Educational achievement in English-speaking countries: Do different surveys tell the same story?

John Micklewright and Sylke V. Schnepf

Abstract

International surveys of educational achievement are typically analysed in isolation from each other with no indication as to whether new results confirm or contradict those from earlier surveys. The paper pulls together results from four surveys to compare average levels of achievement, inequality of achievement, and the correlates of achievement (especially family background) among the six English-speaking OECD countries and between them and countries from Continental Europe. Our aim is to see whether a robust pattern emerges across the different sources: the Trends in International Maths and Science Study (TIMSS), the Programme of International Student Assessment (PISA), the Programme of International Reading Literacy Study (PIRLS) and the International Adult Literacy Survey (IALS).

S^{3}RI Applications Working Paper A04/10

Educational achievement in English-speaking countries: Do different surveys tell the same story?

John Micklewright [*] and Sylke V. Schnepf [+]

[*]School of Social Sciences and S3RI, University of Southampton
[+]Southampton Statistical Sciences Research Institute (S3RI)

May 2004

Abstract

International surveys of educational achievement are typically analysed in isolation from each other with no indication as to whether new results confirm or contradict those from earlier surveys. The paper pulls together results from four surveys to compare average levels of achievement, inequality of achievement, and the correlates of achievement (especially family background) among the six English-speaking OECD countries and between them and countries from Continental Europe. Our aim is to see whether a robust pattern emerges across the different sources: the Trends in International Maths and Science Study (TIMSS), the Programme of International Student Assessment (PISA), the Programme of International Reading Literacy Study (PIRLS) and the International Adult Literacy Survey (IALS).

Acknowledgements

This work is part of research supported by a grant from UNESCO Institute for Statistics, Montreal. Some of it develops unpublished joint work with Giorgina Brown, ISTAT, Rome, and Robert Waldmann, University of Tor Vergata, Rome. The views expressed are our own and should not be associated with UNESCO. The paper was written for the conference on 'Supporting Children: English-Speaking Countries in International Context', Princeton University, 8-9 January 2004. We are grateful to conference participants for comments and especially to our discussants, Janet Currie and Cecilia Rouse.

1. Introduction

Recent years have seen several international surveys of learning achievement of children and 'functional' literacy of adults (the ability to function in modern society). Samples of individuals are administered standardised tests with the aim of comparing countries' levels of achievement or literacy and the factors that influence them. These surveys focus on what people actually know or can do. This contrasts with data on 'attainment' that measure progression to different levels of national educational systems. Attainment is difficult to compare across countries due to institutional differences in how education is organised. And in any case, reaching a given level of education may correspond to very different levels of learning in an absolute sense from country to country. The international achievement surveys, with their purpose-built design for cross-national comparison, seem to cut through these problems.[1]

But which achievement survey to use? Each survey aims to assess something different (e.g. maths or reading) or to assess knowledge in a different way (e.g. in relation to an 'international' curriculum versus the ability to apply knowledge in everyday settings). They each refer to particular age groups or school grades. And they each have been the subject of criticism on one ground or another. A rounded picture of educational achievement of a nation's young people requires the surveys' results to be compared. But each survey is typically analysed in isolation with no consideration as to whether its results support or contradict those from another. In this paper we pull together the evidence from four different surveys to see if a robust picture exists of achievement and literacy in six English-speaking countries: the USA, Canada, Australia, New Zealand, the UK and Ireland. Average national performance, the extent of within-country differences, and the correlates of performance are all considered.

Why focus on the six English-speaking countries? First, these countries are frequently referred to loosely as a group in discussion of various aspects of living standards and social policy, but it is less often that they are considered together explicitly. Second, their common language helps the comparison among them of the survey results. Although the achievement surveys have instruments that are designed to work equally well in any language, inevitably there are concerns that full comparability is not obtained. Notwithstanding these concerns however, we also compare the English-speaking countries with other OECD members. The

[1] For convenience, we use the term 'achievement' in the paper to cover functional literacy as well.

issue of language differences is indeed one reason why a search for robust patterns across a variety of surveys is important in any such comparison.

