardization). I then use it to calculate the mean achievement (and standard error) for each category of the SES variable of interest (parent's education, income categories, etc..). "Category means are plotted at their percentile ranks and cubic models are fit through the points using weighted least squares. Finally, achievement at each country's 90th and 10th SES percentiles is interpolated from the model" (Chmielewski, 2016). The result is an SES gap from an ordinal variable of interest.

As mentioned before, PISA does not provide a single achievement indicator. Instead, I calculate the median of all plausible values for each student [7], resulting in one single score.

To standardize the test score I fit a linear model

$$y_i = \alpha + \beta_1 * AGE_i + \epsilon_i, \quad \epsilon_i \; N(0, \sigma^2) \tag{1.1}$$

for each wave, where y_i is the median student test score for student i and AGE_i is their age measured in months (following the same strategy as Reardon (2011) [8]) weighted by the student sample weights from PISA [9].

I then calculate $\hat{\gamma}_i$ by

$$\hat{\gamma}_i = \frac{\hat{\epsilon}_i}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}} \tag{1.2}$$

Where $\hat{\epsilon}_i$ is the residual for student i, $\hat{y}_i$ is the predicted test score for student i and the denominator is the root mean square error of the model.

This new standardized variable has a mean of zero. Standardizing the median test score solves the problem of comparability between different tests and across waves as the test scores have now the same metric across time. However, if the variance of the test scores changes over time, then standardizing the overall score at each country-wave pair actually makes the transformation biased. That is, by standardizing test scores the variability is forced to be zero across all waves. But if the true deviations of the median academic achievement changes over time, then the estimated trend in the SES gaps will

[7] Since each plausible value is a random draw from a theoretical latent normal distribution of possible student achievement scores, the mean should be precise in getting a central measure of the latent distribution.

[8] This does not mess up the analysis by masking age-specific gaps as all students in the sample are 15 year olds. Controlling for age is simply to adjust for monthly differences in ages.

[9] I also tried to run the model for each country-wave separately but the results were very similar and it was more computationally expensive

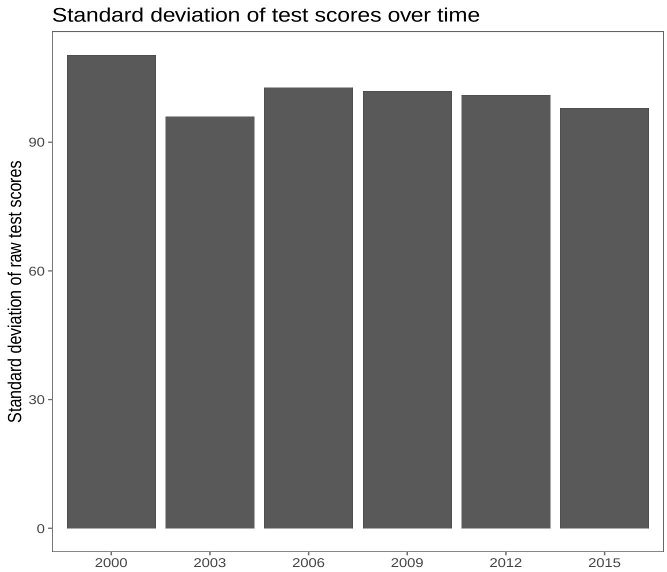

Figure 1.1: Standard deviation of test scores across all waves

be underestimated, or viceversa.

I plot the standard deviation of the mathematics test score for all waves in figure 1.1. The plot suggests that it is something I should not be deeply concerned with. The standard deviation of each wave seems to be following a similar pattern with a not so drastic exception of the year 2000.

Another concern is whether test scores measured at different waves have different amounts of measurement error. If that is the case, then the amount of bias will not be the same in each measure of the gap. This can be misleading and suggest erroneous interpretations regarding trends of the gaps over time (Reardon, 2011). PISA has tried to make sure the tests are comparable across waves but it is still necessary to adjust for this imprecision (OECD, 2012). Accordingly, each PISA survey provides a reliability indicator for each of the tests which can be used to adjust for the reliability of the scores.

In order to correct for this I calculate λ_i which is just $\hat{\gamma_i}$ adjusted by the reliability indicator of each wave. More formally, I calculate it through

$$\hat{\lambda_i} = \hat{\gamma_i} * \frac{1}{\sqrt{r}} \tag{1.3}$$

where r is the reliability score of the test score in that PISA wave [10]. Note that I

[10] Other procedures multiply each country by their own reliability measure for each year-subject pair (Chmielewski, 2016). The reliability estimates are calculated using Item Response Theory (IRT) analogues of traditional estimates of person separation reliability such as internal consistency. Unfortunately, PISA 2000 did not provide any reliability measure separately for each country and PISA 2015 has to yet release their own. At the moment of writing this chapter, they were unavailable. For these reasons, I implement

implement equation (1.3) separately by test scores and waves because there is a separate reliability indicator for each one. Once that is adjusted, the test scores should be roughly free of any bias in the trend that may arise from differential reliability of the tests.

In order to calculate the SES gaps it is necessary to estimate the thresholds for the 90th and 10th percentile. I calculate the thresholds using the SES index separately for each country-wave combination using the specific student sample weights of each one. I then generate a dummy of 1 for those above (including) the 90th percentile and 0 for those below (including) the 10th percentile for each country-wave pair. This means that all students that are below the 90th percentile and above the 10th percentile are excluded from the analysis.

I then fit

$$\lambda_i = \alpha_j + \beta_j * SES_i + \epsilon_i, \text{ for } i = 1, 2, ..., n \text{ for each country } j \tag{1.4}$$

where SES_i is whether the student is at or above the 90th percentile (coded as 1) or whether it is at or below the 10th percentile (coded as 0). I allow α and β to vary by country j in order to obtain gaps for each country. I implement this model separately for each wave and weight by the wave-specific student sample weights. The previous model allows to calculate the achievement gap for each country by extracting the β's and α's for each country. I also calculate the standard error of this difference and generate uncertainty intervals.

I fit a multilevel rather than a linear model because by allowing the SES dummy to vary by countries, the gaps which have very little statistical power *borrow* strength from the other country samples by pooling information together. This is important because including the SES *dummy* reduces the sample size considerably given that only students above or below the 90th and 10th percentile are included in the analysis.

Once the achievement gaps are calculated for every country in the six waves, the final dataset contains very few observations; around 183 to be exact. Considering this shortcoming, modeling the differences in the achievement gap between countries might lack enough statistical power to generate stable and unbiased estimations. Once we consider the risks of modeling under such low statistical power, the best approach is to use a fully bayesian hierarchical linear model with informative priors. The benefits of this methodology are twofold: first, it is more intuitive when assessing uncertainty intervals as they

the analysis following the original work of Reardon (2011) truly represent the probabilities of the estimate being contained in the uncertainty interval 95% of the time. Moreover, it allows to specify prior information based on theoretical and empirical knowledge to counteract measurement error and uncertainty in such low sample size settings.

The empirical literature on tracking has concentrated on a very narrow definition of tracking by focusing only on the age of selection (Hanushek et al., 2006). I use a more fine grained definition of tracking, which is possible through the work of Bol and Van de Werfhorst (2013). Aside from the age at first selection into tracks, I also use the number of tracks in the country, the percentage of the entire curriculum that is tracked and a vocational index [11].

In this model I regress the achievement gap of a country on a dummy variable where 1 equals only 1 track vs more than 1 track, whether the age of first selection is 15 or more versus below 15, a dummy stating whether the percentage of tracked curriculum is above 0 (that is, any curricular tracking), and the standardized vocational index [12]. This model allows the intercept to vary by country because there is not enough power to allow the coefficients to vary by country. I am reluctant to include more variables in the model, first to keep preserve parsimony, and secondly to prevent overfitting due to the sample size. To be clear, this model allows to explain differences in achievement between countries and not modeling the evolution over time directly.

The final model can be expressed as

$$y_i = \alpha_{ji} + X_i\beta + \epsilon_i, \text{for } i = 1,\, 2,\, ...,\, n \text{ years within each country } j \tag{1.5}$$

where $X_i\beta$ is the matrix of $\beta's$ and α_{ji} is the varying intercept. Moreover, the group-level error parameter can be defined as

$$\alpha_j = \mu_\alpha + \eta_j \tag{1.6}$$

where μ_α is the mean intercept for all countries and η_j is the country-level deviations.

The priors for all country intercepts, all of the $\beta's$ in the matrix X_i and the group-level variance parameter are assigned a t distribution with 3 degrees of freedom and scale parameter of 10, expressed as

[11] This index is a factor loading from a principal factor analysis of the percentage of students in upper secondary vocational education (taken from two sources, to reduce measurement error). I take the data from Bol and Van de Werfhorst (2013).

[12] I create dummies to avoid multicollinearity. The final model has a maximum VIF of 2.2.

$$\alpha_j \sim \text{Student t}(3, 10)$$
$$\beta \sim \text{Student t}(3, 10)$$
$$\eta_j \sim \text{Student t}(3, 10)$$

These priors are chosen because they are weak enough to allow variation under the present uncertainty and allow the estimated quantities to be driven by the strength of the data. However, they force the estimates under plausible values closer to zero rather than infinity.

A more summarized specification is to actually standardize all the tracking indicators into one index, something that Bol and Van de Werfhorst (2013) calculate for their data. Using this standardized tracking variable and the same bayesian multilevel model, I model the tracking and vocational indexes as before but add an interaction between the two, given that the second hypothesis looks to test whether these two features explain the differences between countries.

This model has the same priors for the varying intercept and the group-level error term but since these two indexes are a composition of the previous variables, I slightly adapt the prior to be normally distributed with varying means for the three $\beta's$ and standard deviation of 0.1, such that:

$$\text{Tracking index} \sim \text{Normal}(0.3, 0.1)$$
$$\text{Vocational index} \sim \text{Normal}(-0.2, 0.1)$$
$$\text{Tracking index * Vocational index} \sim \text{Normal}(-0.15, 0.1)$$

For the tracking index, past research has shown it to be positive (increase inequality) (Hanushek et al., 2006), so I provide a mean for the distribution of 0.3 with a standard deviation of 0.1. This allows for small positive effect and a zero effect about 0.01% of the time. The vocational index is less clear, so I provide a mean of -0.2. That is, vocational enrollment would decrease inequality as it helps children from the lower tracks achieve other types of jobs otherwise unavailable. For the interaction term I provide a mean of -0.15, an even smaller estimate to be conservative and let the data drive the interaction. This last distribution allows for positive values in the interaction term in about 10% of the time.

Finally, until now the methodology has outlined how we can explain the differences in the 90/10 achievement gap between countries where each observation inside the country records the achievement gap in the 15 year-span under study. This approach was adopted given that the main independent variable, curricular tracking, hardly changes over time making it very difficult to model the change over time directly. For this reason, the previous models explained differences between countries rather country-specific evolutions over time. However, there is another way of estimating the evolution of the achievement gap, which I develop next.

Instead of explaining the evolution over time directly, the third model calculates the cumulative change over time for each country, summarizing the 6 year-time span of achievement gaps into one cumulative inequality gap. That is, each country has one final gap which is calculated through the cumulative addition/subtraction of the 90/10 SES gap over the six time points. Given that the variable is continuous, I fit a linear model to test whether the independent variables are linearly related to the 'cumulative inequality gap'. This linear model has standard weakly informative priors as the ones used in section 1.4.2 given that this specific question has not been researched before.

This approach dramatically reduces the total sample size to 32, effectively the number of total countries available [13]. It is for this reason that I only include at most 2 covariates in the model, the tracking and vocational index, which summarizes the previous models and keeps it parsimonious. The results are presented for the 90/10, 80/20, and 70/30 achievement gaps to test for robustness. In the next section I present the results.

