Children's cognitive ability and the socioeconomic gradient:

changes over age in nine countries

John Jerrim and John Micklewright

1. Introduction

There has long been interest in how differences in child outcomes that are associated with parental background may grow as childhood progresses. Parents from higher socio-economic backgrounds with better levels of education and higher incomes may invest more time and goods into their children.[1] The resulting differences in outcomes, it is argued, emerge early at the pre-school level and then are re-enforced in childhood and the teenage years, despite the potential equalizing effect of compulsory education. The differences may be further compounded on entry to tertiary education and beyond into adult life. Various authors have produced evidence confirming this pattern. Leon Feinstein (2003) considered cognitive ability of British children at two points during the pre-school years and at ages five and ten. The difference in average percentile ranks between children from high and low socio-economic status, defined on the basis of parental occupation, widened from around 13 points at just under two years to nearly 30 points at age 10. Alissa Goodman et al. (2009) use more recent data to look at differences for English children, showing these to grow for much of childhood. In the US, James Heckman (2008) reports differences in average percentile rank in maths tests between children from the bottom and top family income quartiles, rising from about 14 points at age six to 23 points at age twelve.

Other chapters in this volume provide further evidence for one or more countries. Our contribution is to compare the socioeconomic gradient in children's cognitive ability at two ages for almost all of the countries covered by other chapter authors: USA, Canada, England, Scotland, Italy, France, Germany, Denmark and Sweden. Our measures of cognitive ability are from reading tests at age 10 shortly before the end of primary schooling and at 15 when all children are still in compulsory secondary schooling. We do not

have panel data like the authors just cited, so we are not considering the change in socio-economic gradient for the same children. But at each of the two ages, the data sources we use are specifically intended to allow comparison across countries since they are cross-national surveys with common sample designs and survey instruments. And, on the face of it, the two surveys have some socio-economic background variables in common, thus allowing the type of comparisons across two points during childhood that we wish to make.

We first consider how large the gradients are at the two ages and whether they differ significantly across countries. We then measure the changes in the gradients across the two ages within each country. Do they steepen and is the change significantly greater in some countries than in others? If the changes with age do differ across countries, we want to comment on why that may be, and especially whether the changes seem related to institutions that kick in between the two ages. One example of such institutional change is the sorting of children into different types of secondary school by ability level, known as 'tracking'. This takes place to varying degrees in a number of European countries, most notably Germany where the implications in terms of subsequent possibilities of entering higher education are most extreme.

This approach can be thought of as a 'difference-in-differences' analysis of the impact of institutions. Imagine one were to compare across countries the socioeconomic gradients in test scores at just one age, say 15, with the extent of tracking in each country. The unobservable societal factors that lead countries to use tracking may also steepen socioeconomic gradients in scores. The apparent effect of tracking seen when comparing countries could therefore be due to those unobservable factors. The difference-in-differences approach controls for this, by comparing the change in gradients between two ages with the change in institutions operating at these ages.

We are not the first people to think of this strategy; Andreas Ammermueller (2006) and Fabian Waldinger (2007) use some of the same data sources with similar objectives. However, there are important differences between these studies and our own. We should also acknowledge the work of Eric Hanushek and Ludger Woessmann (2006), who related the change in institutions between ages to the change in the variance of test scores, rather than to the change in the covariance of scores with socio-economic background, which is our focus. We consider the results of these studies later in the chapter.

The data we use are from the Progress in Reading Literacy Study (PIRLS) for 10 year olds and the Programme for International Student Assessment (PISA) for 15 year olds. We describe these data in the next section, focusing on the measurement of socio-economic background, where our choices are driven by the need for comparability across the two surveys. We then present results from simple regression models of reading test scores using each of the two surveys. We find that there are large differences in reading scores in all countries between children of low and high socio-economic background at both age 10 and age 15. Yet estimates of the changes in the socio-economic gradient between the two ages are not robust to our choice of model specification and hence we are unable to apply the difference-in-differences approach to assess the impact of institutions. In Section 7, we compare our findings on the socio-economic gradient with those from the papers just cited. A final section concludes.

2. Survey data on reading ability and socio-economic background

We want to compare socio-economic gradients in reading ability in two different surveys, PIRLS and PISA. We use PIRLS data for 2001 and PISA data for 2003.[2] The exception is for Denmark where we use PIRLS data for 2006, when the next round of the survey was conducted, as there are no Danish data for 2001. Sample size in PIRLS for our nine countries ranges from around 2,700 for Scotland to 8,200 for Canada, with a median of about 3,500. For PISA the extremes are again Scotland at 2,700 and Italy at 11,600, with a median of about 4,500. The Canadian sample in PIRLS refers to the provinces of Ontario and Quebec only and we therefore restrict the PISA sample for Canada to this basis as well (this is about a quarter of the total Canadian sample). Both PIRLS and PISA have two stage designs in which schools are sampled and then pupils within schools. In PIRLS a whole class is selected at random while in PISA a fixed number of pupils, 35, are sampled randomly within the school. We allow for this clustering in design in all our estimates of standard errors.

The first issue is the two surveys' assessment of reading ability. It is not clear how they really differ in approach. PISA emphasises measurement of skills that can be applied in real-life situations. The PIRLS organisers argue that the approaches in the two surveys are similar, both being based on 'an expanded notion

of literacy' (Jay Campbell et al. 2001: 85). We assume that the reading tests are sufficiently comparable for our purposes.[3] Among our nine countries, the mean scores at the two ages have low positive correlation and the standard deviations have low negative correlation (r = +0.19 and -0.29 respectively).[4] England, the USA, and Scotland (in that order) have the highest variance at age 10 and Germany, the USA, and Italy at age 15.

The measures of reading scores have a metric that appears superficially comparable across the surveys as scores are scaled in both PIRLS and PISA by the survey organisers to have a mean across all participating countries of 500 and a standard deviation of 100. But the pool of participating countries differs across the surveys. We transform the scores into a metric that is comparable between the surveys: national Z-scores. That is, in each survey we adjust an individual's reading test score by subtracting the country mean and dividing by the country standard deviation. Note that it is the *national* mean and standard deviation that are used in this standardisation: we strip out differences in the variance of scores across countries at each age. We then assess the extent to which socio-economic background is associated with score dispersion, where the extent of this dispersion is standardised to be the same everywhere.

The alternative is to leave these differences in national variances in the data, by adjusting the raw scores in each survey by the mean and standard deviation in the *pooled* sample of the nine countries that we study (each country's data are given an equal weight in the calculation of the pooled statistics). With this 'international' metric at each age, the standard deviation for Germany rises by a quarter between ages 10 and 15, and that for Italy by a tenth, while those for England and Scotland fall by about a tenth and fifth respectively. We return to consider this alternative metric later in the chapter.

