

First paper - Draft

Jorge Cimentada

June 21, 2017

Contents

1 Literature Review	1
2 Research questions	4
3 Methods	5
3.1 Data	5
3.2 Data analysis	6
4 Analysis	9
4.1 Descriptives	9
4.2 Results	9

1 Literature Review

Recent research on educational inequality has found that differences in test performance between High-SES and Low-SES kids has been growing quickly over the years. The literature on educational inequality has mainly concentrated on the United States ([Reardon, 2011](#)) but other international evidence is emerging with other similar findings for other count. The United States is usually the case study of interest, firstly, because it is the only country where cognitive testing is very widespread across surveys. And secondly, this trend has allowed to have testing records as early as 1940 (check reardon) until present day. Using this information, ([Reardon, 2011](#)) is the first to investigate the evolution of the cognitive gap and the results are very 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 (cite study of white-black gap -waldfogel study). The widening of the achievement gap has been happening in parallel to the growth of income inequality. Although very suggestive, it is hard to link both things causally.

[Reardon \(2011\)](#) finds that the increase has occurred predominantly from the 1970's until the 2000's. In fact, hard numbers suggest that the gap increased by about 40-50%. The author also estimates the rate of change using data from the 1940's and finds an even higher increase of about 75%. Since the studies before the 1970's lack a lot of rigor in comparability and sampling, the author

computes everything for the sample before/after 1970's separately¹. Finally, [Reardon \(2011\)](#) concludes that the U.S gap between the 90th and 10th income percentile is at about 1.25 standard deviations. [Bradbury et al. \(2015\)](#) don't find results which are very far away from this. They find that for 14 year olds in the United States, the gap is above 1 SD but lower than 1.25. [Duncan and Magnuson \(2011\)](#) find similarly to the previous studies and confirm a gap of 1.25/1.50 standard deviations.

One important drawback of these studies is that they don't present the uncertainty of this estimate. Not necessarily to gauge their statistical significance, but to simply asses how much we can trust their accuracy. Despite the recent surge in understanding achievement gaps as a function of social background, sociologists have already amassed a great deal of findings concerning inter-generational inequalities, one notable example being that it's very clear that large inequalities in parent's education tend to reproduce large inequalities in children's education ([Breen and Jonsson, 2007](#)). Virtually in all countries, be it developed or developing, there is inequality of opportunity. But there is considerable variation in the magnitude of this effect. For example, the Scandinavian countries, particularly Denmark and Sweden, prove to be very mobile countries ([Esping-Andersen and Wagner, 2012](#); [Breen and Jonsson, 2007](#); [Shavit and Blossfeld, 1993](#)).

With very recent data, [Reardon and Portilla \(2015\)](#) have unexpectedly uncovered a new finding: the reversal of the trend. From 1998 until 2010, it seems like the income achievement gap² decreased at a rate of 0.01 and 0.008 standard deviations per year. In contrast to ([Reardon, 2011](#)), we find that in a 30-year span the gap was systematically increasing at a rate of 0.02, something not dramatically different from the previous estimates. The American situation is a very distinctive one relative to Europe. The reasons why we're seeing a reversal in the trend could be numerous and should be studied very close to context. For example, the American educational system lacks any formal curricular differentiation *la* European style. In contrast, [Reardon \(2011\)](#) suggest that the reversal is likely due to the high increase of preschool enrollment. They suggest that in this same period (1998 - 2010) the income achievement gap in early schooling enrollment decreased substantially.

Some authors have taken this analysis to an international context in order to discover between-country trends. The comprehensive work of [Bradbury et al. \(2015\)](#) performs a comparative analysis of Australia, United Kingdom, United States and Canada. Their research design is very 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 14³. The core finding behind the book is that the American achievement gap is much wider than in any other comparison countries. As mentioned earlier, they find that once the achievement gap is present in early school entry, it doesn't seem to narrow very much over the life course. They find that the achievement gap is very stable all the way from early education to secondary school (although their studies cover up until age 14).

One limitation of their study is that they concentrate on countries which have a very particular educational structure, namely the fact that there is little formal stratification in terms of curricula. These four countries have no major jump in school selection, quite the opposite to the European average countries. 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 and Woessman (tracking study), use the TIMSS and PISA surveys to gauge whether highly tracked countries do indeed increase inequality after students pass the age at first selection of tracking. Moreover, Mickleweight & Schnepf (2007) using similar data but different design finds that countries which have a high level of 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.

¹Even after weighting appropriately.

²The income achievement gap is the difference between the 90th and 10th income percentile in the test of interest

³To the best of my knowledge this is the only study that uses panel data to study achievement gaps, let alone to do this between countries

Despite their high level of rigour, their analysis is based on four surveys that have significant differences in terms of questions, sampling and populations and cannot be easily compared. Their findings are very reliable but should be taken as suggestive at best. For this reason we should also pay particular attention to studies such as Chmielewski and Reardon (2016) and Chmielewski (2017) which have attempted not only to compare gaps between countries, but to evaluate whether there is a general increase in educational inequality in many countries using similar studies. These studies tackle a completely different question from the above, but they do provide support for the overall finding that the achievement gap is certainly not narrowing over time.