1.5 Results

1.5.1 Evolution of the achievement gap

Table 1.1 shows a description of the sample size and mean score of the top and bottom SES groups for only the first and last time point. One main concern from the planned analysis is that using the top 90th percentile and bottom 10th percentile would result in a small sample size. This table suggests that the data has a reasonable number of respondents to actually estimate gaps accurately. Moreover, we can see that in all instances the bottom SES group has a lower score than all top SES groups.

We can see that in some countries such as Finland and Sweden the average low SES score is actually above the average score of 0, suggesting they are equal countries. However,

[13] I have been presenting a selected number of countries given the lack of space, but for every model/estimation I include all countries available.

		Low SES			High SES		
Year	Countries	N	Avg score	S.E	N	Avg score	S.E
2000	Australia	294	0.22	0.07	270	1.22	0.09
2000	Austria	294	-0.04	0.11	235	1.14	0.15
2000	Canada	1970	0.30	0.05	1357	0.90	0.07
2000	Denmark	230	-0.01	0.13	228	1.16	0.17
2000	Finland	261	0.42	0.12	267	0.89	0.15
2000	France	253	0.03	0.04	258	1.20	0.05
2000	Germany	250	-0.64	0.04	303	1.80	0.05
2000	Hungary	291	-0.46	0.09	254	1.65	0.12
2000	Italy	278	-0.31	0.05	278	0.80	0.06
2000	Netherlands	129	0.46	0.08	158	1.24	0.11
2000	Poland	209	-0.31	0.04	166	1.25	0.06
2000	Spain	318	-0.28	0.05	340	1.09	0.07
2000	Sweden	241	0.12	0.10	246	1.06	0.13
2000	United Kingdom	531	0.15	0.04	415	1.32	0.06
2000	United States	222	-0.43	0.02	164	1.74	0.03
2015	Australia	1694	-0.15	0.04	1235	1.28	0.05
2015	Austria	676	-0.30	0.07	705	1.53	0.10
2015	Canada	2215	0.20	0.03	1863	0.99	0.05
2015	Denmark	1013	0.09	0.08	595	1.08	0.10
2015	Finland	575	0.07	0.08	584	1.16	0.11
2015	France	570	-0.37	0.02	615	1.66	0.03
2015	Germany	545	-0.09	0.02	582	1.45	0.03
2015	Hungary	466	-0.72	0.07	589	1.79	0.09
2015	Italy	947	-0.20	0.03	1043	1.23	0.04
2015	Netherlands	521	0.13	0.04	525	1.24	0.06
2015	Poland	446	0.10	0.03	448	1.24	0.05
2015	Spain	608	-0.25	0.03	703	1.24	0.04
2015	Sweden	527	-0.20	0.06	542	1.32	0.09
2015	United Kingdom	1387	-0.03	0.03	1388	1.15	0.04
2015	United States	585	-0.39	0.01	544	1.22	0.01

Table 1.1: SES sample size and ISCED composition

Countries	# of tracks	Age of selection	% of curric tracked	Std. Voc	Std. tracking
Australia	1	16	0.15	0.97	-1.04
Austria	4	10	0.67	1.7	1.82
Canada	1	16	0	-1.72	-1.32
Denmark	1	16	0.25	0.45	-0.87
Finland	1	16	0.25	0.74	-0.87
France	2	15	0.25	0.39	-0.47
Germany	4	10	0.69	0.89	1.86
Hungary	3	11	0.67	-0.7	1.42
Italy	3	14	0.38	0.95	0.17
Netherlands	4	12	0.45	1.26	0.94
Poland	3	15	0.38	0.3	-0.08
Spain	1	16	0.17	0	-1.02
Sweden	1	16	0.25	0.69	-0.87
United Kingdom	1	16	0.15	0.47	-1.04
United States	1	16	0	-1.84	-1.32

Table 1.2: Curricular tracking statistics for selected countries

countries such as Spain, Poland, Hungary and Germany have, on average for the low SES group, cognitive scores that are much lower than their countries average. I also see some countries with major changes from year 2000 to 2015, with, for example, Australia decreasing the low SES gap from 0.22 SD to -0.15, well below the country average. In order to make sure the composition of the top SES group and bottom SES group is as expected, I present table 1.1 in the appendix. This table confirms that both groups are extremes in terms of education, with the top SES having highly educated parents while the bottom group has lower educated groups. More concretely, 96% of the top SES group have bachelor's degrees or above while *none* of the respondents in the bottom SES group have parents with bachelor's degrees or above. In addition to this, these groups not only differ in education but on other dimensions such as immigration and human capital measured as books in the household. These results confirm that both groups measure what they are supposed to: both extremes of the SES gradient. This table is presented for all countries pooled for years 2000 and 2010 to confirm that the results from the modeling section are not due to changes in the composition of the groups. The results indeed show that the composition is roughly the same between the two time points.

In table 1.2 I present the data related to the tracking setup of a few selected countries. This data (all countries not presented here, only selected ones) is the one used in the modeling section and it is taken from Bol and Van de Werfhorst (2013).

It is evident that there is ample variation in the tracking structure of countries. Germany and Austria have highly tracked curriculums, early age of selection and high number of tracks, whereas countries such as the U.S. and Canada have the opposite. The last two columns of the table show a standardized index created by Bol and Van de Werfhorst (2013). The first one measures the degree of vocational specificity of a country and the last one the degree of curricular tracking. Both have a mean of zero and positive values mean higher tracking and vocational enrollment while negative values mean the opposite. Looking at the results more closely we see similar patterns: European countries have higher vocational specificity and curricular tracking than Anglo-Saxon countries. Two curious cases are Hungary and the Netherlands. The first one shows low vocational, while the other has high levels of vocational enrollment. Yet Hungary also has high levels of curricular tracking. In summary, table 1.2 shows by and large that countries have developed different tracking structures and roughly speaking there is a large divide between European and Anglo-Saxon countries, something missing in Bradbury et al. (2015).

Next I look at the evolution of the achievement gap. I start by looking at the achievement for a selected number of developed countries in figure 1.2 as I cannot plot all PISA participating countries. I plot only mathematics for each country (dark dots) but also a quadratic trend spline for *both mathematics and literacy* pooled. The pooled trend spline is calculated for both tests in order to test whether the results hold under stricter specifications. By comparing the dots (only Mathematics) and the trend spline (Mathematics and Literacy) we can gauge how far reading is from the Mathematics result. The trend spline is calculated by averaging the medians of the two scores at each time point.

As we can see from the results, some countries have increased their achievement strongly. For example, France, Austria and surprisingly Sweden have very steep slopes. France experienced an increase in inequality by roughly 0.9 SD, Austria by 0.6 and Sweden by 0.6. This pattern happens similarly for literacy. For example, France has an increase of 0.6. For such a short period of time, the magnitude of these increases are reasonably big.

Given that no one has estimated the evolution of the gap I cannot cross-check how other empirical estimations put France at. However, the work of Micklewright and Schnepf (2007) is the closest reference available which also finds that France was a low dispersion country in 2000; there is no evidence on what happened over time. Fortunately, the work of Bernardi and Ballarino (2016) did study social origin inequalities (broadly speaking, not in terms of achievement gaps) in France and found that they increased since the 2000's.

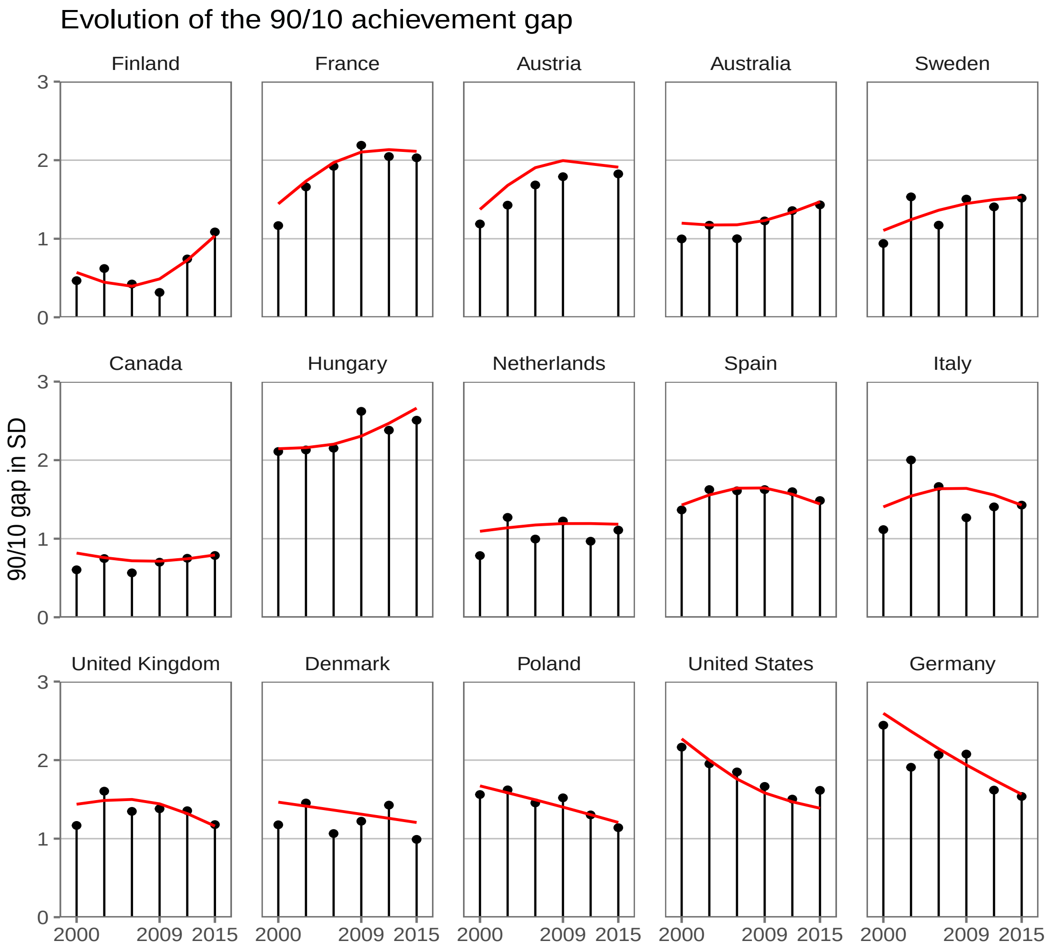

Figure 1.2: 90/10 achievement gaps in mathematics expressed as standard deviations for selected countries between 2000 and 2015

Other countries have reasonable increases such as Finland and Hungary, with increases of nearly 0.6 and 0.4 standard deviations respectively. Aside from these countries, there are other countries which experience no changes at all, specifically, Canada, Netherlands and Spain. Canada excels here not only because the gap has been stable over time, but because it has the smallest gap of all countries presented here. It is nearly 0.5 SD in 2000 and it increased only by 0.2 in 2015.

On the other hand, there are other countries which experience a decrease in the SES achievement gap. Poland decreased by about -0.4 and Denmark by -0.2. However, the most notable cases are the U.S. and Germany. These two countries show high levels of dispersion in the year 2000 with SES gaps of over 2 SD. But in the 15-year time trend both countries reduced the gaps by -0.6 and -0.9 respectively. Their distinctively large gaps in 2000 also show up in the work of Micklewright and Schnepf (2007). This corroborates the findings of Reardon and Portilla (2015), which found a decreasing gap for kindergartners. The decline is evident also for 15 year olds, suggesting it might be more of an institutional change rather than a specific grade-level policy. Do note that the trend line is for mathematics and literacy pooled. If it were only for mathematics, then the trend line would be even more pronounced. Despite this, pooling the two subjects gives more strength to the analysis by showing that both tests are in the same direction. I also plot the 80/20 and 70/30 SES gaps and find similar patterns (see figure 1.4 in the appendix).