The choice of socioeconomic background variables is a major issue. We need measures that are comparable across the two surveys. There seem to be only two candidates: the number of books in the home and parental education.[5] Parental education would normally be our first choice given the importance of human capital explanations of earnings, occupation and incomes and the ability to link the analysis to an extensive existing literature on intergenerational transmission of educational advantage. However, it poses two problems. First, parental education is missing for large numbers of children in PIRLS, in which information on this variable is collected through questionnaires administered to parents. No parental

questionnaires were issued in the USA (true of both the 2001 and 2006 survey rounds). Parental education is missing for about half of children in England and a third in Scotland and Germany, due mainly to no parental questionnaire being completed (again true in 2006). The data are not missing at random. In each case, average reading scores for children with missing parental data are, roughly speaking, about a third of a national standard deviation below the national mean. The missing data problem for parental education is much less in PISA (information was collected from the children) but where it occurs the test scores are also very different on average – between a half and one standard deviation below the mean. Second, we have concerns about the comparability of the data across countries and surveys. In both PIRLS and PISA, information on parental education is coded into levels of the International Standard Classification of Education (ISCED). This should provide a comparable variable but as we explain later we have some doubts.

We are drawn therefore to the number of books in the home. This is a standard background variable in international surveys of children's cognitive ability, including the well-known Trends in Maths and Science Study (TIMSS) as well as PIRLS and PISA. It has been used prominently by a variety of authors e.g. Andreas Ammermueller and Jörn-Steffen Pischke (2009), and Gabriela Schütz et al. (2008), as well as in the work by Ammermueller (2006) and Waldinger (2007) that we referred to in the Introduction. Results on the association of test scores and books in the home are drawn on in leading review articles by Stephen Machin (2009) and Eric Hanushek and Ludger Woessmann (2010). Schütz et al. give a spirited defence of the use of this variable as a measure of socio-economic background compared to parental education, and are echoed by Hanushek and Woessmann. They contend that a large number of books indicates a home environment that encourages academic effort and that it is a reasonable proxy of both the social and the economic background of the parents, arguments also made by Mariah Evans et al. (2010). Schütz et al. point to the far lower missing data problem in the international surveys for books in the home than for parental education, express concern over the true comparability across countries of the ISCED measure of education, and argue that books in the home is a 'considerably stronger' (2008: 286) predictor of student performance. Schütz et al. also compare the books variable in PIRLS with a household income variable, which is available for a few

countries. They interpret their results as supporting the validity of cross-country comparisons that use books in the home as an indicator of socio-economic background.

Information on books in the home is requested of the child in both PIRLS and PISA so the lack of a completed parental questionnaire does not generate missing values. On average across our nine countries, 97 percent of children respond to the question concerned. With the exception of an additional category at the top of the range in PISA, the variable has the same categories in the particular PIRLS and PISA survey rounds that we use.[6] On the face of it therefore, books in the home as reported by the child looks to be an attractive option. For the moment we take this variable at face value but we then return to re-consider its validity. To trail our findings: we are less sanguine than previous authors.

We begin by taking the child's report of books in the home as our sole measure of socio-economic background. We then compare the information on books collected from children with that collected from parents, before investigating the robustness of the results to using in addition some of the information on parental education.

3. Books in the home and test scores

The books variable has five categories, ranging from 0-10 books to over 200 – see Figure 1.[7] Excluding missing values, the proportion of the PIRLS sample in the lowest category ranges from 3 percent (Sweden) to 15 percent (Italy), and in PISA from 5 percent (Sweden) to 13 percent (Italy). The figures for the top category are much higher, ranging from 13 percent (Italy) to 33 percent (Sweden), and in PISA from 21 percent (France) to 36 percent (Sweden).

Figure 1 plots mean reading test scores at each age, in national z-scores, against the books variable categories. The data are presented as line graphs in order to fit all the information on the page but the distortion of the horizontal scale should be noted – neither graph implies anything near a linear relationship between the actual number of books and the test scores. Three features stand out. First, the variation in mean scores within each country at each age is large. Children reporting only 0-10 books in the home have reading scores on average about 0.75 of a standard deviation below the national mean in both surveys; those

reporting 200 or more books are on average around 0.3 (PIRLS) or 0.4 (PISA) standard deviations above the mean. Second, as these figures illustrate, the rise in scores with the number of books appears in general to be greater at age 15. The prima facie evidence therefore suggests a *steepening* in the socio-economic gradient. Third, and perhaps most striking, the steeper profile at 15 seems to result from a strengthening in the association of books in the home with reading scores at the top of the books distribution, a finding not uncovered in previous studies. Mean reading scores for 10 year olds typically fall slightly between the categories 100-200 and over 200 books. But the same is clearly not true for the 15 year olds. The suggestion is that between ages 10 and 15, the children from the most advantaged backgrounds pull away and it is notable that this appears to be the case in all nine countries. Fourth, and following on from this, the similarities in the patterns across countries appear at first sight more obvious than the differences.

< Figure 1 about here >

We now investigate these patterns in more detail in a simple regression framework (Model 1). For each country we estimate two regressions, one for each age group, in which the dependent variable is the test score (measured in national z-scores) and the explanatory variables are:

books categories (dummies, base 11-25)
language spoken at home is not that of the test (dummy)
interactions of the books and language dummies
age (in months)
gender (dummy, base male)

The regression model therefore contains only very limited controls. (We also include dummies for missing values for books in the home, language spoken at home, and interactions of missing language with the books dummies.) We take the base for the books dummies as 11-25 books in preference to the bottom category of 0-10 books since the latter is very sparsely populated in some countries and may not provide the most useful contrast with the quite well populated top category. Table 1 reports for each age group the estimated

coefficients on the dummy for the top category of books, over 200, together with the difference between the estimates (age 15 coefficient minus age 10 coefficient), which is the basis for the ordering of the countries. An on-line appendix to accompany this chapter gives the full results of each regression model.

< Table 1 about here >

In line with Figure 1, the average reading test score of children reporting over 200 books in the home differs from that in the base category of 11-25 books by between a quarter and three-quarters of a national standard deviation at age 10 (average 0.54). In all countries, this difference is larger at age 15 (average 0.82), indicating a rise in the socio-economic gradient when measured in this way. In all but Canada, the change between ages 10 and 15 is by over 0.2 of a standard deviation – not insubstantial – and is significantly different from zero at the 5 percent level or better (t-statistics vary from 2.4 to 5.0). In three countries the differences are about 0.4 of a standard deviation. Although this represents a substantial variation across countries, the hypothesis that the increase is the same in Italy or the USA at one end of the range and Denmark and Germany at the other can only just be rejected at the 5 percent level.[8] Finally, reflecting Figure 1, the increase in socio-economic gradient between ages 10 and 15 is driven almost entirely by the steepening between the top two categories of books.

4. Doubts over the child's report of books in the home

We have noted that the number of books in the home reported by the child is a standard measure of socio-economic background collected in international surveys. Despite the arguments which have been made in its favour, we feel some unease with this variable, especially over its reporting by young children. Children who respond to PIRLS and PISA fill in self-completion questionnaires about their home background. The 15 year-olds in PISA are told 'there are usually about 40 books per metre of shelving' before being asked to tick one box indicating the number of books in their home. The 10 year-olds in PIRLS are given more guidance. A description in words and a diagram accompanies each category of books. For example, 26-100 books is described as 'enough to fill one bookcase' and the diagram shows a full bookcase with four shelves (11-25

books is described as enough to fill one shelf). Despite this guidance, the potential for measurement error seems obvious.