Chmielewski and Reardon (2016), using the Programme for International Student Assessment (PISA), the Trends in International Mathematics and Science Study (TIMSS) and PIRLS (the Progress in International Reading Literacy Study), assess whether there are patterns of cross-national variation in the achievement gap. In other words, is the size of achievement gap different for many countries? Although there is potential to study countries over time using these surveys, they choose not to. They do so because their question of interest (income categories) was only asked in three waves; in the end they only had three countries repeated in all three waves.

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 a little over 1 standard deviation in the year 2005 (for 15 year olds) and Germany has a decreasing gap from around 1.25 to 1 standard deviation. They go even further and link this achievement gap to several country-level indicators related to income inequality, school differentiation, central exams, etc ... The correlations are indeed very suggestive but clearly we must be extremely cautious in drawing causal changes from these two variables. But it is important to stress that their design is quite different from the work of Reardon (2011). Reardon (2011) takes studies in the U.S starting from the 1940's until 2015⁴ and makes them comparable across time. This gives the author a very long time series to build a reliable achievement gap (over 40 years). In their study, Chmielewski and Reardon (2016) change the aim of the study to model between-country differences from a cross-sectional perspective.

Chmielewski (2017), 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 the year 1964 until 2015 in order to discover differences between and across countries. With over 50 years of data, and over 100 countries, Chmielewski (2017) finds that there seems to be a general world increase in the achievement gap. However, once she disentangles the relationship by country, she finds a reasonable amount of heterogeneity, with some countries seeing the achievement gap closing, others no change at all, while others record a steady increase. One clear limitation of their study (as well as Reardon (2011)) is that they adjust for the age of each child in all studies. Although for their purposes is the right thing to do⁵, they are masking age-specific achievement gaps by controlling for age, such as Reardon (2011) did.

The evolution of High/Low SES gaps for preschool children might be much less marked than the same gap for high school children. The explanation, although very debated, has been gaining much support in recent years. In countries with high levels of curricular differentiation, the transition from early schooling into the tracking system has been found to increase inequality of learning (Wossman and Hanushek diff-in-diff). Moreover, the vast sociological literature on educational transitions systematically finds that tracking tends to foster between-track inequality rather than erode their differences by tackling their specific needs (Van der Werfhorst and Mijs 2011). Based on this, we cannot simply assume that the achievement gap has been neither constant across cohorts (because there have been tracking reforms in many countries, introducing as well as elimination tracking structures) nor

⁴Each study is independent of each other meaning that they might've been given to 6 year olds as well as to 12 year olds. Although the author adjusts for age, the trends can't be generalized to age-specific, but rather in overall terms.

⁵The differences in achievement could simply be due to changes in cognitive abilities across the lifetime. However, as we've noted before, Corak and Waldfogel (2015) find that the achievement gap is very stable across the life time

the same between ages, because tracking/no tracking might exacerbate the achievement gap.

With this being said, this paper introduces one novelty in the literature which is to evaluate the evolution of the High/Low SES achievement gap in the past 15 years for all PISA participant countries. This is different from previous work because it concentrates solely on 15 year old children, and it attempts to capture the evolution of the achievement gap for each country. The advantages of this study are numerous. First, we concentrate on the evolution of the gap for only 15 year olds. This is different from all the work of Chmielewski and Reardon, in which they pool all ages to get average changes, to study the properties of specific age-groups. As we've seen before, there are reasons to think that specific age-groups have seen changes in the achievement gap. Moreover, in almost all countries with a tracked curriculum children are either at or in the process of tracking by the age 15, meaning that we will be able to link whether tracked countries are the most variable in their evolution of achievement gaps.

The rule of thumb to choose the 90th, 50th and 10th percentile is arbitrary, others have used, for example, the 5th, 50th and 95th. (Micklewright & Schnepf, 2007)

Calculate the difference between 90/10 and then the ratio of this with the annual gain of one year of schooling (Mickleweight & Schnepf did it). They also find that France is a lower dispersion country (but in my results this is changing a lot). They also find that Germany and US are very dispersed countries (quite the opposite)

Mickleweight & Schnepf find that France has a very small gap in 2000, something we find very clearly in our data, and the U.S has a very big gap in 2000, something we also find in our data. Mickleweight & Schnepf find that average achievement is negatively related to dispersion.

Is efficiency related to opportunity here? Is efficiency what explains the inequality of opportunity?

Talk about how the achievement gap has been widening in Malaysia and South Korea (or Japan?) Note how many of these studies haven't really concentrated on who is getting better or worse: top or bottom? Talk more about how countries with high social mobility has been linked to smallest achievement gaps

Talk about Durpiez and Dumay and how there's not relationship between inequality income - inequality achievement in contrast to reardon and anna who find some relationship.

2 Research questions

This paper is interested in two questions. As explained in the literature review, we already know that U.S achievement gaps have been growing over time, and some international evidence suggests that it might be increasing in other countries as well. All of these studies pool all students together from different ages and estimate global trends.