Analyzing figure 1.2 the reader may get the impression that these trends are not very steep and they should not be relevant in practical terms. However, note that the Y axis is measured in standard deviations. Small changes are actually large in practical terms. Take the case of Sweden. The slope does not look that steep but in reality it increased the gap from 1 SD in 2000 to around 1.5 SD in 2015. With that information in mind, the trends of Poland, United States, France and Germany are concerning.

I find that the initial gap for the U.S. in 2000 is 2.16 standard deviations of magnitude varying between 2.13 and 2.2. This gap is much higher than found in previous studies. There is one difference and one limitation relative to previous research. First, past research has never really concentrated on the 15-year old achievement gap but rather on student achievement gaps while adjusting for age which forces the age effect to disappear. Secondly, the literature cited until now has neglected to present uncertainty estimates for their achievement gap estimates. These two reasons, either working together or in isolation could explain the difference in the magnitude of the U.S. gap. However, the methodological approach used here is very similar to Reardon (2011), with the main caveat that the author adjusts for age.

Figure 1.3 takes a more direct approach and looks at the percentage change from the first and last time point available. Each data point has been computed together with its 50% and 95% uncertainty interval [14].

Generally speaking, most countries increased their achievement gap over time. France had an average increase of about 80% since 2000 varying down to 40%, whereas Germany had a similar figure but decreasing. Many of the countries that did not show a positive slope, such as Hungary or Australia, had in fact increases of about 40% in their gap. In contrast, the U.S. and Poland had also significant decreases of about 40%. The benefit of

[14] Each of these uncertainty intervals were computed using a 500-replicate bootstrap

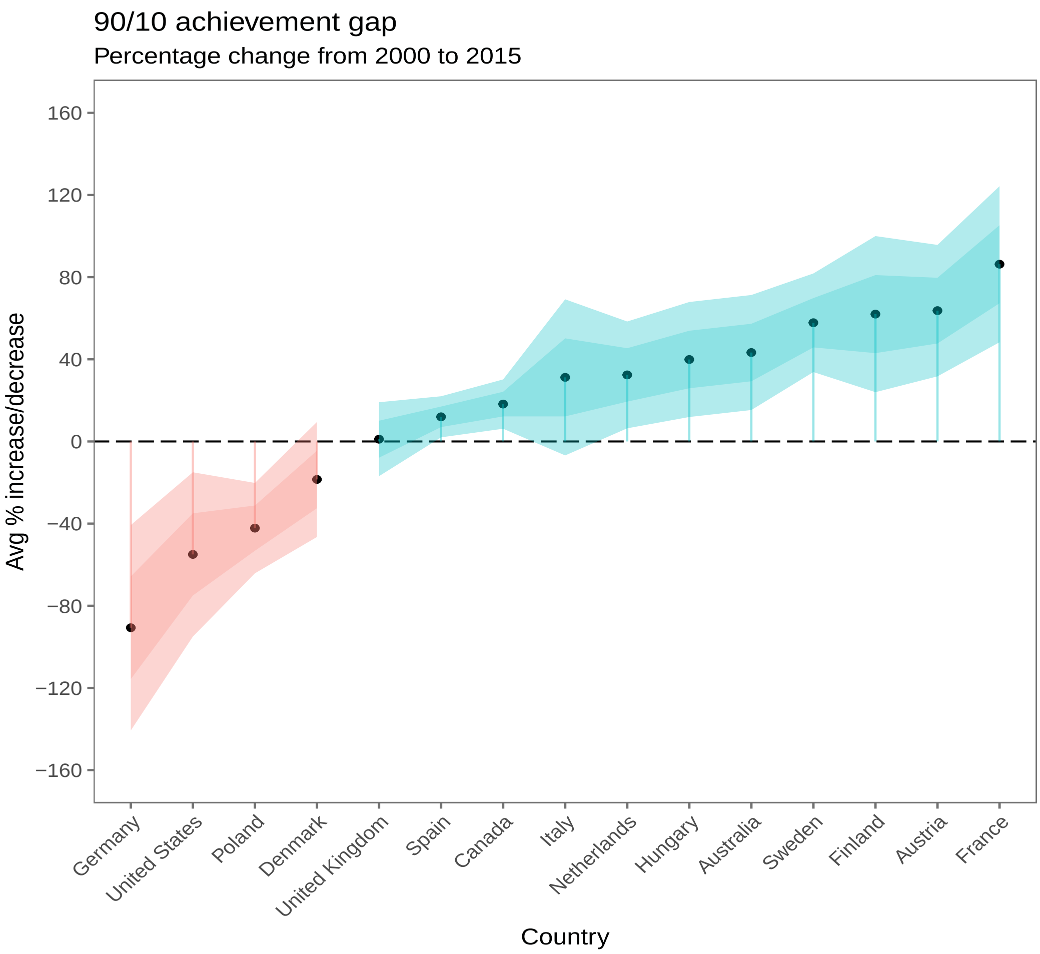

Figure 1.3: Percentage change in mathematics between years 2015 and 2000 for selected countries. Red regions represent decreases while blue regions represent increases in the achievement gap

presenting these estimates this way is that the reader can actually assess the uncertainty of each calculation, and there is evidence that they have wide variability. Despite this, most countries show a clear sign of either decreasing or increasing. I plot the same graph for the 80/20 (figure 1.5) and 70/30 (figure 1.6) SES gaps in the appendix and the results hold for these achievement gaps as well.

Something reassuring is that mathematics and reading (figure 1.1 and figure 1.2 in the appendix) follow basically the same trend across the presented countries. This means that the result is not an artifact of chance alone.

The first question in the research/hypothesis section asked how does the achievement gap behave in terms of its evolution over time and between countries. The results show that the gap is very different between countries. There is ample variability across countries with some countries that have different educational systems experiencing similar achievement gaps. Countries as disparate as the U.S. and Germany follow similar patterns while other such as France and Austria experience steep increases. Equally important, there are strong and visible changes in the evolution of the gap. This highlights the fact that achievement gaps are very variable and very context-specific to their educational systems. Moreover, the results of Reardon (2011) are different from the ones presented here. It could be for two reasons. First, these results show the achievement gap from the year 2000 onwards which is the year where Reardon (2011) had no more data points. Moreover, Reardon (2011) has surveys for many different age groups and adjusts for age, eliminating the age-specific effects. These results are the first to document such starking contrasts for the achievement gap of 15 year-olds.

However, it is important to disentangle where this gap is originating from. Is this because the top are improving while the bottom decreases? Or is it that the bottom is catching up? These results are important because they help pinpoint whether countries experiencing similar changes are indeed originating from the same source. Next, I plot the same graph but show the divergent patterns between high/low SES origins.

1.5.2 Source of the achievement gap

Figure 1.4 shows the evolution for low and high SES groups separately. The middle line is a linear trend showing whether one of the groups is growing/narrowing faster than the other. For example, in the U.S. the top seems to be equalizing much stronger than the bottom is catching up. The UK seems to be following the same path as the U.S. as well.

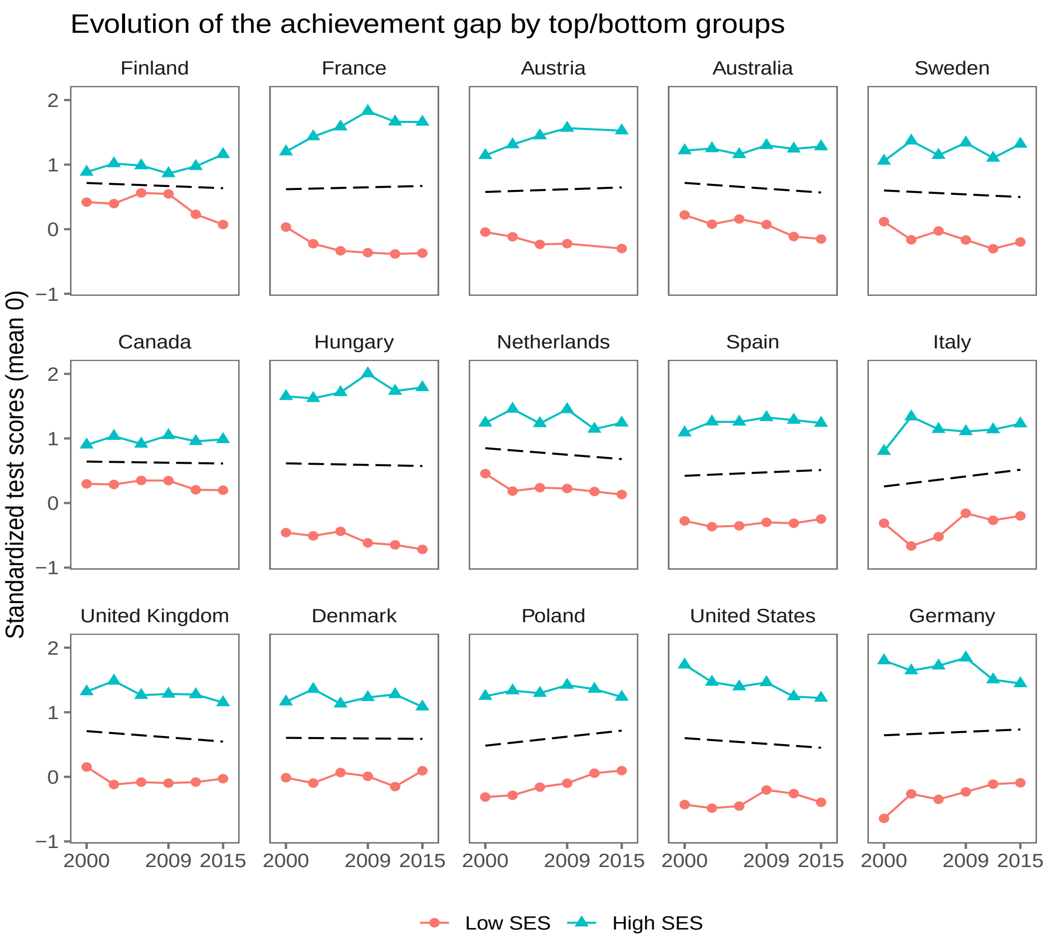

Figure 1.4: Evolution of the achievement gap for separate SES groups

On the other hand, in Poland the gap seems to be closing as the bottom SES group is catching up much faster. The Netherlands shows a similar pattern as the U.S. and the UK patterns but upon closer inspection the explanation is different. The slope is negative (like in the previous two countries) but that is because the low SES group is decreasing at a greater rate than the high SES is decreasing.

Moving on to the flat summary lines, these should be interpreted with caution as it does not mean that the gap is not changing. Denmark, Belgium, Netherlands, Norway and Spain show a flat line as the gap is growing little, if at all. On the other hand, Sweden, France, Finland, New Zealand, Austria, among other countries, show a flat line as groups are distancing themselves at a similar rate. These results highlight the importance of not only summarizing average achievement gaps: the source of the gap varies greatly between countries and it is easy to see how each of these patterns contributes to the overall inequality of a country.

After analyzing the trend of each country and where it is coming from, interesting results start to emerge. The U.S. is closing the achievement gap at a rapid rate, but it is because the top SES group is going down faster than the bottom is catching up. In contrast, countries such as Germany and Poland show a closing gap as well but it is because the bottom group is catching up faster than the high SES group is coming down. In fact, the Polish case is the most egalitarian of all: the bottom group is catching up at a rapid pace and the top group is maintaining their high levels of performance, fostering a truly equalizing effect.

In countries where the gap is widening, both groups are distancing themselves at a similar pace. Finland, France and Austria, show this exact pattern. The results seem to show a solid pattern: when the gap widens it is because both SES groups are distancing from each other. But when the gap is being reduced the source can vary between both groups narrowing or just one of the two leading the decrease.

In a more convenient way, we can inspect whether there is a lot of variance in the rate at which top/bottom groups are changing in figure 1.5.