In particular, our interest focuses on the reporting of the top two categories of books in PIRLS. We have noted that it is the lack of difference in average reading test scores between these two groups, shown clearly in Figure 1, that drives the conclusion from Table 1 that the gradient of test scores with books in the home rises between ages 10 and 15.

We investigate the validity of the children's reports in PIRLS in two ways. First, where a parental questionnaire was completed, we can compare the report by the children with information given by the parents since the latter were also asked to report the number of books in the home. We have noted the major problem with missing data from the parental interview in PIRLS in some countries: no data for the USA and very large numbers missing in England, Scotland and Germany. But where information from both children and parents are present a comparison can be made. Note that children and parents are asked slightly different questions: the children are asked not to include 'your school books' while parents are asked to exclude 'children's books' and then are asked a separate question about these. The same categories of books are given in the parental and child questionnaires. We compare the distribution of books reported by children and parents, restricting attention for the latter to the question excluding children's books. We pool data for the eight countries for which there is any parental data. There is a reasonable degree of agreement between the two variables but it is far from complete. In 40 percent of cases where the child reported more than 200 books, the parent indicated a lower category. Where the child reported 100-200 books, the parent reported the same category in only a quarter of cases with roughly equal numbers reporting over 200 and reporting a lower category. The rank correlation coefficient for the two variables is only 0.49. Figures for individual countries vary from 0.39 for Scotland to 0.54 for France.

We then calculate average reading tests scores for the children at age 10 in each category of books as reported by the parents – see the series for 'Parent reports' in Figure 2, which shows averages across countries. The graph also gives the averages in each category reported by the children, restricting the analysis to the *same* children for whom there is information on books reported by the parents e.g. only about half of

children in the UK. This is the series 'Child reports (subset)'. For comparison, we also show the figures for all children, including those for whom there are no parental reports of books: 'Child reports (all)'. The contrast in pattern between the parent and child reports is striking. There is no flattening out or decline in average scores between 100-200 books and over 200 books in the case of the parents' reports, while average child test scores in the lowest category of books reported by the parents is substantially above that for children in the lowest category of their own reports.

< Figure 2 about here >

Our second validity check is to compare the children's reports of books to the parents' reports of household income, which are available for certain countries in PIRLS. Is income significantly higher on average in households where over 200 books are reported than for households in the category 100-200 books? Among our nine countries, there are income data in PIRLS for Canada, England, Germany, and Sweden. The measurement of household income is certainly not ideal. The information is collected in banded form with a single question on all forms of income to one parent, typically the mother. This method of collection is likely to result in substantial measurement error (John Micklewright and Sylke Schnepf 2010). There is also considerable item non-response. The income levels indicated by each band differ from country to country and we estimate for each country a censored regression model of (log) income on dummies for each category of books (we use ordinal probit for Canada where we have been unable to establish the boundaries of the income bands). We estimate two models, one using the child's report of books and one using the parent's report. In all four countries we easily reject the null of no difference in average income between the top two books categories when using the parent's report. But we fail to reject the null at the 5 percent level in England and Canada when using the child's report and in both Germany and Sweden the estimated difference in income between the two categories is substantially reduced compared to results with the parent's report (from 18 percent to 9 percent and from 19 percent to 12 percent respectively). We view this as further evidence that the children's reporting of books in PIRLS towards the top of the distribution may be prone to error, calling into question the apparent flattening out in average test scores that drives the results for change over age in Table 1.[9]

We then estimate the regressions for test scores at ages 10 and 15 using the parents' reports of books rather than the children's reports, including dummy variables for missing values where parental reports are missing. This is Model 2. Another change is that we enter the variable for the parents' reports as a continuous variable, rather than as a series of dummies. This imposes the constraint that the marginal increase in the reading test score is constant between each level. The levels represent greatly varying ranges of book numbers, so this is quite different to assuming that the marginal increase in score with a given number of books is constant, which the data show clearly not to be the case. The constraint imposed is suggested by the pooled sample average test scores for PIRLS in Figure 2 (parent reports), and by the figures for PISA in Figure 1. We test the constraint formally. In no country for PIRLS do we reject the null of a constant marginal effect at the 5 percent level. In PISA we do reject at this level in Sweden, Denmark and France but Figure 1 does not suggest that we do the data a great disservice. We again include the controls present in Model 1, e.g. language in the home. A dummy variable is included for missing parental reports of books in the home in PIRLS: in England, Scotland and Germany we have around a third to a half of the sample with no parental information on books.

Table 2 reports the key result from these regressions, the estimated coefficient for each country on the categorical books variable at each age, 10 and 15, and its change between the two ages. There are results for the USA only at age 15. To aid comparison with Table 1 we have multiplied the coefficients and their standard errors by three, thus showing the estimated increase in score with a three category change in the books variable. In some cases there are marked changes from Table 1. Italy moves from being the country with the largest change in gradient between the two ages in Table 1 to being the country with the smallest change in Table 2 – a change essentially of zero. In only four of the eight countries is the change significantly different from zero and the average change is substantially lower than in Table 1. Germany, the classic case of tracking by ability at secondary age, is one of the countries where the change is insignificant. Germany has one of the steeper gradients at age 10, significantly greater than that in Canada, Sweden, Denmark, and England. But at age 15, the gradient in Germany is not exceptionally steep. It should be remembered that we have removed from the data the differences in the national variances in reading scores at

each age, a point that we will return to later and one that is particularly relevant for Germany. England and Scotland show the largest changes in the gradient in Table 2 between the two ages, about a quarter of a national standard deviation.

< Table 2 about here >

5. Using information on parental education

We now turn to the other measure of socio-economic background available at both 10 and 15, parental education. In principle, the data are comparable across PIRLS and PISA as well as comparable across countries within each survey. Information for each parent on the highest level of education attained is recorded in both surveys with the ISCED classification, albeit with some small differences in aggregation of the detailed categories. The information is obtained from parents in PIRLS and from children in PISA. This means that at age 10 there is no parental education recorded for the USA and a major missing data problem for several other countries. Among the latter, missing data are much more common for children with lower reading scores. Where present, we take whichever value of parental education is higher, the father's or the mother's.

The distribution of children across categories of parental education displays some large differences between the two surveys. For example, the share of the low education group, ISCED 0-2 (compulsory education only or 'initial' vocational education), is much higher in England, Scotland, and France in PIRLS than in PISA.[10] In the case of England and Scotland, this is despite the substantial non-response to the parental questionnaire being more common in PIRLS for children with lower reading scores, which should have the opposite effect. One possibility is that the difference may reflect the reporting of the information in PIRLS by the parents and in PISA by the children, with the 15 year old children of lower educated parents overestimating their parents' education. Another possibility is that the translation of national educational systems into ISCED categories differs between PIRLS and PISA. Hilary Steedman and Steven McIntosh (2001, footnote 3) report that 'there is disagreement between international organisations about the allocation

of different education and training outcomes to the categories of the ISCED scale' and go on to use the UK as an example.