We take a different approach and we want to study age-specific trends in achievement gaps. We're interested in studying a) the size of the achievement in a comparative perspective and its relationship to overall achievement levels, b) which countries are experiencing changes in the achievement gap, c) the rate at which each country-gap is widening/narrowing and d) establishing whether the gaps are widening/narrowing because particular groups are getting ahead/behind.

We develop each question separately for more detail.

a) We want to test the notion that better performing countries have lower levels of achievement gap. (Van de Werfhorst and Mijs, 2010) emphasize that there is empirical evidence that suggests this. This pattern is not so obvious. For example, countries with high levels of tracking could maximize

student performance, specially the high SES students, raising their overall performance and thus raising the national performance score. But if the bottom performers are not gaining at the same rate, then the achievement gap will inevitable grow resulting in a high performing countries with widening achievement gap.

b) 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 only 30 years, ([Reardon and Portilla, 2015](#)) stress that they also found a significant decrease in only 15 years of data. This lends credibility that our estimations will show variation over time. As usual, their estimations are only for children going from 4th to 8th grade. We concentrate on age-specific groups, in this case, 15 year olds.

c) We want to compare the percentage change at which the gap widened/narrowed from the first to the last year. This will give us a general idea of the overall change over time, and will allow us to compare our estimates to the actual literature ⁶

d) The widening/narrowing of the achievement gap has a source, which has been often studied to be related to everything from educational spending, income inequality, time allocation to students and preschool enrollment. The literature has concentrated very narrowly on whether the gap is increasing because the top performers are getting ahead, because both are distancing or because the bottom is falling behind. We shall pay particular attention to identifying the rate at which the top/bottom groups are evolving over time. This type of analysis is particularly useful in jotting down the mechanisms through which the gap is evolving. If the bottom performers are stagnated whereas the top is gaining ground, the efficiency-trade off dilemma of tracking might be much more credible than other explanations.

3 Methods

3.1 Data

In order to study the above mentioned questions I will be using 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 the year 2000 up until 2015, where recently, over half a million students were tested in mathematics, literacy and science in over 70 countries, both developed and developing ones.

PISA collects data through a two-stage stratified sampling design. With the help of official 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 PISA 2000 is 1.27388×10^5 , for PISA 2003, 2.76165×10^5 , for PISA 2006, 3.9875×10^5 , for PISA 2009, 5.15958×10^5 , for PISA 2012, 4.8081×10^5 , and for PISA 2015 5.19334×10^5 . 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 time to visualize changes in the structure of an educational system.

In order to measure family background PISA includes several variables. Firstly, the 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

⁶Although no study has performed this age-specific achievement gap for comparable tests over such a long time. Our results will serve as comparison for other studies that use age-specific groups, such as 4th graders.

waves have a different coding. Moreover, another limitation is the fact that parent's education is measured with the ISCED classification, something that has changed over time. For example, until PISA 2009, the ISCED 1997 classification was currently used, but the next wave already used the new ISCED 2011 classification. Both these classification schemes have equivalent look-up tables, yet this requires a detailed inspection of the codings.

Another social background variable PISA collects is the International Socio-Economic Index of Occupational Status (ISEI). This variable captures the economic status of the family, the closest it can without asking for income information. It has been scaled for comparability between waves and some authors have used it for inequality studies finding expected results to be consistent with social origin literature (Anna study about world-wide inequality). 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.

But one of the most relevant variables for our study is the 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 students parents, converted into years of schooling; the PISA index of family wealth; the PISA index of home educational resources; and the PISA index of possessions related to "classical" culture in the family home (Education at a Glance, OECD, Paris, 2002, Glossary). This variable, aside from capturing all relevant dimensions of SES, such as education, occupation, and material resources, takes care of transforming the variables into comparable measures across waves. In fact, we use the newly-released ESCS index () which rescaled all ESCS indexes to make them suitable for over-time analysis ⁷.

Aside from SES, the other most relevant variables are test scores. PISA provides plausible values instead of one single indicator of test score performance. As explained in the PISA manual (PISA 2012, 2015), these are imputed values that resemble individual test scores and have approximately the same distribution as the latent trait being measured. Suppose we have i , the average student test score in mathematics for student i . Instead of estimating i alone, plausible values estimate a distribution of possible s , for student i , together with the likelihood of each i based on the respondents answers on the test. This is defined as the posterior distributions of s for student i . The reason why we use this procedure is because estimating a single estimate i is plagued with measurement error, among other types of bias (see Wu, 2005). The plausible values in the PIAAC data are effectively five (ten for PISA 2015) random draws from this distribution. These variable is are continuous, ranging from 0 to 500, with a mean of 250.

3.2 Data analysis

The aim of this paper is to identify and disaggregate 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). The actual calculation of the achievement gap varies substantially and different strategies have been implemented. For example, Mickleweight & Funsche calculate the difference in achievement by crudely subtracting the gap 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. Guan Kung Saw (2015), for instance, used parental education as a proxy of SES, whereas Soo-yong Byun and Kyung-keun Kim (2010) use a similar SES index as ours, but created by them.