Figure 1.5 shows that there is a significant number of countries where high SES is increasing inequality faster than low SES is catching up (top-left panel) but there is also a fair share where low SES is catching up faster, such as the U.S. and Germany (bottom-right panel). The plot reveals an overall negative relationship where countries in which the low SES group is catching up, the high SES group goes down. This is interesting as

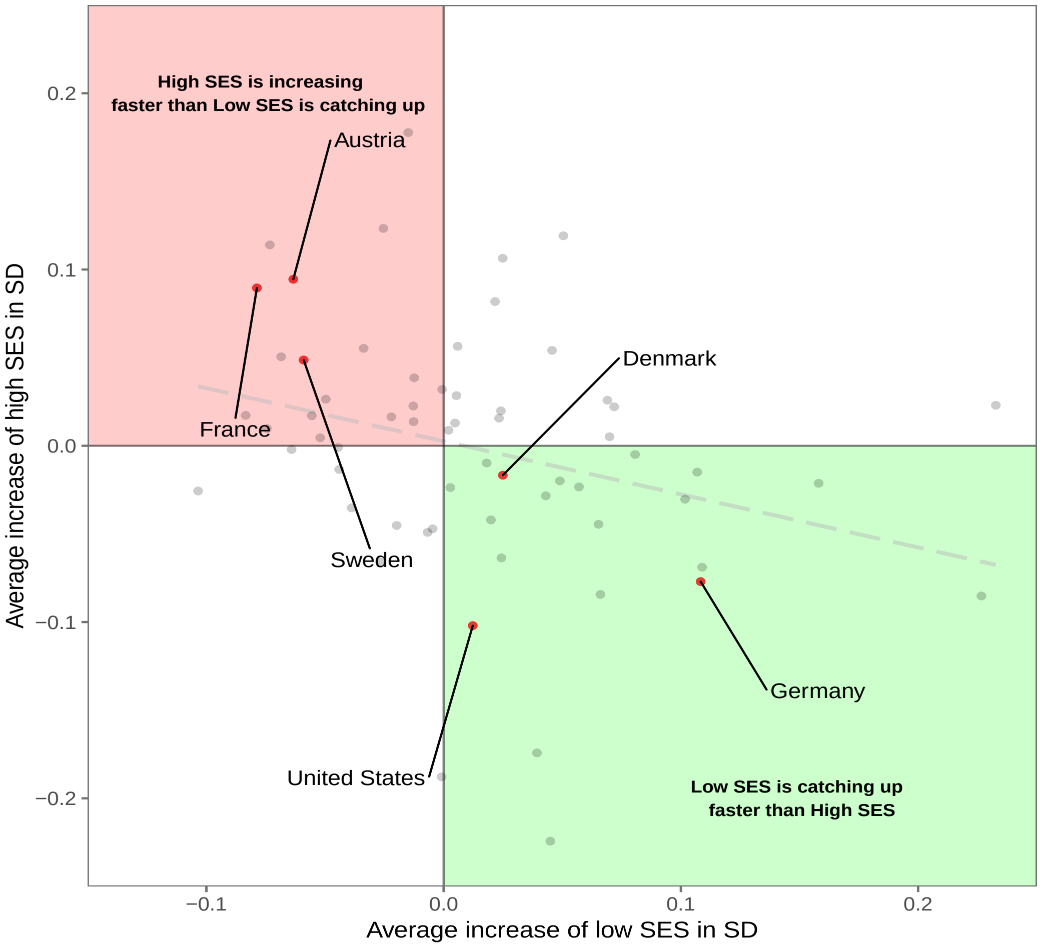

Figure 1.5: Rate at which top/bottom SES groups are changing for the mathematics 90/10 SES gap

it suggests that there is a trend for countries to have the bottom SES groups climbing upwards, a trait of meritocracy. However, this means that the top group comes down, rather than stagnate or increase.

It is also important to note that countries clustered in the top-left panel share very distinct tracking profiles. For example, Sweden has very little tracking compared to France or Austria, which is a high profile tracker with over 3 tracks and separating students as early as 10 years of age. Similarly, countries in the bottom right panel have also very different tracking setups, for example, Germany and the U.S.. However, for the countries shown on the back there are several countries that share a similar degree of curricular tracking. All in all, there seems to be some relationship between the evolution of the achievement gap and tracking but the evidence is mixed.

The plot also shows very few countries where low SES are decreasing and high SES is also decreasing (bottom left panel). This can be interpreted as either one of the two SES groups increases while the other decreases. Interestingly enough, there seems to be also very few countries where both SES groups are increasing (top right panel), suggesting that on very few cases do SES groups increase their performance on par. I plot all remaining countries in the back, together with a trend line and the results show that the relationship holds for all PISA countries.

The second research question was interested in finding out whether the achievement gap was changing because the top was gaining ground, the top was falling behind or because of a dynamic interaction between the two. The hypothesis stated that in countries where there is high degree of tracking we should expect the gap to be increasing because both groups are distancing from each other. Conversely, in countries with little tracking we should expect for the gap to be narrowing. The results presented here shows mixed evidence in favor of the hypothesis. There does not seem to be a clear cut pattern between the tracking profiles of each country and the evolution of their achievement gaps. In other words, the results shown here confirm that the curricular tracking setup of a country does not imminently mean a certain degree of equality/inequality, although it is related.

More concretely, the evidence suggests that countries where the top SES group is increasing and the bottom is decreasing (top-left panel) have different curricular tracking setups. Austria is an early tracker with several tracks, while Sweden is a late tracker with only 1 track. On the opposite panel (bottom-right), we can see countries where the bottom group is catching up while the top is coming down. The results are also similar as we see countries such as Denmark and Germany which are at opposite poles in terms of tracking, while the U.S. is distinctive in that it has little formal curricular stratification. The results do not show definitive evidence of top-to-bottom equality in all of these countries (except for Poland). However, an interesting finding clearly arises from these high-level plots: regardless of the different tracking setups, there is seems to be a trade off between SES groups, either the bottom comes up and the top comes down or the top goes even further up and the bottom further down.

To the best of my knowledge this is the first time this evidence has come up in the literature. The reason why this pattern seems to be present in over 30 countries is unknown and should be followed in future research.

1.5.3 The role of curricular tracking in explaining the achievement gap

We now move to the modeling section of the chapter, where I test a possible explanation to differences in the achievement gap between countries. The reason why countries differ in their evolution of inequality is still unknown. There is ample evidence showing the inequality between countries, to a certain extent, can be explained by the degree of tracking (Hanushek et al., 2006). But I am also interested in what explains the between country evolution of inequality, that is, the explanation as to why in certain countries it is increasing more than others over time.

Using the first bayesian model explained in the methodology section, I present the results of the first model in table 1.3.

	Model 1	Model 2	Model 3	Model 4	Model 5
Only 1 track		-0.54 (-0.84, -0.24)	-0.33 (-0.64, -0.03)	-0.32 (-0.65, 0.03)	-0.21 (-0.54, 0.13)
Age selection >= 15			-0.39 (-0.69, -0.07)	-0.4 (-0.71, -0.1)	-0.59 (-0.92, -0.26)
% of curric tracked				-0.03 (-0.57, 0.55)	0.57 (-0.13, 1.27)
Vocational Index					-0.26 (-0.46, -0.06)
Between-group variance:	0.26	0.2	0.18	0.18	0.16
Sample size:	183	183	183	183	183
Number of groups:	32	32	32	32	32
Intercept	1.58 (1.44, 1.73)	1.76 (1.6, 1.92)	1.93 (1.7, 2.15)	1.96 (1.36, 2.54)	1.58 (0.93, 2.24)

Table 1.3: Explaining 90/10 achievement gap - Multilevel model with intercept varying by country

I present uncertainty intervals next to each coefficient. The first model includes an empty model to record the initial variance between countries, which is 0.26. The second model only includes the number of tracks in the country. The variable shows that countries that have one track have on average 0.55 SD of less inequality than other countries. This covariate alone explains about 23% of the variance between countries of the 90/10 achievement gap as the variance decreased from 0.26 to 0.20. The third model includes the dummy for whether the country has an age selection of 15 or more and the results show that countries with late selection into curricular tracking have about -.39 less SD in inequality. Also note that the coefficient for the number of tracks was reduced, suggesting these two variables are explaining similar things. This is expected given that late tracking is usually associated with fewer tracks, such as in the Scandinavian countries. The fourth model now includes whether the country has any degree of curricular tracking, meaning whether the country has above 0% of the curriculum tracked. The coding of the variable comes from visual inspection of the distribution, where some countries, such as the U.S. and Canada, have untracked curriculums. For this last variable there does not seem to be a relationship, but after including the vocational index in the fifth model, the coefficient is completely reversed with countries with tracked curriculums having about 0.59 more SD in inequality than countries with no curricular tracking.

Also note that vocational enrollment is associated with less inequality. Not surprisingly, the age of selection coefficient also increased due to the correlation between age of selection and vocational enrollment. Countries with high vocational enrollment are also those with early age selection, such as Germany and Austria. These two variables should be inspected further, perhaps with an interaction, given that the coefficients are sizable and correlated. This final model explains about 40% of the variance between countries in the 90/10 achievement gap as the total variance was reduced from 0.26 to 0.16. This is not a trivial percentage and suggests that these institutional factors are playing an important role in defining country differences in the achievement gap.

It is also important to highlight that as the model increases complexity (addition of variables), the estimated coefficients become more uncertain. It is true that all the coefficients point in the expected direction, even with a weakly informative prior, which suggests that the data has reasonable strength in influencing the estimation. However, the uncertainty intervals suggest that they vary considerably. For example, the coefficient for the percentage of curriculum that is tracked varies all the way from -0.13 to 1.28. This suggests that the coefficient is most likely to be positive as most of the region lies inside the positive spectrum yet there are possibilities that it varies down to -0.13. In contrast, the coefficient for the age of selection varies between -0.92 and -0.26. This coefficient shows a wide interval but only concentrated on the negative region, suggesting that it is highly probable that it is negative and strong.

It is important to highlight that we should not dismiss this finding as irrelevant simply because any of the estimates includes zero, as the uncertainty intervals need to be interpreted as a measure of how much we can trust our results. The explanatory measures presented here such as the variance between countries should be also interpreted with caution. The country-level variables presented here are only one set of possible explanations. Other explanatory variables might, and should, be correlated with the curricular tracking variables, making these estimates endogenous to other explanations.

All of these countries have a different curricular tracking structure, with some setups being more egalitarian than others. Using the previous model I make a simulation and predict the level of the 90/10 achievement gap for all possible curricular tracking setups. In figure 1.6 I plot all these combinations. The X axis shows the number of tracks that a country has, the Y axis has the predicted 90/10 achievement gap and below each country name it is specified whether the country has an age selection of 15 or more or below 15. In addition, to confirm how accurate the model is, I also plot the actual level of inequality in the actual curricular tracking setup of the country (the red star). Let us take Germany for example.

The star point for Germany is in the '>1 track' and '<15 age' setup, meaning that Germany has a tracking system that has more than 1 track and an age of selection of below 15. We can see that the model predicts the level of inequality for that set up accurately (height of the bar), relative to the actual level (the star point). If the German tracking setup would be in the ideal '1 track' and age of selection above or equal to 15, then inequality would be lower.

First, this plot highlights that the model seems to fit the data reasonably well as the estimated prediction stars line up very close to the actual achievement gap of the country. All in all, the model does a reasonable job at capturing the achievement gap of each country.