We therefore include information on parental education into regressions of the reading test scores with some caution. We adopt a conservative approach, using just one dummy variable for a highly educated parent: those in ISCED levels 5A and 6 (university or college education). This is Model 3. The rest of the specification is as for Model 2 (except that we interact the language dummies with the parental education dummy rather than the books categories). As for parental reports of books in the home, in England, Scotland and Germany this means that we have a substantial fraction of the sample at age 10 with no information on parental education and the USA again has to be excluded from the age 10 regressions. The parental education dummy has a powerful and well-determined effect in the regressions: holding constant books in the home and the other controls, tests scores for children with at least one college or university educated parent averages 0.35 of a national standard deviation higher at age 10 across the eight countries and 0.26 higher at age 15. These figures are roughly double that for one category of the books variable at age 10 and equal to it at age 15. In this rather narrow sense, parental education is 'more important' than books at age 10 and 'as important' at age 15.

We use the regression results to predict at each age the difference in reading scores associated with an increase in books in the home of three categories plus a change in parental education from neither parent having university or college-level education (ISCED 5A/6) to having at least one parent educated at this level – see Table 3. The differences average 0.88 of a national standard deviation at age 10 and only a little higher, 0.95, at age 15. In only two countries, Scotland and England, is the change in the difference between the two ages as much as 0.2 of a standard deviation and in neither country is this change very well determined. The change is small, whether positive or negative, and insignificant (less than a standard error) in five countries.

< Table 3 about here >

Figure 3 plots the 'change' figures from Tables 2 and 3 against one another. Including parental education, albeit in a very limited way, produces a substantial difference in the picture for Canada and Sweden, while for several countries the results are similar or very similar.

< Figure 3 about here >

6. Summary

Taking together Tables 1-3, based on our three different regression models, what have we found? First, using *child reports of books in the home* alone as the measure of socio-economic background, there are statistically significant increases in gradients between ages 10 and 15 in all countries except Canada (Table 1). In these eight countries, the rise in test scores when moving from a low category (11-25 books) to the top category (200+) increases between the two ages by between about 0.2 and 0.5 of a national standard deviation. This range is quite large but the standard errors around the figures are such as to preclude most firm statements about the variation across countries.

Second, when switching to *parent reports of books in the home* at age 10 (and living with the large number of missing values in some countries), the picture is somewhat less clear (Table 2). Moving from the same low category of books to the top category, and now constraining the marginal change in score across categories to be constant, leads to smaller increases in the gradients than before.

Third, adding a second measure of socio-economic background, information on *parental college or university completion* (for which there is also a substantial missing value problem), has little effect on the results for most countries but reduces again the number of countries where the increase in the apparent 'effect' of socio-economic background is positive and significant (Table 3).

As far as individual countries are concerned, it is not always easy to summarise the picture. Across the three sets of results (i) there is little evidence of a significant rise in socio-economic gradient for Canada (barely significant at the 5 percent level in Table 2 only); (ii) not much evidence for a rise in Germany (statistically significant in Table 1 only) or in Italy (strongly significant rise in Table 1 only); (iii) significant rises in each table of results for England and Scotland (caution is needed in Tables 2 and 3 given the extent of missing data) and, rather less obviously, Denmark. The absence of a clear and sharp rise in Germany, with its system of secondary school tracking by ability, is notable.

All our results have been obtained using test scores that have been transformed into *national* z-scores. We have removed from the data the differences across countries in the national variances at each age. How would our results change if we left those differences in the data, and used the alternative approach we mentioned in Section 2 of adjusting the recorded scores in PIRLS and PISA by the mean and standard deviation at each age in the *pooled* sample of the nine countries that we study? With this 'international' metric the mean score is equal to zero and the standard deviation is equal to one in the pooled sample at each age. But the means and standard deviations at each age vary across countries.

We noted in Section 2 that using this international metric for reading scores at each age, the standard deviation for Germany rises considerably between ages 10 and 15 while those for England and Scotland fall. To be clear: the reading tests at ages 10 and 15 are different tests so these changes in dispersion are not measured with a single 'absolute' metric for reading ability that is the same at the two ages. The changes in the standard deviations imply only that *relative to other countries* the dispersion of scores in Germany is higher at age 15 than it is at age 10 while the opposite is true for England and Scotland.

To recover key results for scores based on this international metric at each age is simple. For a given country, we need only to multiply our regression slope coefficients estimated with data transformed into national z-scores, by the country's standard deviation in the data re-scaled using the international metric at the age in question. Figure 4 shows how use of this alternative metric changes the picture obtained of change in socio-economic gradient between ages 10 and 15. The vertical axis measures the changes shown in Table 2, based on national z-scores. The horizontal axis measures changes using the international metric. The rise in score dispersion for Germany using the international metric between ages 10 and 15 has the effect of substantially increasing the change in the socio-economic gradient. On the other hand, the fall in dispersion for England and Scotland between the two ages has the effect of dampening down the change in gradient. Which is the more appropriate measure to use, a national or international z-score transformation? Arguments can be made both ways. Our results in Tables 1-3 based on national z-scores focus on the relationship between socio-economic background and scores when score dispersion is set to be the same in each country at each age.

< Figure 4 about here >

7. Comparison with other authors' results

In the Introduction we mentioned work of other authors who have used the international surveys PIRLS, PISA and TIMSS to try to shed light on the role of institutions in determining how the inequality of children's learning achievement changes with age. How do their results compare to ours?

In a widely cited paper, Hanushek and Woessmann (2006) used published results from all three surveys to argue that early tracking of children into different types of schools by level of ability increases inequality in achievement. They focused on the change in the variance of test scores between primary and secondary schooling, rather than in the covariance of scores with socio-economic characteristics. However, they noted that 'one channel for increasing inequality is re-enforcing the effects of family background' and argued that this would be a valuable direction for further research.[11]

Subsequent authors have taken up this challenge. Ammermueller (2006) used data on reading scores from PIRLS 2001 for 10 year olds and PISA data for 2000 for 15 year olds to estimate regression models that are similar to ours for 14 countries, including six covered in our study. The main family background variables were the children's reports of books in the home (in both surveys), parental education, and parental attitudes to their children's schooling. The substantial problem of missing data, especially in parental education and attitudes at age 10, was carefully documented and was addressed by imputing values.[12] Ammermueller concludes that the impact of family background increases between primary and secondary education 'in almost all countries' and that social origin becomes more important with age in countries with 'a differentiated schooling system with various school types and a large private school sector' (p27). Among other differences from Ammermueller's study, we have not included the parental attitude variables in our analysis since we wish to capture to the full association of test scores with socio-economic background, including that coming through attitudes to education. Nor does it seem to us that the attitude variables are comparable across the surveys.

Waldinger (2007) questions the earlier studies, arguing that his results 'cast serious doubt' on conclusions that tracking increases the importance of family background on learning achievement**.** He uses data from PIRLS, PISA and TIMSS and reports that 'slight changes' in the measurement of the extent of tracking, in choice of samples, and in specification of regression models 'renders the [earlier authors'] results... insignificant' (p.6). Two notable differences between Waldinger's analysis and that of Ammermueller, which Waldinger does not highlight, are that he uses the parental reports on books in the home in PIRLS rather than the child reports (the large problem of missing values is not mentioned) while he does not use information on parental attitudes to education. Like is not being compared with like. A similar conclusion of a lack of robustness and no apparent effect of tracking on inequality in achievement is reached by Maciej Jakubowski (2010). He focuses on a re-examination of the Hanushek and Woessman study, using the survey microdata to define samples that he argues are more comparable than those used to generate the published figures which formed the basis of the earlier analysis.