⁷These rescaled indices can be found in the [PISA website](#) under *Rescaled Indices for Trend Analyses*.

Reardon (2011) and subsequently (all reardon papers), used a different method developed by Reardon (2011) which we partially adopt in this paper. SES achievement gaps are measured as the difference in standardized achievement between the 90th and 10th percentiles of the chosen SES variable.

We start by standardizing achievement for each country, year and subject. Because PISA does not provide a single achievement indicator, we take the median of all plausible values for each student⁸, resulting in one single score.

We then fit a linear model

$$Y_i = \alpha + \beta_1 * AGE_i + \epsilon_i, \epsilon_i \sim N(0, \sigma^2) \quad (1)$$

for each wave, where Y_i is the median student test score and AGE_i is their age measured in months (because all children should be 15 year olds) weighted by the student sample weights.

We then adjust \hat{Y}_i by

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

Where ϵ_i is the residual for student i and the denominator is the root mean square error of the model.

Standardizing the median test score solves the problem of comparability of gaps measured with different tests, and across waves. However, if the variance of academic achievement changes over time, then standardizing the overall score at each wave of testing confounds changes in the gap with changes in the variance of test scores. That is, by standardizing we're forcing the variance of test scores to be zero across all waves. But if the true variance of academic achievement grows over time, then the estimated trend in the achievement gaps will be underestimated, and vice versa.

Inspection of the data suggests that it's something we shouldn't be deeply concerned with. The standard deviation of each wave seems to be following a very similar pattern with not so drastic exception of the year 2000.

Another concern is that if the gaps at different waves are measured with tests that have different amounts of measurement error, then the amount of bias will not be the same in each measure of the gap. This can be very misleading and suggest erroneous interpretations regarding trends in the of the gaps over time (Reardon, 2011). PISA has tried to make sure the tests are comparable across waves (PISA 2012 technical report), but we still have to adjust for this imprecision.

In order to correct gap estimates for measurement error, I adjust student test scores by

$$\hat{Y}_i = \hat{Y}_i * \frac{1}{\sqrt{r}} \quad (3)$$

where r is the reliability score of the wave.⁹ Each PISA survey provides a reliability indicator

⁸Since each plausible value is a random draw from a theoretical latent normal distribution of possible student achievement scores, the median should be precise in getting a central measure of the latent distribution.

⁹Other procedures multiply each country by their own reliability measure for each year-subject pair (Anna analysis of worldwide gap). 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 paper, they were unavailable. For this reason we implement the analysis following the original work of Reardon (2011)

which we use accordingly. This yields estimates of the true gaps, and eliminates any bias in the trend that may arise from differential reliability of the tests.

After constructing the adjusted test score measurement, we estimate the SES dummy by calculating the weighted 90th and 10th percentile scores for the SES composite index, and then generating a dummy of 1 for those above (including) the 90th percentile and 0 for those below (including) the 10th percentile.

We then fit

$$\hat{Y}_i = \alpha_{j[i]} + \beta_{j[i]} * SES_i + \epsilon_i, \text{ for } i = 1, 2, \dots, n \text{ within country } j \quad (4)$$

where SES_i is whether the student is in the top/bottom SES dummy and we allow both α_i and β_i to vary by country j . We implement this separately for each wave and weight by the student sample weights.

Finally, we calculate the SES achievement gap for each country by extracting the fitted α and β for each country j and calculating the difference of the predicted score of High and Low SES.