On the substantive side, this graph reveals some interesting patterns. Generally speaking, we can see that the simulation predicts that virtually all countries which are not in this 'ideal' tracking, that would switch to the 'ideal' setup, would experience a reduction of the achievement gap. The opposite is also true. Countries which are in the 'ideal' setup

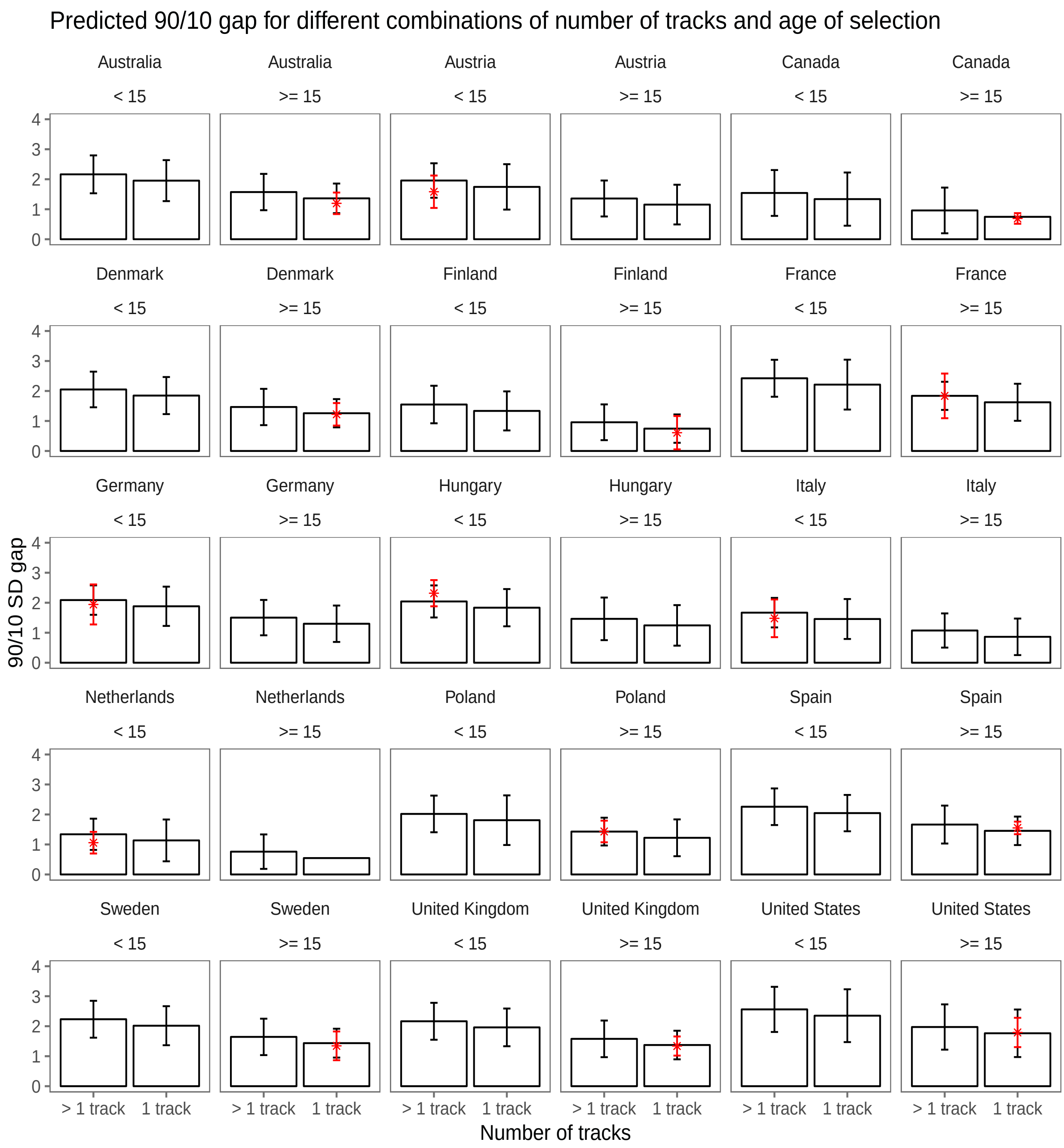

Figure 1.6: Simulation: predicted 90/10 achievement gap for different tracking setups

that would change to the 'worst' setup would see a widening of the achievement gap. On average, I find that if all countries switched to the 'ideal' setup, there would be an average reduction of 11% of the achievement gap with some countries experiencing even a 30% reduction. This is an important estimate considering that Reardon (2011) found that the US gap increased by about 40-50% in 30 years. Had all countries in the 'ideal' tracking switched to the 'worst' curricular tracking, the model predicts an average widening of nearly 51%. For those countries which are not in the 'ideal' tracking setup, I compute the percentage change from switching to the ideal curricular tracking in the appendix in table 1.2. The average reduction in the achievement gap only for these countries is an even stronger 32%.

Table 1.4 shows the results for the second model outlined in the methodology section. This table is different from the previous as it summarizes the tracking index and adds an interaction between the tracking index and the vocational index.

	Model 1	Model 2	Model 3	Model 4
Tracking Index		0.27 (0.17, 0.37)	0.32 (0.21, 0.43)	0.36 (0.25, 0.48)
Vocational Index			-0.19 (-0.3, -0.07)	-0.19 (-0.31, -0.08)
Tracking * Vocational Index				-0.12 (-0.22, -0.02)
Between-group variance:	0.26	0.2	0.18	0.17
Sample size:	183	183	183	183
Number of groups:	32	32	32	32
Intercept	1.58 (1.42, 1.73)	1.59 (1.46, 1.72)	1.66 (1.53, 1.78)	1.7 (1.57, 1.84)

Table 1.4: Explaining the 90/10 achievement gap - Tracking and Vocational interaction with intercept varying by country

The results from the previous models are confirmed, where more curricular tracking is associated with an achievement gap of about 0.36 SD wider and more vocational enrollment is associated with a reduction of about 0.19 SD, half the slope of curricular tracking. Could it be that the curricular tracking influence might be offsetting the effect of the vocational index given that it is half the size of the slope? Both these coefficients are actually significantly different from each other, so it is possible. Before we explore it visually, note that the interaction between the two is not either big or small in terms of its uncertainty. However, it shows to be mostly negative [15].

[15] Given that the interaction term might be overly driven by the prior distribution, I ran the *same* model but assigning a uniform prior for the interaction and the coefficient is 0.10 (-0.25, 0.04). As it is evident, the prior distribution just helps to nudge the coefficient slightly because of the low sample size but the results are very much alike.

Figure 1.7 plots the interaction for different quantiles of the vocational index.

Figure 1.7: Interaction between tracking and vocational values. Legend ordered relative to the lines

The results are interesting. We can see that for lower levels of curricular tracking (left hand side of the X axis) there is a low achievement gap, regardless of whether the country has high or low vocational enrollment (all lines overlap). However, once curricular tracking enters the picture, vocational enrollment can be a strong equalizer (right hand side of the X axis). We see that the bottom quantile of vocational enrollment (top line) has almost twice the achievement gap as the top quantile for vocational enrollment (bottom line). Moreover, the earlier interpretation (that curricular tracking is more important than vocational enrollment for reducing the achievement gap) is evident here. Despite the equalizing power of vocational enrollment, a high level of curricular tracking with the highest level of vocational enrollment still leaves a country with an achievement gap of over 1.5 SD. However, a country with low levels of curricular tracking and *any* level of vocational enrollment is nearly 1.3 SD at best.

This interaction presents high levels of uncertainty considering that the predicted values have wide intervals. This is expected as estimating an interaction term requires more than twice the number of observations to accurately capture a simple main effect (Gelman, 2015). However, even with such a small sample, the interaction shows a clear and evident relationship.

To test whether these results hold for other samples, I compute the same models as before but also for the 80/20 and 70/30 SES gap. Figure 1.3 plots these coefficients for the three SES gaps in the appendix. All models present virtually the same size in the coefficients which shows that the estimates are robust as they replicate under other gaps.

The third research question looked to understand why some countries differed in their achievement gaps and test whether different indicators of tracking and vocational enrollment seemed to explain this cross-country variability. More concretely, the hypothesis suggested that tracking would explain why some countries showed big differences in achievement gaps while others showed small achievement gaps. The results from table 1.3 and table 1.4 seem to be in line with these claims. Tracking alone seems to explain about 25% of the between country variability while including vocational enrollment raises it nearly 40%. Moreover, Figure 1.7 shows a surprising result: curricular tracking seems to be a possible explanation of the 90/10 achievement gap, however it should not be studied in isolation from other features such as the degree of vocational enrollment.

The exact mechanisms through which the tracking setup is explaining these inequalities is still speculative. It is possible that curricular tracking, although not changing, is currently exacerbating inequality as time passes by. Another explanation, which this chapter finds evidence for, is that the curricular tracking setup interacts closely with the vocational setup of a country. This means that tracking alone seems to increase the achievement gap possibly as students in the higher tracks benefit much more than the ones in the bottom groups. However, this seems to be slightly equalized as students enroll in vocational tracks. These are possibly children coming from the bottom tracking groups which find vocational tracks that allow them to excel in specific areas and find steady jobs. These explanations are speculative and further research should carry the task of looking for a more concrete explanation and attempting to replicate these findings.

Finally, the last model described in the methodology section looked to model the evolution over time and test tracking and vocational enrollment as possible explanations.

Table 1.5 presents the results.

	90th/10th	80th/20th	70th/30th
Track Index	1.64 (0.85, 2.45)	1.11 (0.46, 1.77)	1 (0.36, 1.63)
Vocational Index	-1.03 (-1.95, -0.14)	-0.9 (-1.63, -0.16)	-0.87 (-1.63, -0.16)
R-squared:	30%	25%	23%
Sample size:	32	32	32
Intercept	9.37 (8.63, 10.11)	4.59 (3.96, 5.21)	3.88 (3.26, 4.48)

Table 1.5: Linear model regressed on the cumulative achievement gap - Models for three different achievement gaps

We see that a 1 SD increase in the tracking index widens the 90/10 SES gap by 1.66 SD over time, and the vocational index reduces it by 1 SD. These two covariates explain nearly one third of the cumulative gap. As the achievement gap decreases (80/20, 70/30), these coefficients become smaller as expected because the gaps are smaller, but keep their strength and their predictive power. Comparing this model to the previous models, the evidence suggests a similar outlook. That is, tracking and vocational enrollment seem to explain differences in achievement gaps between countries and the preliminary results suggest that it also explains the cumulative yearly gaps indirectly. Moreover, these results are more reassuring of the overall results of the chapter given that this new specification is tackling the evolution of the achievement gap rather than the differences between countries. These results seem to align with the previous findings and make the claim more credible.

Finally, the fourth question and third hypothesis looked to understand not only whether tracking and vocational enrollment explained between-country differences but whether tracking can be also relevant as explaining the evolution of the achievement gap. Although tracking does not change much over time, there are traits that depend on it that do change over time. For example, the quality of instruction for the different tracks might change. This might be more pronounced in countries where there are several tracks and they are very differentiated.

The results from table 1.5 confirm this hypothesis. The results show that tracking seems to explain the cumulative evolution of the achievement gap for the 90/10 gap in Mathematics. The hard numbers point out that tracking and vocational enrollment explain about 30% of the variability on the evolution of the achievement gap. Moreover, the results are replicated for the 80/20 and 70/30 gaps with the results lining up as expected given that as the gaps become broader, the explanatory power decreases. We should expect this because as the gap grows the differences between SES groups become smaller.

The results presented in this chapter seem to all converge towards one clear direction: the 90/10 student achievement gaps seem to be very different between countries and institutional features such as curricular tracking and vocational enrollment seem to be plausible explanations for these gaps. Some countries have gaps which are of over 2 standard deviations while others, like Canada, have gaps of 75% less magnitude. However, when all countries are pooled into a common model, tracking and its various features together with vocational enrollment seem to explain about 40% of the variation in the achievement gaps between countries. These results are not perfectly clear because there is a great deal of uncertainty in the estimation procedure. However, the same estimates replicate almost perfectly with smaller achievement gaps such as the 80/20 and 70/30 gaps.