Our own analysis has looked at a smaller set of countries than any of these other authors. We have used data from PIRLS and PISA but not TIMSS. The focus has also been rather different, as we have concentrated on the measurement of socio-economic background rather than on measures of institutions such as the degree and strength of tracking by ability. We have transformed the score data into a different metric. That said, we think our results are more in line with the questioning stance of Waldfinger and of Jakubowski than with the earlier results of Hanushek and Woessmann and of Ammermueller.

8. Conclusions

In this chapter we have considered socio-economic gradients in reading scores at ages 10 and 15 for most of the countries considered in other chapters of this book. In all countries we find large differences in reading scores at both ages between children of low and high socio-economic positions. We used measures of reading scores that were standardised to have the same dispersion in all countries at each age. The results we have obtained are not very robust to changes in the specification of the simple regression models that we use

to explore the socio-economic gradients. Among the nine countries we consider, we vary from concluding that there are significant increases in the gradient between ages 10 and 15 in eight of them to just two.

In this situation it is clearly difficult to draw any conclusions that might inform policy. Some earlier authors, who have conducted similar cross-country analyses to ours, have concluded that their results show differences between countries in the policy of tracking students into different secondary schools according to ability level. Specifically, they suggest that this leads to a strengthening in the socio-economic gradient of test scores between primary and secondary school ages. Our own results do not support this conclusion.

We have been limited in the measures of family background available to us, given our need for measures that are comparable across the age 10 and age 15 surveys that we have analysed. Each survey alone provides a number of measures but the common set is very small – essentially books in the home and parental education – and we have emphasised problems with these variables. The international surveys of children's achievement require good measures of socio-economic status, and preferably measures that are in common, if the data are to be used to provide good estimates of how the socio-economic gradient varies across countries and across childhood.

References

Andreas Ammermueller (2006) 'Educational Opportunities and the Role of Institutions' ZEW Discussion Paper No. 05-44, Mannheim.

Ammermueller Andreas and Jörn-Steffen Pischke (2009) 'Peer effects in European Primary Schools: Evidence from the Progress in International Reading Literacy Study' *Journal of Labor* Economics 27(3): 315-48.

Giorgina Brown, John Micklewright, Sylke Schnepf and Robert Waldmann (2007) 'International Surveys of Educational Achievement: How Robust are the Findings?' *Journal of the Royal Statistical Society* Series A, 170 (3): 623-46.

Anders Björklund and Kjell Savanes (2010) 'Education and Family Background: Mechanisms and Policies' IZA Discussion Paper 5002, Bonn (forthcoming, *Handbook of the Economics of Education*).

Jay Campbell, Dana Kelly, Ina Mullis, Michael Martin, and Marian Sainsbury (2001) *Framework and Specifications for PIRLS Assessment 2001*, 2nd edn. Chestnut Hill: Boston College.

Mariah Evans, Jonathan Kelley, Joanna Sikora, and Donald Treiman (2010) 'Family Scholarly Culture and Educational Success: Books and Schooling in 27 Nations' *Research in Social Stratification and Mobility* 28 (2): 171-97.

Leon Feinstein (2003) 'Inequality in the Early Cognitive Development of British Children in the 1970 Cohort' *Economica* 70, 73–97.

Alissa Goodman, Luke Sibieta, Elizabeth Washbook (2009) 'Inequalities in educational outcomes among children aged 3 to 16' report for the National Equality Panel.

Eric Hanushek and Ludger Woessmann (2006) 'Does Educational Tracking Affect Performance and Inequality? Differences in Differences Evidence across Countries' *Economic Journal* 116(510): C63-C76.

Eric Hanushek and Ludger Woessmann (2010) 'The Economics of International Differences in Educational Achievement' IZA Discussion Paper 4925, Bonn (forthcoming, *Handbook of the Economics of Education*).

Robert Haveman and Barbara Wolfe (1995), 'The Determinants of Children's Attainments: A Review of Methods and Findings' *Journal of Economic Literature*, 33(4): 1829-1878.

James Heckman (2008) 'Schools, Skills, and Synapses' IZA Discussion Paper 3515, Bonn.

Maciej Jakubowski (2010) 'Institutional tracking and achievement growth: exploring difference-in-differences approach to PIRLS, TIMSS and PISA data' in J. Dronkers (ed.) *Quality and Inequality of Education: Cross-National Perspectives*, Springer.

Stephen Machin, (2009) 'Inequality and education' in W. Salverda, B. Nolan, and T. Smeeding (eds.) *The Oxford Handbook of Economic Inequality*. Oxford: Oxford University Press.

John Micklewright and Sylke Schnepf (2010) 'How reliable are income data collected with a single question?' *Journal of the Royal Statistical Society* Series A, 173 (2): 409-29.

OECD (2001), *Knowledge and Skills for Life – First results from PISA 2000*, OECD, Paris.

Gabriela Schütz, Heinrich Ursprung and Ludger Wössmann (2008) 'Education policy and equality of opportunity' *Kyklos*, 61 (2): 279–308.

Hilary Steedman and Steven McIntosh (2001) 'Measuring low skills in Europe: how useful is the ISCED framework?' *Oxford Economic Papers*, 53(3): 564-81.

Fabian Waldinger (2007) 'Does Ability Tracking Exacerbate the Role of Family Background for Students' Test Scores?', unpublished manuscript, London School of Economics.

Table 1. Differences in predicted reading ability between children in households with 11-25 books in the home and over 200 books in the home (books reported by children at both ages)

	Age 10		Age 15		Change	
Country	**Diff.**	**S.E.**	**Diff.**	**S.E.**	**in Diff**	**S.E.**
Canada	0.575	0.062	0.691	0.065	0.116	0.090
Denmark	0.519	0.067	0.729	0.056	0.210	0.088
Germany	0.656	0.047	0.867	0.054	0.211	0.072
Sweden	0.631	0.058	0.849	0.064	0.218	0.086
England	0.703	0.056	0.929	0.066	0.225	0.087
France	0.599	0.056	0.841	0.063	0.242	0.085
Scotland	0.594	0.073	0.970	0.058	0.377	0.093
USA	0.412	0.075	0.870	0.051	0.459	0.091
Italy	0.208	0.067	0.675	0.069	0.467	0.096
average	0.544		0.825		0.281	

Note. 'Diff.' stands for difference and 'S.E.' for the standard error of this difference. Results are based on an OLS regression using dummy variables indicating the number of books in the home. The specification is described in more detail in the text (Model 1). Reading ability is measured in national z-scores.