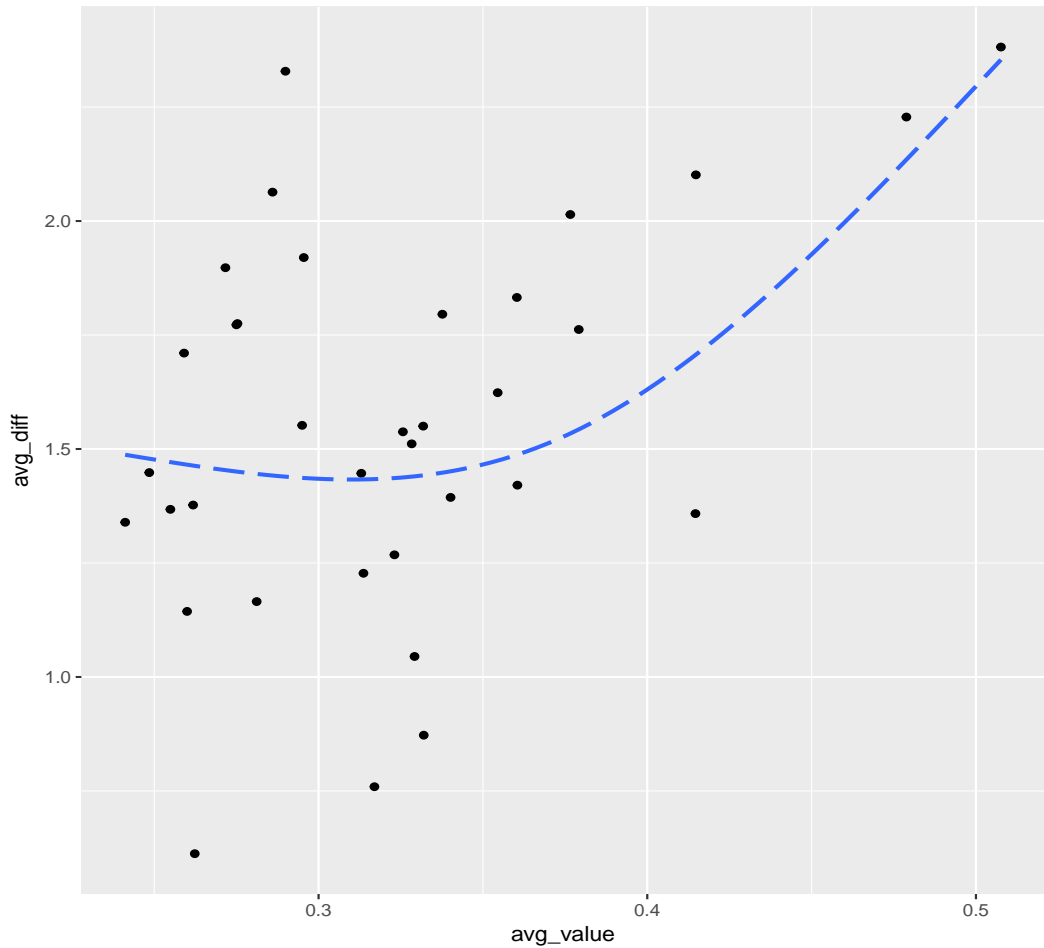
$$\begin{aligned} \text{High SES}_j &= \alpha_j + \beta_j \\ \text{Low SES}_j &= \alpha_j \\ \text{SES gap}_j &= \text{High SES}_j - \text{Low SES}_j \end{aligned} \quad (5)$$

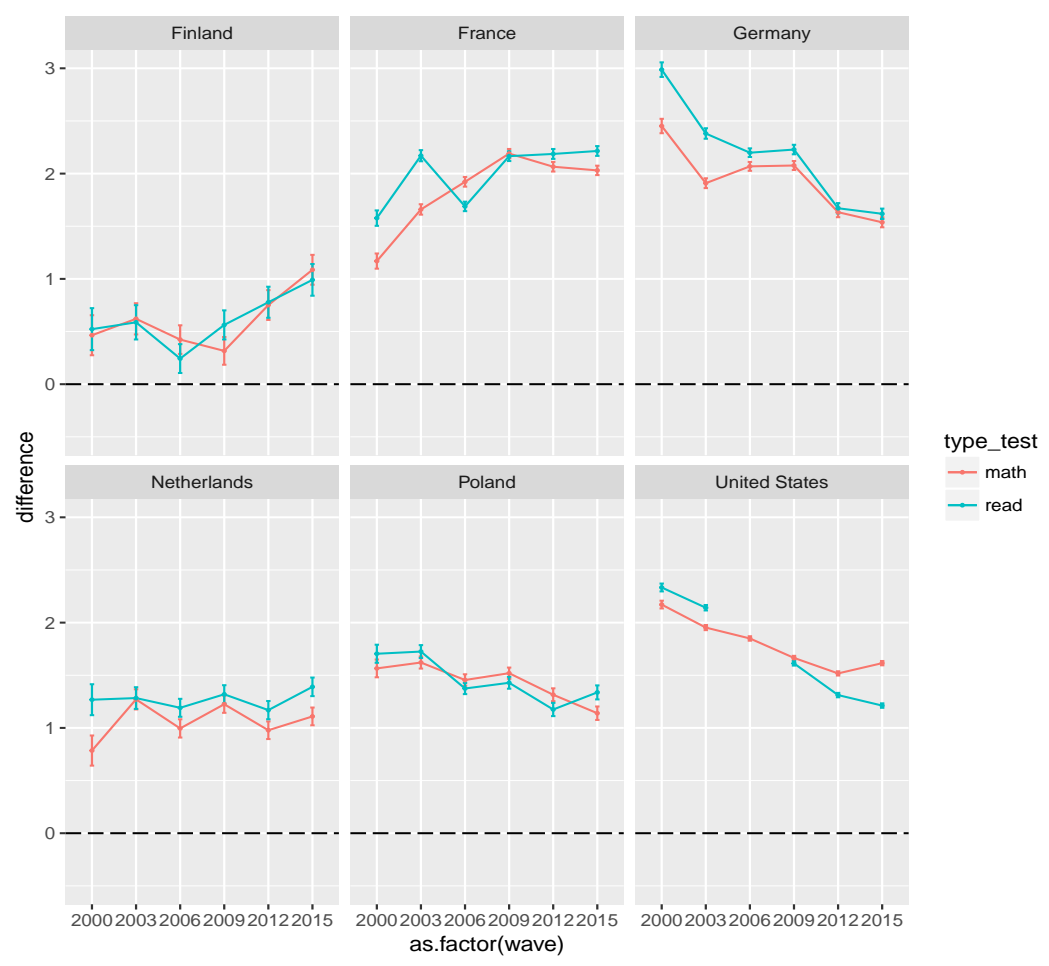
We also calculate the standard error of this difference and use it to create uncertainty intervals in the estimation of time trends. We fit a multilevel rather than a linear model because by pooling the information together, we weight countries appropriately to their sample size. Given that including the SES dummy reduces the sample size considerably, we want to be able to estimate each country-difference as accurately as possible.