Finally, as an indirect measure, the results also show that the tracking structure and vocational index seem to indirectly explain about 30% of the cumulative achievement gap in 15 years for the 32 countries under study. This test is not a smoking gun because it is very difficult to calculate a cumulative achievement gap if gaps increase/decrease every year with a constant positive or negative slope. However, for countries which constantly increase or decrease, then this is a valid model. As the evidence in the visualization of the achievement gaps suggests, there are a handful of countries, if not the majority, which show either increasing or decreasing patterns over time. All in all, the take away from the combination of all the previous results is that despite the uncertainty in all previous models, they all point in the same direction.

1.6 Limitations

The results presented here seem to converge to similar results and prove to be robust throughout different specifications. Despite this, there are several important limitations in the design of the study.

The main limitation of the study is the fact that once the achievement gap is summarized for each country-year combination, the resulting sample size is small and contains an accumulated degree of measurement error as calculating the achievement gap carries forward a number of errors in measurement. I adopt one of the most widely chosen strategies:

bayesian estimation. This allows to counteract the low statistical power and measurement error by providing likely prior distributions. The prior distributions used here were chosen among experimentation, theoretical considerations and previous research. But of course, prior distributions carry some degree of subjectivity. This is a limitation that is important. Further research should attempt to study the findings presented here under a different scenario or under different piors (already tried in the exploratory attempts before the chapter was written down) and present the results.

Another limitation from the last model is that when calculating the cumulative achievement gap some countries might see oscillating increases/decreases year after year, effectively making their cumulative achievement gap meaningless because it does not represent a trend. Most countries used in the analysis present either a cumulatively decreasing or increasing gap as was shown before. However, not all countries present this behavior. Despite the results lining up between the different models, this limitation hinders the model because some countries do not have cumulative trends.

When calculating achievement gaps I use the cognitive scores provided by PISA and follow estimations which have been replicated and validated in other studies. However, there have been major criticisms as to what exactly is the PISA test measuring. Perhaps these are tests that confound true ability with teaching-to-the-test. The previous phenomena is not quite clear for PISA given that the test is not announced with enough time for teachers to prepare. However, it is true that these types of tests have received extensive criticism [16] and further research should keep that in mind. These tests contain measurement error and are not bullet-proof in terms of the underlying latent variable being measured.

1.7 Conclusion

The literature on achievement inequality has recently started studying the evolution of the achievement gap and it has uncovered a wide variety of results. The main finding is that the gap of the U.S. has been widening over time by about 40-50% from 1970 until the 2000. Others find that this recent trend has reversed, at least for an age-specific group since the 2000's. From an international perspective, recent studies have shown that

[16] For example, see the article by the guardian https://www.theguardian.com/education/2014/may/06/oecd-pisa-tests-damaging-education-academics where over 100 academics signed a petition for the OECD to alter the logistics of PISA

there is wide variability in this trend, with some countries showing decreases while others increases. This chapter looks to investigate the magnitude of these gaps across selected countries, and attempts to explain why they might be happening.

The chapter looks at the evolution of the achievement gap and finds starking variation for the 15-year old achievement gap across countries. The U.S. has been closing its 90/10 gap by about 50% with a wide interval going down to nearly 35%. This reduction is also present in Germany, shockingly a country with an institutionalized curricular tracking system. Other countries such as France and Austria experience the contrary with persistent increases in the achievement gap. Moreover, once I disaggregate these patterns into a more fine-grained analysis, I discover that the dynamics are different. The gap of the U.S. is shrinking because the top SES group is coming down at a faster rate than the bottom group is catching up. Germany, which has a similar reduction, portrays a different explanation. Both the top and bottom groups are closing together at a similar rate. These results once aggregated to all countries show a general pattern across all 32 countries: either the bottom catches up faster than the top comes down or viceversa. There seems to be a trade off between these forces. This pattern is interesting and there is still is not an explanation as to why it happens. I leave up to future research to test the mechanisms and explanations through which this trade off occurs.

I propose to study tracking and vocational enrollment as possible explanations to the evolution of the achievement gap and find that the degree of curricular tracking of a country seems to explain this phenomena rather strongly. Tracking alone explains nearly 30% of between-country variability in achievement gaps. However, the results suggest that it should not be studied without factoring in the degree of vocational enrollment. I find that if a country has a low level of curricular tracking, vocational enrollment will not help decrease the achievement gap. Conversely, once curricular tracking enters the picture, vocational enrollment can help ease the burden of inequality and reduce the achievement gap significantly. In a similar vein, I use this model to perform a simulation for each country: if every country switched to little or no curricular tracking, what would be the resulting reduction in the 90/10 achievement gap? I find that if all countries switched, the average gap would shrink by about 11%, with outlier countries like the Netherlands and Hungary experiencing reductions of 48% and 46% respectively. I contrast the former simulation with another counterfactual scenario: what would be the average gap increase if all countries switched to a system where curricular tracking predominates? I record an average widening of 51%, something that suggests that increasing curricular tracking is more than three times worse than reducing it.

I conclude by showing that curricular tracking and vocational enrollment are playing an important role in explaining the evolution of the achievement gap. I find that these results are suggestive that institutional features, more specifically, stratified curriculums, might be playing a sizable role in the promotion of inequality. After having ran a battery of tests and different empirical strategies, the results are consistent and point out that we should pay attention to the source of the achievement gaps and study whether particular SES groups are contributing more to the gap than others.

Future research should attempt to replicate these results under further empirical tests. If other studies manage to replicate and corroborate these results, then policymakers should begin to consider vocational enrollment not only as a 'labor market' solution to youth unemployment but also as a compensatory mechanism for countries which have highly tracked curriculums.

Bibliography

Bauchmuller, Robert, Mette Gortz, and Astrid Wurtz Rasmussen (2014). "Long-run benefits from universal high-quality preschooling." In: *Early Childhood Research Quarterly* 29.4, pp. 457–470.

Bernardi, Fabrizio and Gabriele Ballarino (2016). *Education, occupation and social origin: a comparative analysis of the transmission of socio-economic inequalities*. Edward Elgar Publishing.

Bjorklund, Anders and Markus Jantti (2009). "Intergenerational income mobility and the role of family background." In: *Oxford Handbook of Economic Inequality, Oxford University Press, Oxford*, pp. 491–521.

Bol, Thijs and Herman G Van de Werfhorst (2013). "Educational systems and the trade-off between labor market allocation and equality of educational opportunity." In: *Comparative Education Review* 57.2, pp. 285–308.

Bradbury, Bruce, Miles Corak, Jane Waldfogel, and Elizabeth Washbrook (2015). *Too many children left behind: The US achievement gap in comparative perspective*. Russell Sage Foundation.

Breen, Richard and John Goldthorpe (1997). "Explaining educational differentials: Towards a formal rational action theory." In: *Rationality and society* 9.3, pp. 275–305.

Breen, Richard and Jan O Jonsson (2007). "Explaining Change in Social Fluidity: Educational Equalization and Educational Expansion in Twentieth-Century Sweden 1." In: *American Journal of Sociology* 112.6, pp. 1775–1810.

Byun, Soo-yong and Kyung-keun Kim (2010). "Educational inequality in South Korea: The widening socioeconomic gap in student achievement." In: *Globalization, changing demographics, and educational challenges in East Asia*. Emerald Group Publishing Limited, pp. 155–182.

Campbell, Frances A, Craig T Ramey, Elizabeth Pungello, Joseph Sparling, and Shari Miller-Johnson (2002). "Early childhood education: Young adult outcomes from the Abecedarian Project." In: *Applied developmental science* 6.1, pp. 42–57.

Chetty, Raj, David Grusky, Maximilian Hell, Nathaniel Hendren, Robert Manduca, and Jimmy Narang (2016). *The Fading American Dream: Trends in Absolute Income Mobility Since 1940.* Tech. rep. National Bureau of Economic Research.

Chetty, Raj, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez (2014). *Where is the land of opportunity? The geography of intergenerational mobility in the United States.* Tech. rep. National Bureau of Economic Research.

Chmielewski, Anna K (2016). "Changes in Socioeconomic Achievement Gaps in International Comparison, 1964-2012." In: *2016 APPAM International Conference.* Appam.

Chmielewski, Anna K and Sean F Reardon (2016). "Patterns of Cross-National Variation in the Association Between Income and Academic Achievement." In: *AERA Open* 2.3, p. 2332858416649593.

Coleman, James S et al. (1966). "Equality of educational opportunity." In:

Cunha, Flavio, James Heckman, Lance Lochner, and Dimitriy V Masterov (2006). "Interpreting the evidence on life cycle skill formation." In: *Handbook of the Economics of Education* 1, pp. 697–812.

Duncan, Greg J and Katherine Magnuson (2011). "The nature and impact of early achievement skills, attention skills, and behavior problems." In: *Whither opportunity*, pp. 47–70.

Duyme, Michel, Annick-Camille Dumaret, and Stanislaw Tomkiewicz (1999). "How can we boost IQs of "dull children"?: A late adoption study." In: *Proceedings of the National Academy of Sciences* 96.15, pp. 8790–8794.

Erikson, Robert, John Goldthorpe, and Lucienne Portocarero (1979). "Intergenerational class mobility in three Western European societies: England, France and Sweden." In: *The British Journal of Sociology* 30.4, pp. 415–441.

Esping-Andersen, Gøsta, Irwin Garfinkel, Wen-Jui Han, Katherine Magnuson, Sander Wagner, and Jane Waldfogel (2012). "Child care and school performance in Denmark and the United States." In: *Children and Youth Services Review* 34.3, pp. 576–589.

Fryer J, Roland G and Steven D Levitt (2004). "Understanding the black-white test score gap in the first two years of school." In: *Review of Economics and Statistics* 86.2, pp. 447–464.

Gamoran, Adam (2001). "American schooling and educational inequality: A forecast for the 21st century." In: *Sociology of education*, pp. 135–153.

Ganzeboom, Harry BG (2010). "A new International Socio-Economic Index (ISEI) of occupational status for the International Standard Classification of Occupation 2008 (ISCO-08) constructed with data from the ISSP 2002-2007." In: *Annual Conference of International Social Survey Programme, Lisbon*. Vol. 1.

Ganzeboom, Harry BG and Donald J Treiman (1996). "Internationally comparable measures of occupational status for the 1988 International Standard Classification of Occupations." In: *Social science research* 25.3, pp. 201–239.

Gelman, Andrew (2015). *The connection between varying treatment effects and the crisis of unreplicable research: A Bayesian perspective.*

Hanushek, Eric et al. (2006). "Does educational tracking affect performance and inequality? Differences-in-differences evidence across countries." In: *The Economic Journal* 116.510, pp. C63–C76.

Heckman, James (2006). "Skill formation and the economics of investing in disadvantaged children." In: *Science* 312.5782, pp. 1900–1902.

Jaeger, Mads Meier and Anders Holm (2007). "Does parents' economic, cultural, and social capital explain the social class effect on educational attainment in the Scandinavian mobility regime?" In: *Social Science Research* 36.2, pp. 719–744.

Kautz, Tim, James Heckman, Ron Diris, Bas ter Weel, and Lex Borghans (Dec. 2014). *Fostering and Measuring Skills: Improving Cognitive and Non-Cognitive Skills to Promote Lifetime Success*. Working Paper 20749. National Bureau of Economic Research. DOI: 10.3386/w20749. URL: http://www.nber.org/papers/w20749.

Magnuson, Katherine and Jane Waldfogel (2008). *Steady gains and stalled progress: Inequality and the Black-White test score gap*. Russell Sage Foundation.

Micklewright, J and SV Schnepf (2007). *Inequality and Poverty Re-Examined*. Oxford Univ. Press. Chap. Inequality of learning in industrialized countries.