Table 2. Differences in predicted reading ability associated with a three category increase in the number of books in the home (books reported by parents at age 10 and children at age 15)

	Age 10		Age 15		Change	
Country	**Diff.**	**S.E.**	**Diff.**	**S.E.**	**in Diff.**	**S.E.**
Italy	0.741	0.048	0.726	0.057	-0.015	0.072
Germany	0.774	0.039	0.837	0.045	0.063	0.057
France	0.777	0.048	0.885	0.051	0.111	0.069
Canada	0.555	0.051	0.690	0.048	0.135	0.069
Sweden	0.618	0.057	0.786	0.039	0.168	0.069
Denmark	0.558	0.045	0.771	0.045	0.213	0.063
England	0.606	0.069	0.837	0.051	0.231	0.087
Scotland	0.672	0.069	0.924	0.042	0.252	0.078
USA	-	-	0.854	0.013	-	-
average	0.663		0.807		0.145	

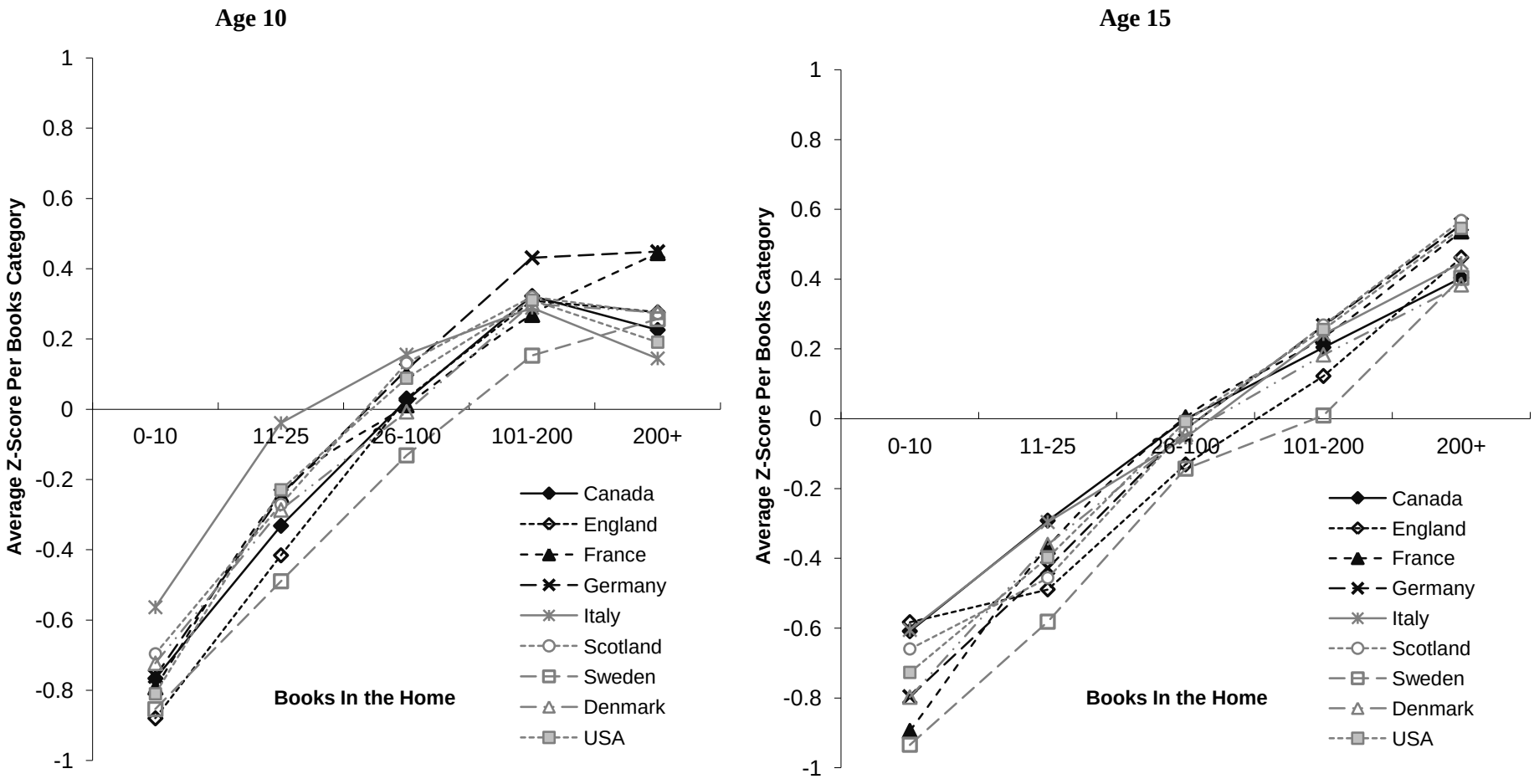
Note. The average at age 15 does not include the difference for the USA. 'Diff.' stands for difference and 'S.E.' for the standard error of this difference. Results based on an OLS regression using a continuous variable with five values indicating numbers of books in the home. The specification is described in more detail in the text (Model 2). Reading ability is measured in national z-scores.

Table 3. Differences in predicted reading ability associated with a three category increase in the number of books in the home and a change to a parent having college or university education (books and parental education reported by parents at age 10 and by children at age 15).

	Age 10		Age 15		Change	
Country	**Diff.**	**S.E.**	**Diff.**	**S.E.**	**in Diff.**	**S.E.**
Canada	0.863	0.057	0.807	0.055	-0.056	0.079
Sweden	0.845	0.062	0.791	0.047	-0.054	0.078
Italy	0.927	0.056	0.919	0.070	-0.008	0.089
France	0.977	0.054	1.003	0.061	0.026	0.081
Germany	0.950	0.048	0.998	0.050	0.048	0.069
Denmark	0.762	0.061	0.907	0.056	0.145	0.083
Scotland	0.860	0.084	1.055	0.053	0.195	0.100
England	0.860	0.087	1.085	0.057	0.225	0.104
USA	-	-	0.999	0.046	-	-
average	0.881		0.946		0.065	