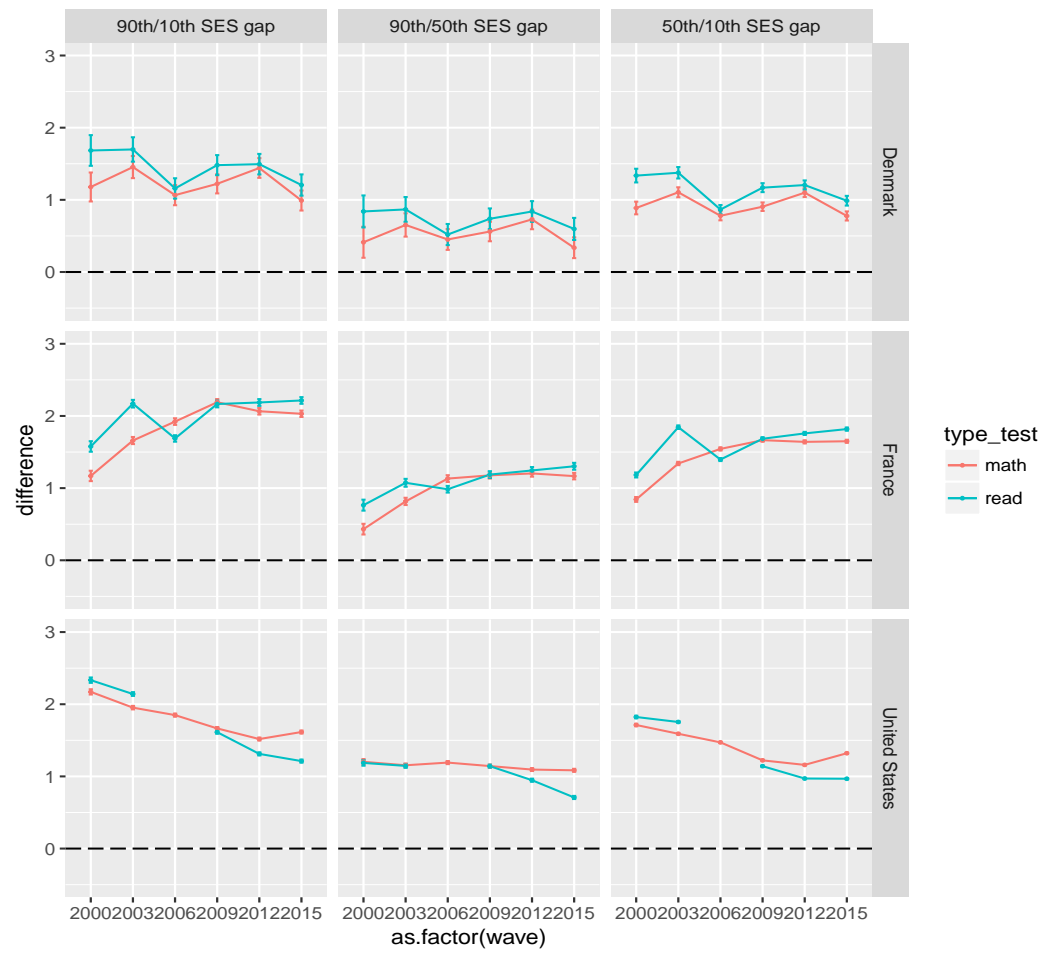
4 Analysis

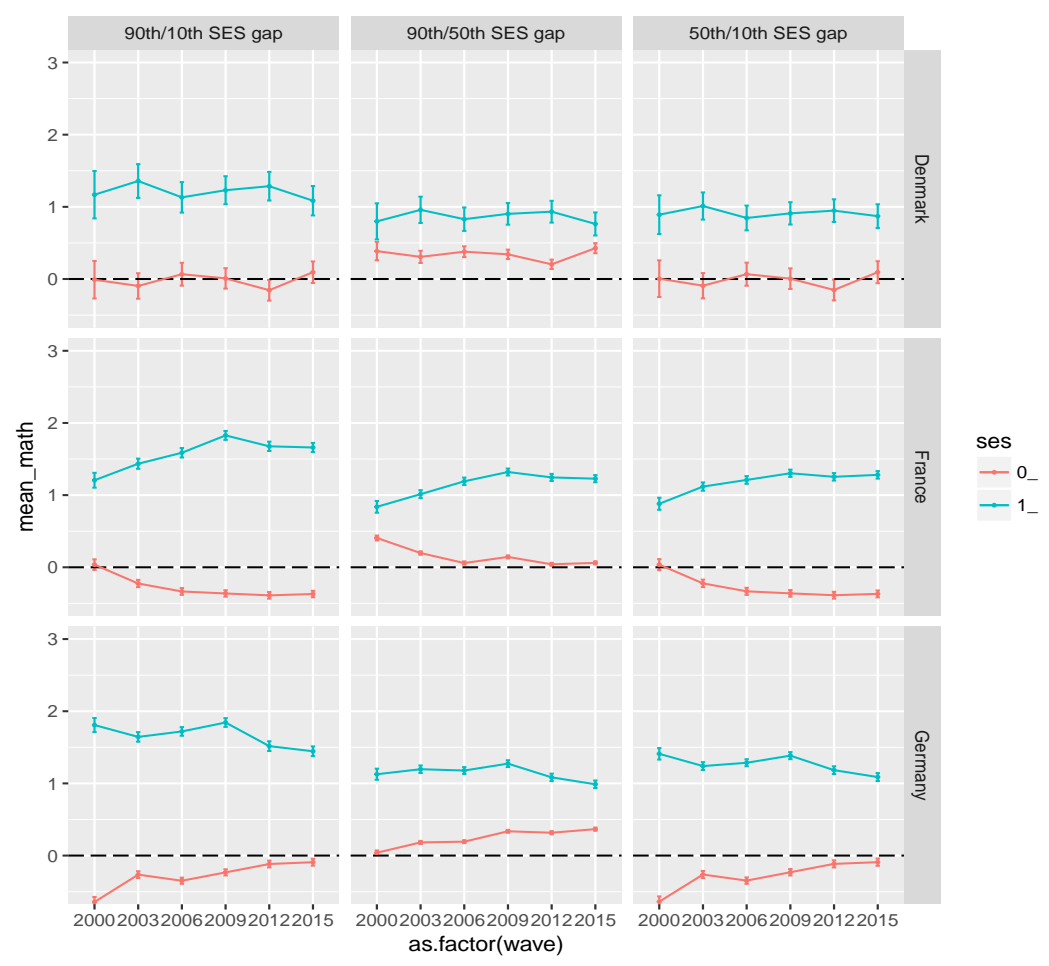
4.1 Descriptives

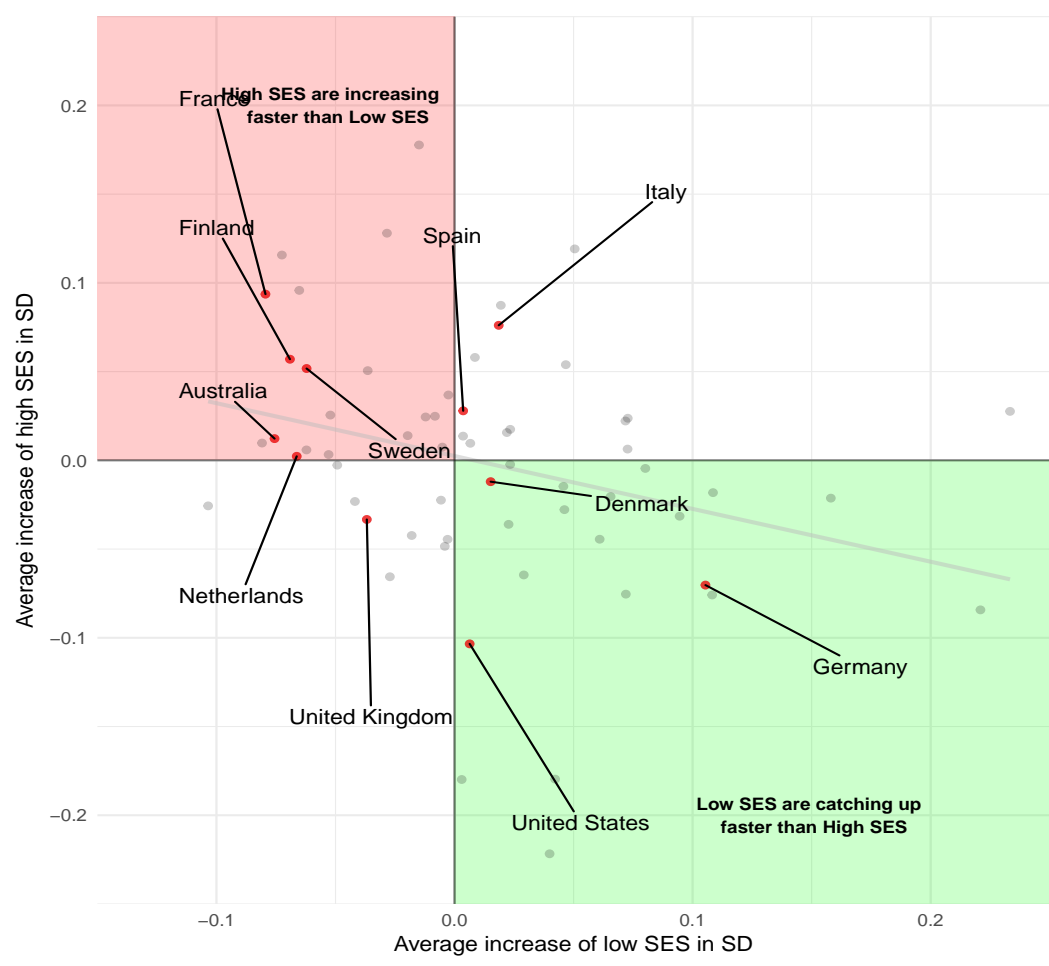
4.2 Results

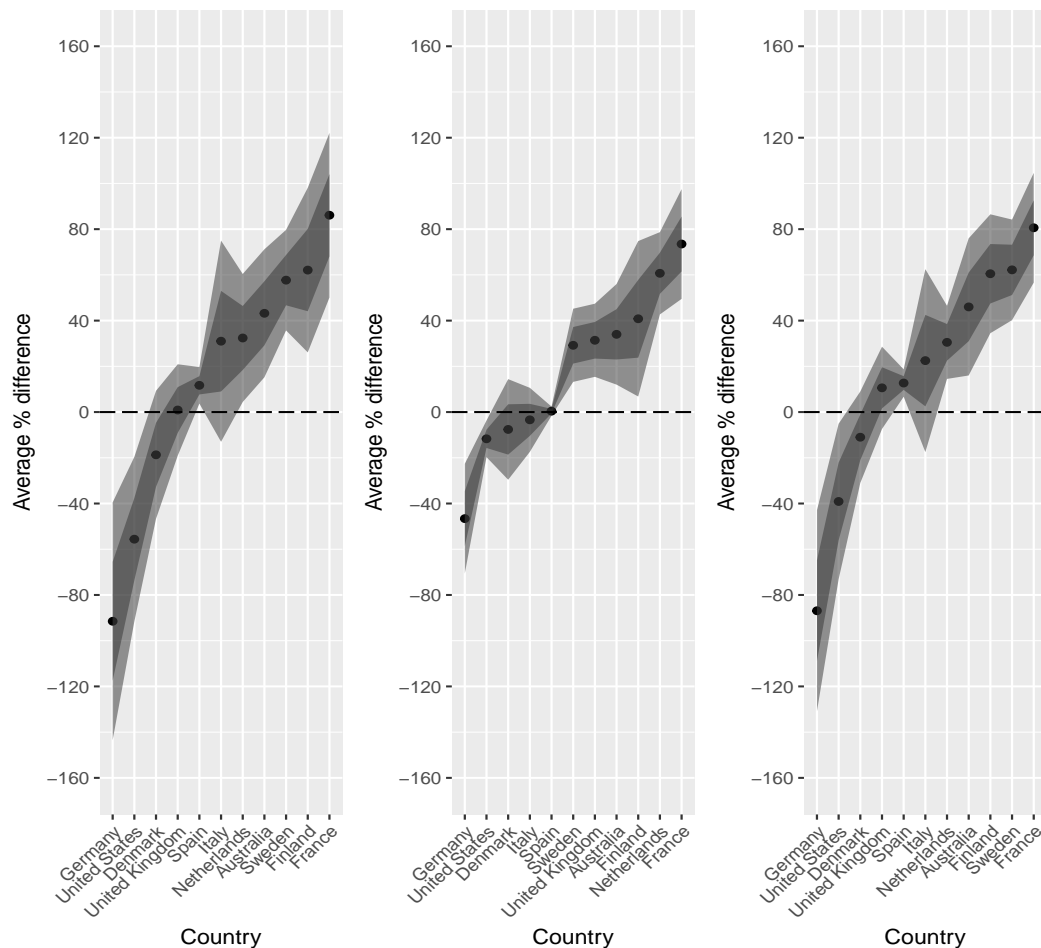












References

- Bradbury, B., Corak, M., Waldfogel, J., and Washbrook, E. (2015). *Too many children left behind: The US achievement gap in comparative perspective*. Russell Sage Foundation.
- Breen, R. and Jonsson, J. O. (2007). Explaining change in social fluidity: Educational equalization and educational expansion in twentieth-century sweden 1. *American Journal of Sociology*, 112(6):1775–1810.
- Duncan, G. J. and Magnuson, K. (2011). The nature and impact of early achievement skills, attention skills, and behavior problems. *Whither opportunity*, pages 47–70.
- Esping-Andersen, G. and Wagner, S. (2012). Asymmetries in the opportunity structure. intergenerational mobility trends in europe. *Research in Social Stratification and Mobility*, 30(4):473 – 487. Consequences of Economic Inequality.
- Reardon, S. F. (2011). The widening academic achievement gap between the rich and the poor: New evidence and possible explanations. *Whither opportunity*, pages 91–116.
- Reardon, S. F. and Portilla, X. A. (2015). Recent trends in socioeconomic and racial school readiness gaps at kindergarten entry. *Stanford, CA: Center for Education Policy Analysis*. Retrieved December, 20:2015.
- Shavit, Y. and Blossfeld, H.-P. (1993). *Persistent Inequality: Changing Educational Attainment in Thirteen Countries*. *Social Inequality Series*. ERIC.
- Van de Werfhorst, H. G. and Mijs, J. J. (2010). Achievement inequality and the institutional structure of educational systems: A comparative perspective. *Annual review of sociology*, 36:407–428.