OECD (2002). *Education At A Glance: OECD Indicators 2002*. OECD Publishing. ISBN: 9264198903.

— (Mar. 2012). *PISA 2009 Technical Report*. OECD Publishing.

— (Dec. 2016). *PISA 2015 Results (Volume I)*. OECD Publishing.

Piketty, Thomas and Emmanuel Saez (2003). "Income inequality in the United States, 1913–1998." In: *The Quarterly journal of economics* 118.1, pp. 1–41.

Reardon, Sean F (2011). “The widening academic achievement gap between the rich and the poor: New evidence and possible explanations.” In: *Whither opportunity*, pp. 91–116.

Reardon, Sean F and Ximena A Portilla (2015). “Recent trends in socioeconomic and racial school readiness gaps at kindergarten entry.” In: *Stanford, CA: Center for Education Policy Analysis. Retrieved December* 20, p. 2015.

Saw, Guan Kung (2016). “Patterns and trends in achievement gaps in Malaysian Secondary Schools (1999–2011): gender, ethnicity, and socioeconomic status.” In: *Educational Research for Policy and Practice* 15.1, pp. 41–54.

Shavit, Yossi and Hans-Peter Blossfeld (1993). *Persistent Inequality: Changing Educational Attainment in Thirteen Countries. Social Inequality Series.* ERIC.

Van de Werfhorst, Herman G and Jonathan JB Mijs (2010). “Achievement inequality and the institutional structure of educational systems: A comparative perspective.” In: *Annual review of sociology* 36, pp. 407–428.

Waldfogel, Jane (2006). “What do children need?” In: *Juncture* 13.1, pp. 26–34.

Wu, Margaret (2005). “The role of plausible values in large-scale surveys.” In: *Studies in Educational Evaluation* 31.2-3, pp. 114–128.

1.8 Appendix

	2000				2015			
	Low SES		High SES		Low SES		High SES	
	Count	Percentage	Count	Percentage	Count	Percentage	Count	Percentage
Gender								
- Female	150	60%	146	48%	300	55%	301	52%
- Male	98	39%	154	51%	245	45%	281	48%
- NA	2	1%	3	1%	0	0%	0	0%
Highest edu in HH								
- No schooling	4	2%	0	0%	25	5%	0	0%
- Primary education	30	12%	0	0%	1	0%	0	0%
- Lower secondary education	68	27%	0	0%	286	52%	0	0%
- Upper secondary education	33	13%	0	0%	6	1%	0	0%
- Post-secondary non-tertiary education	3	1%	2	1%	15	3%	1	0%
- Bachelors or above	0	0%	282	93%	1	0%	560	96%
- NA	112	45%	19	6%	211	39%	21	4%
Father born in country of test								
- Yes	112	45%	278	92%	284	52%	523	90%
- No	105	42%	23	8%	244	45%	56	10%
- NA	33	13%	2	1%	17	3%	3	1%
Number of books in the HH								
- 0-10	74	30%	0	0%	171	31%	3	1%
- 11-100	135	54%	15	5%	307	56%	27	5%
- 101-250	23	9%	51	17%	39	7%	94	16%
- 251 or more	12	5%	232	77%	14	3%	454	78%
- NA	6	2%	5	2%	14	3%	4	1%

Table 1.1: Descriptives of selected variable for 2000 and 2015 for all countries pooled

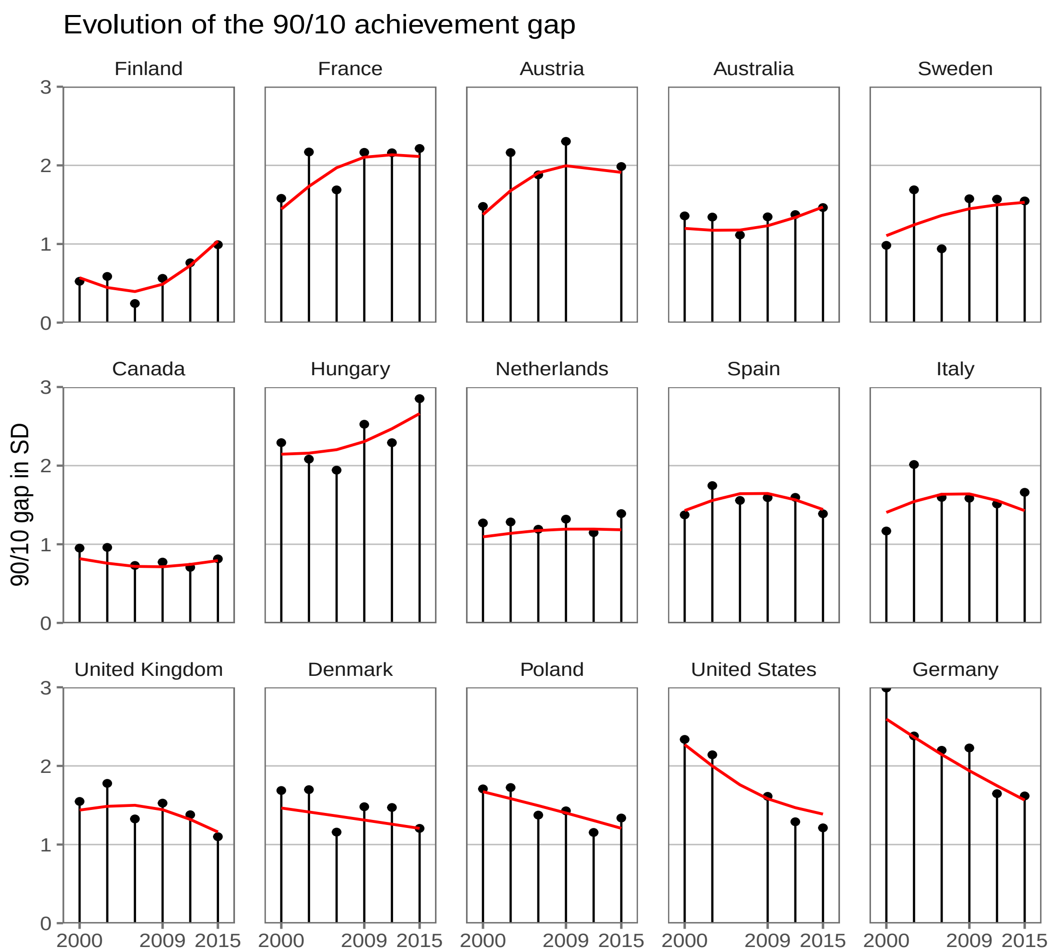

Figure 1.1: 90/10 achievement gaps in literacy expressed as standard deviations for selected countries between 2000 and 2015

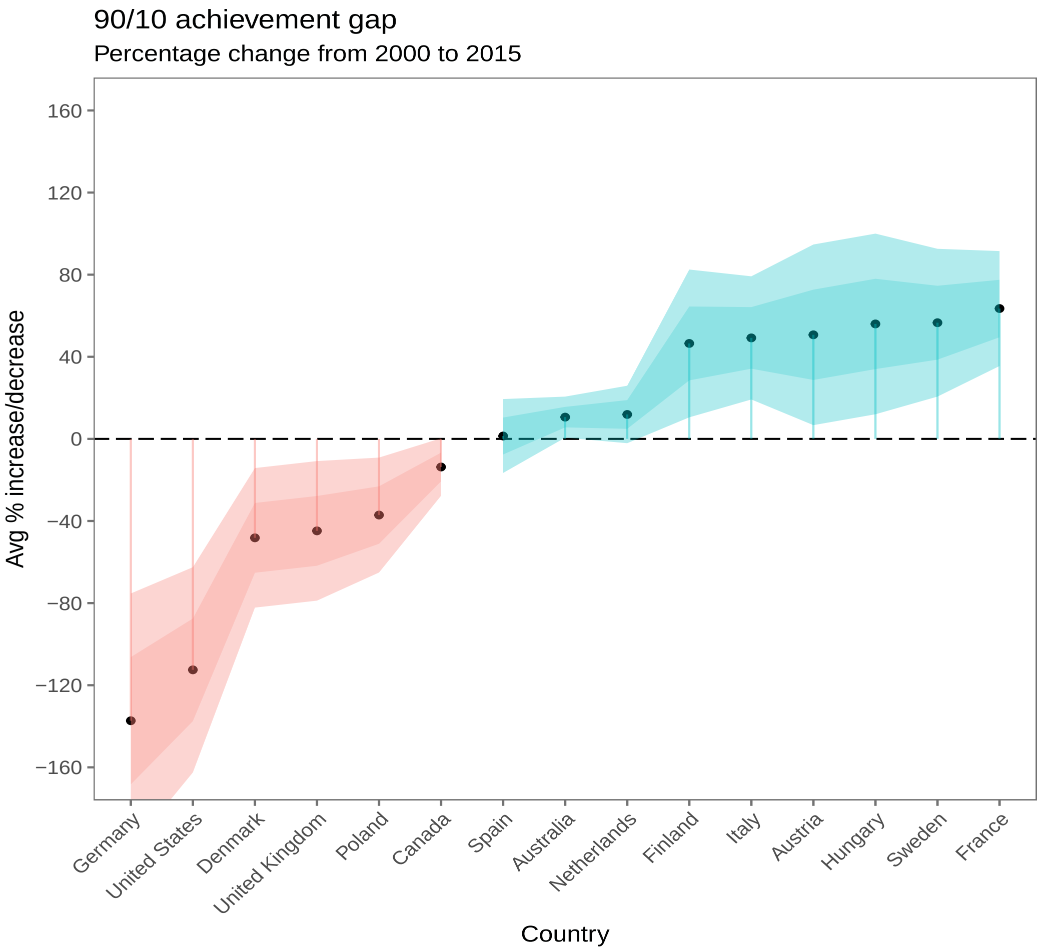

Figure 1.2: Percentage change in literacy between years 2015 and 2000 for selected countries. Red regions represent decreases while blue regions represent increases in the achievement gap

Country	Number of tracks	Age of selection	Inequality	Predicted	% reduction
Austria	More than one track	Less than 15	1.58	1.16	28%
France	More than one track	15 or more	1.84	1.62	12%
Germany	More than one track	Less than 15	1.94	1.30	34%
Hungary	More than one track	Less than 15	2.32	1.24	47%
Italy	More than one track	Less than 15	1.48	0.86	42%
Netherlands	More than one track	Less than 15	1.06	0.55	49%
Poland	More than one track	15 or more	1.43	1.22	15%

Table 1.2: Simulation: reduction of achievement gap if countries switched to 'ideal' tracking