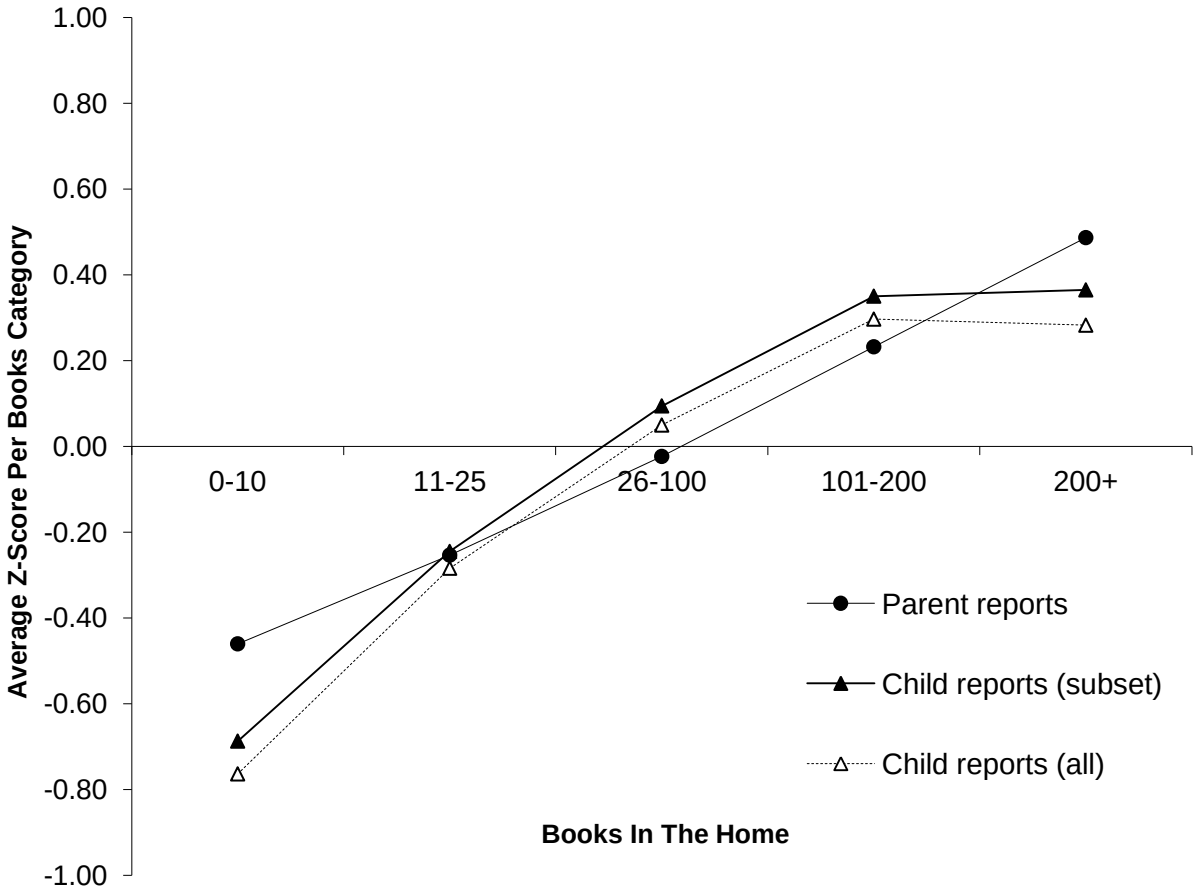
Note. The average at age 15 does not include the difference for the USA. 'Diff.' stands for difference and 'S.E.' for the standard error of this difference. Results based on an OLS regression using a continuous variable with five values indicating numbers of books in the home and a dummy variable for at least one parent having college or university education. The specification is described in more detail in the text (Model 3). Reading ability is measured in national z-scores.

Figure 1. Average reading score (measured in national z-scores) by categories of books in the home reported by the child



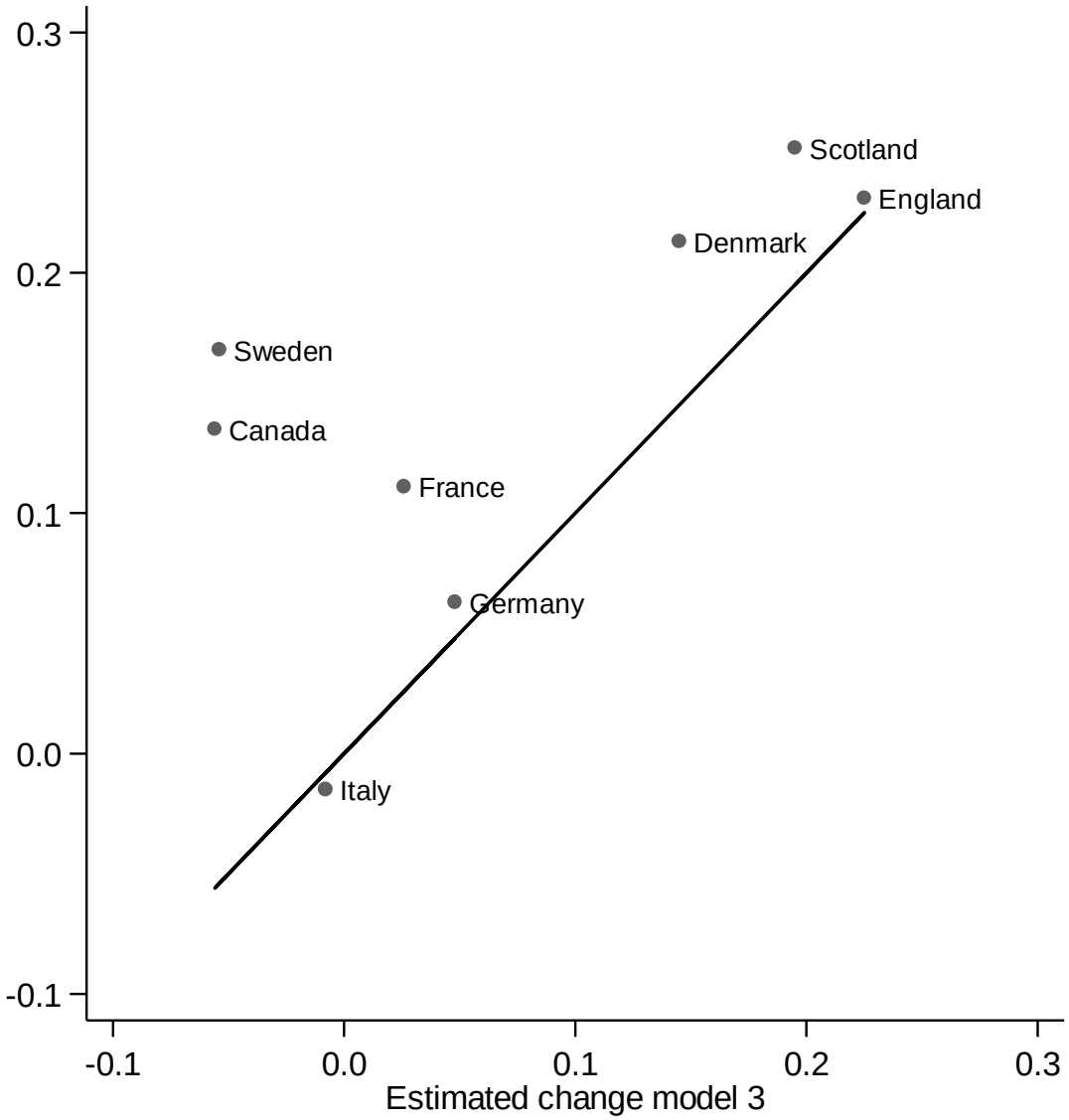
Note. See the text for further explanation.

Figure 2. Average reading score (measured in national z-scores) at age 10 by categories of books in the home: child reports of books vs. parent reports (averages of country figures)



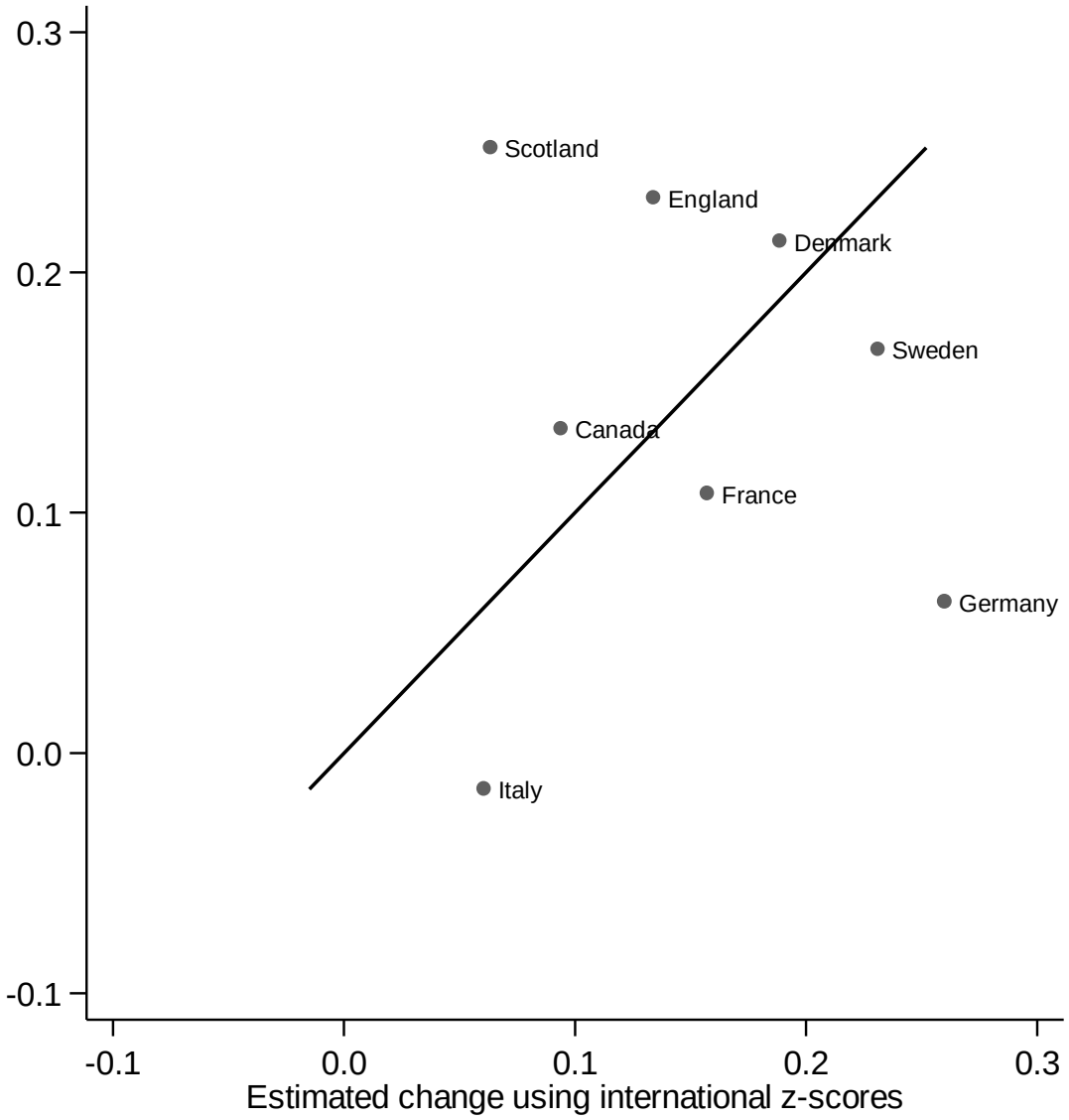
Notes: Unweighted averages of figures for each country excluding the USA. The solid line for children refers to those children with parents who also report books at home. The dotted line refers to all children including those with no data on books at home reported by the parents.

Figure 3. How the inclusion of parental education alters the results: a comparison of results based on Models 2 and 3



Note. The graph plots the 'Change in Diff.' figures from Table 2 and Table 3 on the vertical and horizontal axis respectively.

Figure 4. How a switch in reading score metric alters results: a comparison of results based on Model 2 when measuring the test score with international rather than national standard deviations



Note: The graph plots the 'Change in Diff.' figures from Table 2 on the vertical axis. These figures are based on Model 2 and use reading scores measured in national z-scores. The horizontal axis shows the analogous figures for the 'change in difference' in socio-economic gradient between ages 10 and 15 figures based on Model 2 when the reading test scores at both ages are measured in the alternative metric of international standard deviations. Further explanation is given in the text. The 45 degree line shows where there is no difference in results.

[1] Robert Haveman and Barbara Wolfe (1995) provide a simple framework for considering these investments. Their framework underlines that the apparent effect of parental education and income in part reflects unobserved ability inherited by children. See also the review in Anders Björklund and Kjell Savanes (2010).

[2] PIRLS assessed children in the upper of the two grades with the most 9-year olds at the time of testing. This corresponds to the fourth grade and an average age of just over 10 years for most of our countries. PISA assesses 15 year olds irrespective of their grade. PISA also measures ability in maths and science. Reading was a 'minor' subject in 2003, implying less question time devoted to testing, but we nevertheless choose this year so as to maximise comparability of socio-economic measures between PISA and PIRLS.

[3] In both surveys, children's answers to the reading test questions are summarized by the survey organizers into a score using an 'item-response model'. The intuition is that true reading ability is unobserved, and must be estimated from the answers to the test. Five 'plausible values' are generated for each individual, each estimating the individual's true proficiency. Except where indicated, we estimate parameters (means, standard deviations, or regression coefficients) and their standard errors with each plausible value and then average the five estimates. We also use survey weights. The 2003 PISA data contain plausible values of reading scores for all responding children but there is some question in our minds over whether all children actually took the assessment in reading, given its status as a minor subject that year, or whether the recorded scores for some children are just estimates from the item response model. We have checked results for the sub-sample of children that we are fairly sure took the reading test and they are similar to those obtained with the whole sample.

[4] Giorgina Brown et al. (2007) provide more systematic comparison of central tendency and dispersion of scores in PIRLS, PISA, TIMSS and IALS.

[5] There are also variables measuring migrant status and language spoken in the home that seem comparable, but we see these (and use the latter) as control variables and not measures of socio-economic background as such. Both surveys contain measures of parental occupation but our assessment is that the coding systems are insufficiently comparable for our purposes.

[6] By contrast, the categories differs somewhat in PISA in 2000, when reading was the 'major' subject under investigation, which is the reason why we do not follow Ammermueller (2006) in using the data from this survey round.

[7] There is an additional closed-interval category at the top of the distribution in PISA, 201-500 books, which we combine with the 500+ category.

[8] Were we to take the sparsely populated 0-10 books category as the base, the rises in gradient between the two ages would appear *smaller*, this group being further adrift at age 10 from the 11-25 category than at age 15.

[9] There are income data reported by parents in the PISA survey round for Germany in 2006 when parental questionnaires were introduced for a small number of countries. We estimated a censored regression model for the data and found a substantially steeper gradient with children's reports of books in the home than in the PIRLS data for 2001.

[10] Taking only those children for whom there is information on parental education recorded, the PIRLS shares are higher by around 30 percentage points in England, 20 points in France and 15 points in Scotland. The differences between the distributions for PIRLS 2001 and PISA 2000 are documented in Ammermueller (2005, Table A3).

[11] Consider the following simple model for achievement, A, as a function of family socio-economic background, S, and unobserved factors, u: $A = \alpha + \beta.S + u$. The model implies that $var(A) = \beta^2.var(S) + var(u)$. One driver for the change in inequality of achievement, measured by var(A), is a change in the slope coefficient β. The OLS estimate of β is equal to $cov(A,S)/var(S)$.

[12] Notes 2 and 7 describe differences between the PISA rounds for 2000 and 2003, the year we use. It is unclear how the analysis of the PISA 2000 data dealt with the differences in books categories from PIRLS 2001. The test score data do not appear to have been re-standardised to allow for the difference in the pool of countries participating in each survey.