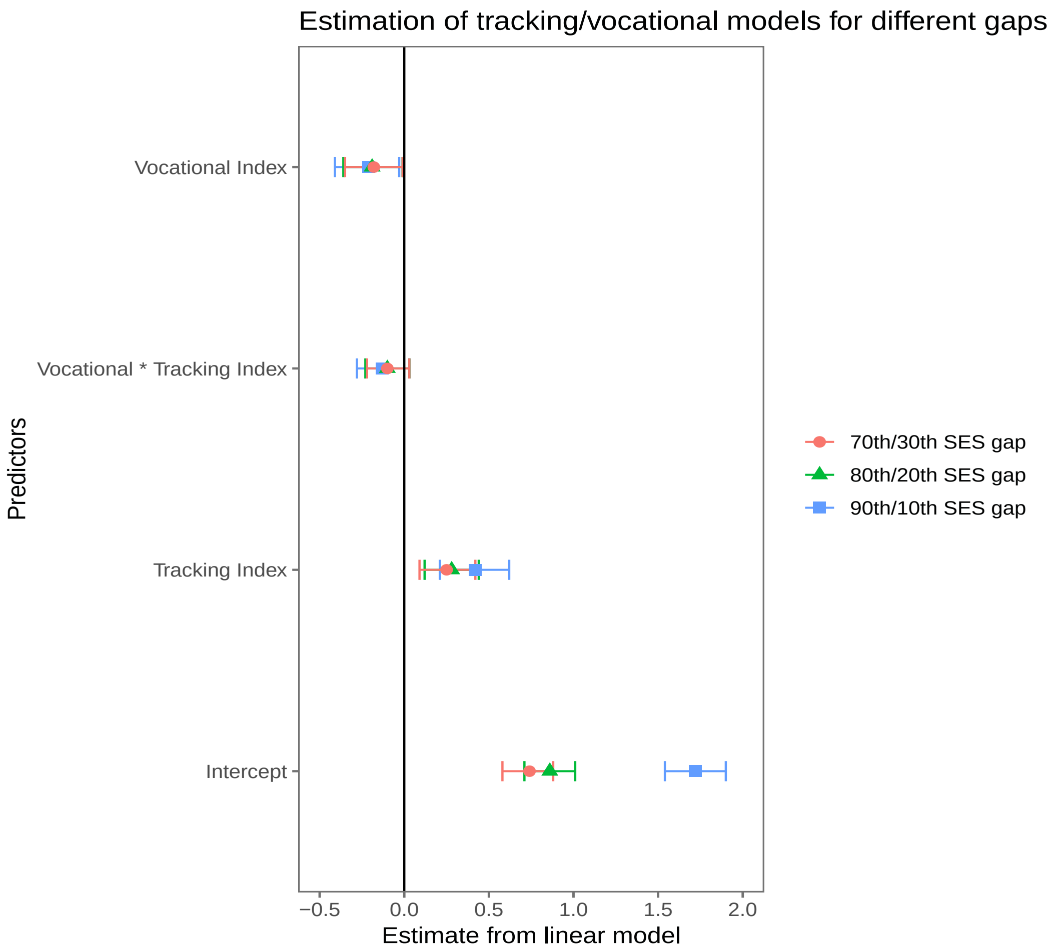

Figure 1.3: Explaining achievement gaps - Model comparisons

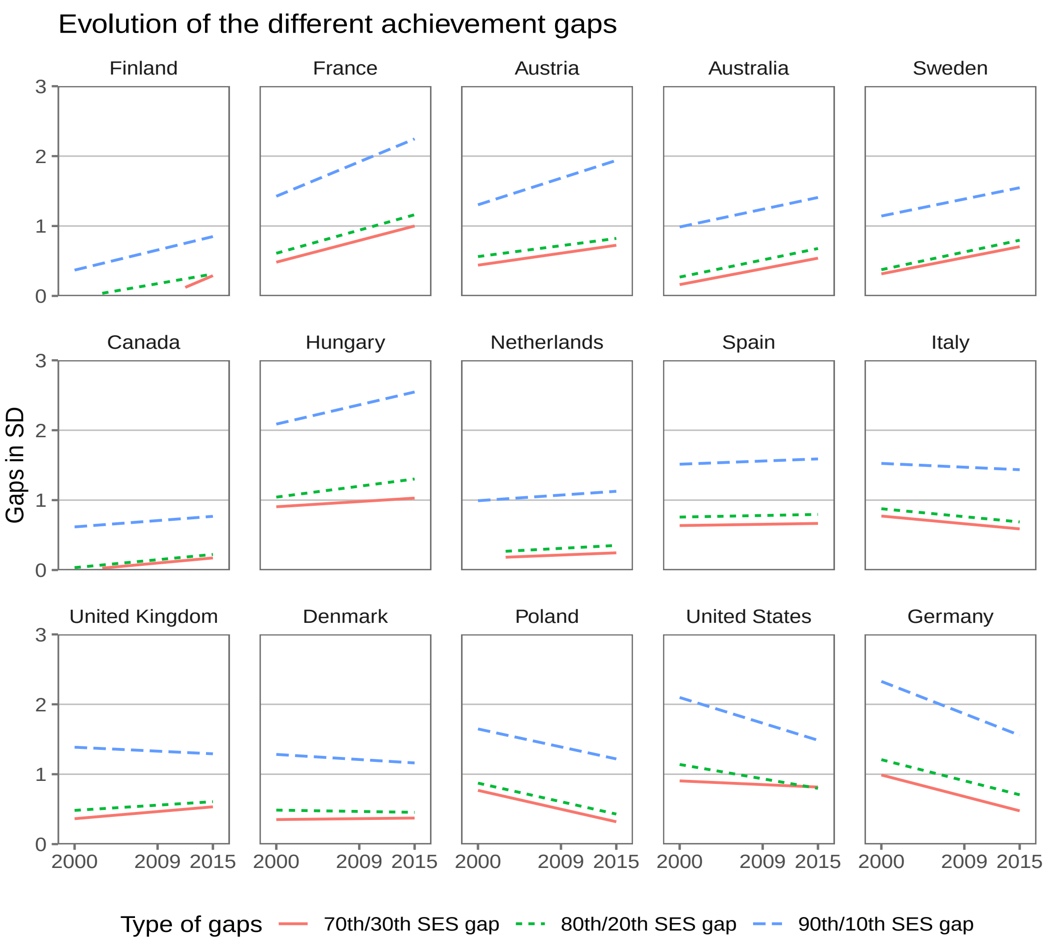

Figure 1.4: Evolution of the achievement gap for several gaps

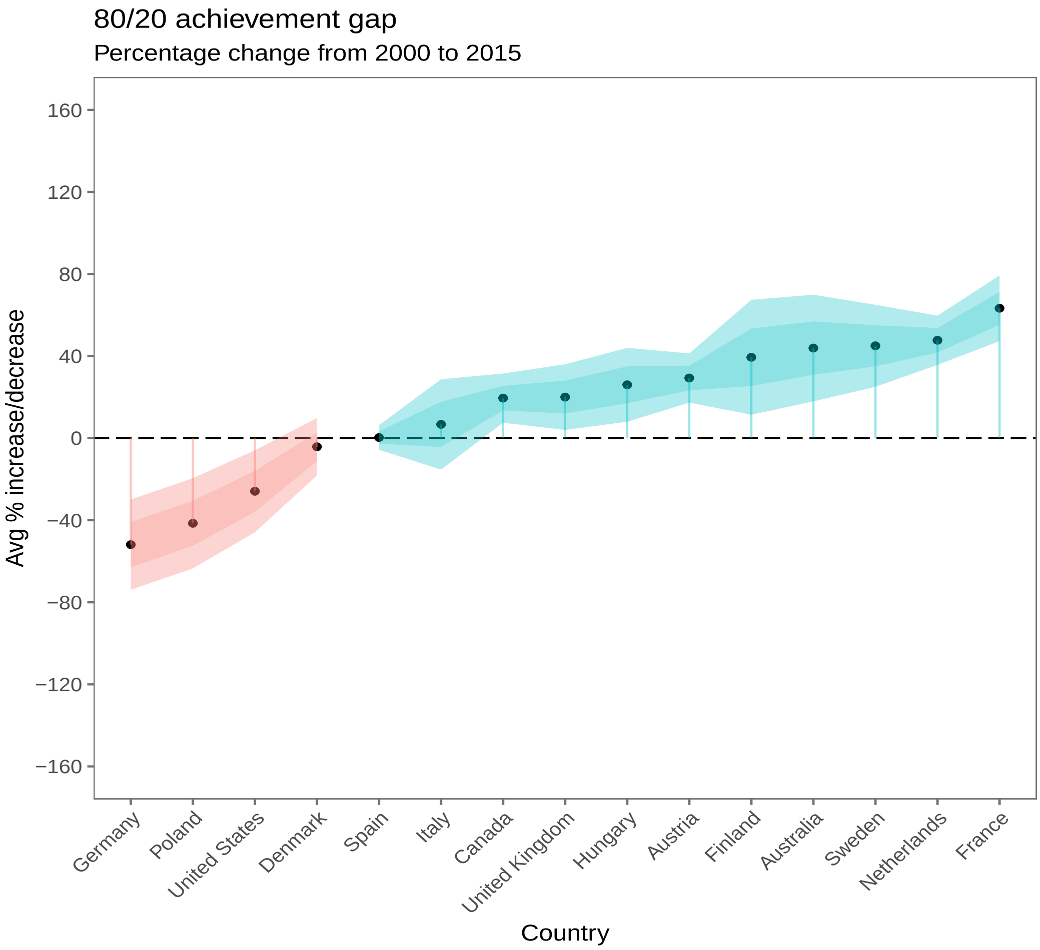

Figure 1.5: Percentage change in the 80/20 achievement gap from 2000 to 2015

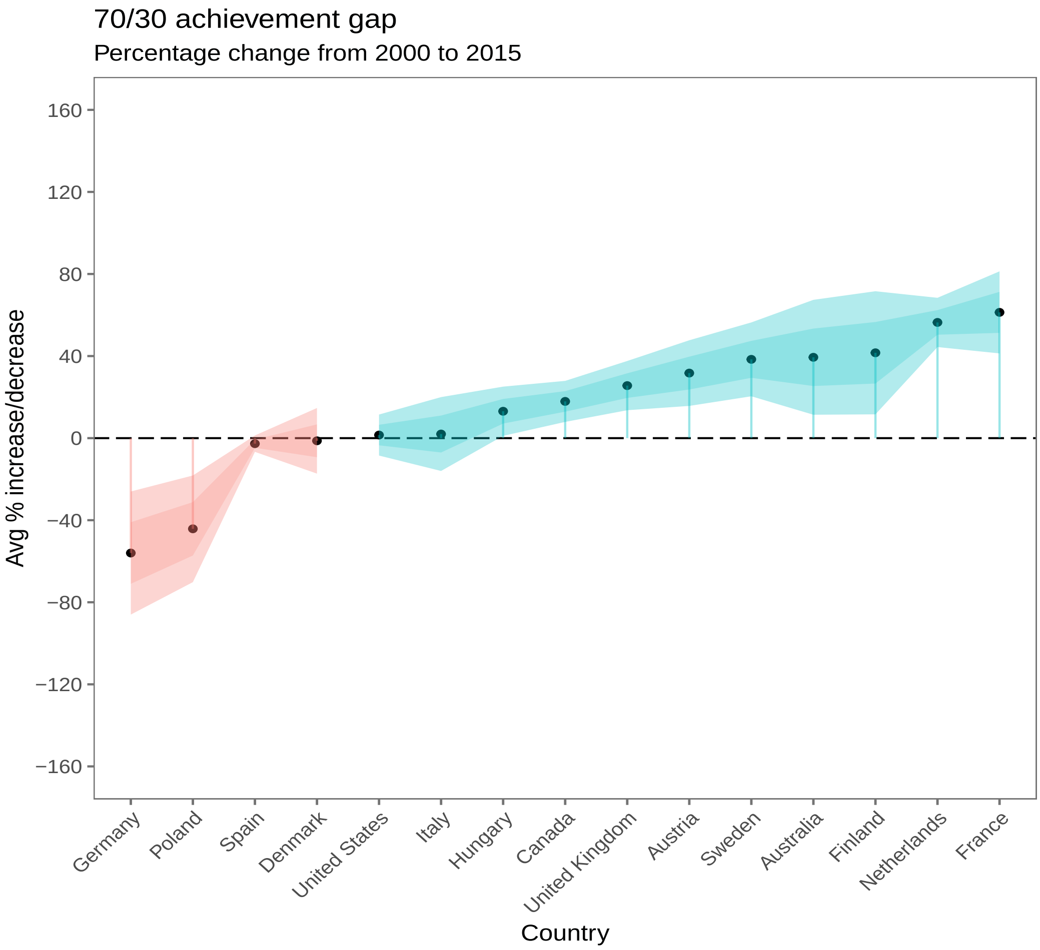

Figure 1.6: Percentage change in the 70/30 achievement gap from 2000 to 2015