Section 2 briefly introduces the four surveys we draw on: the Trends in International Maths and Science Study (TIMSS), the Programme of International Student Assessment (PISA), the Programme of International Reading Literacy Study (PIRLS), and the International Adult Literacy Survey (IALS). The first three focus on school-age children while from IALS we select young persons aged 16-24. Section 3 summarises the surveys' results in terms of countries' average levels of achievement and the differences in achievement within them. Do the English-speaking countries stand out on either aspect of achievement, what is the variation among them, and is the pattern of results similar across surveys? We also compare the absolute levels of achievement of those young people at the top of the national distributions and likewise for those at the bottom. Section 4 addresses the association of achievement with family background, again looking for common patterns or differences between the six countries and comparing them with other OECD countries. Section 5 discusses the possible relationship of achievement with public expenditures on education. Section 6 summarises our results and considers the importance (if any) of the findings.

2. The international achievement surveys

Table 1 lists our sources of data. Three of these surveys relate to school-aged children, and are conducted in schools. Their sample designs involve the selection of a sample of schools and then a single class (TIMSS and PIRLS) or a random sample (PISA) of pupils within each school. TIMSS is perhaps the best well known. We use data on 7th and 8th graders from the 1995 and 1999 rounds of the survey (taking the data from the later year if a country participated in both rounds).[2] The PISA data relate to an age group rather than a grade, which is an important difference. Some countries promote all children at the end of the year to the next grade irrespective of their achievement, while others insist on a certain competence being reached before passage upwards is allowed. (Several grades can be represented in an age group and several ages can be found in a grade, with the mixes varying from country to country.) The most recent of the three, PIRLS, focused on younger children – 9-10 year

[2] About one third of the questions to 8th-graders in 1999 were exactly the same as those put to 7th and 8th-graders in 1995. The others were intended to give results that were comparable. The 1995 round also collected data on 3rd and 4th graders and on children in the last grade of secondary school but we do not use those data here.

olds.[3] All three surveys collected information on the schools and on parental background (typically via questions to the children) as well as on the children's learning achievement. In contrast, IALS was a household survey collecting information on people of working age (16-65). We restrict attention to the young people aged 16-24, of which there are on average about 700 per country.[4]

The surveys differ widely in the type of achievement that they try to assess. IALS and PISA have some similarities, despite being aimed at very different age groups. (Both were organised by the OECD.) IALS was designed to measure the extent to which people are able to use literacy skills to perform everyday tasks, through assessment in three areas: prose literacy (understanding and using information from texts), document literacy (locating and using information contained in various formats) and quantitative literacy (applying arithmetic operations to numbers embedded in printed material). PISA assessed ability in reading, science and maths, attempting to determine to what extent "education systems in participating countries are preparing their students to become lifelong learners and to play constructive roles as citizens in society" (OECD 2001). Like IALS, the aim was to measure broad skills, trying to look at how students would be able to use what they have learned in real-life situations.

While covering a similar age group to PISA and two of the same subjects (maths and science), TIMSS focuses more on measuring mastery of internationally agreed curricula.[5] This may seem a narrow approach. But at least the concept of a curriculum agreed by educationalists is one that a lay person can begin to understand, even though the content of that curriculum is subject to debate. The 'life-skills' approach of PISA and IALS, on the other hand, seems more slippery. It may also be easier to carry out measurement of achievement against a standard in a culture-free way in TIMSS.[6] In any event, to the non-specialist it is not easy to see what the differences between PISA and TIMSS entail in practice, making it hard to choose between the two sources. There are also differences in the way information is

[3] 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 about 10 years for most of the countries.
[4] Details on the surveys can be found in their reports: Mullis et al (2000), Mullis et al (2003), OECD and Statistics Canada (2000) and OECD (2001).
[5] TIMSS 1995 covered 3rd and 4th grades, 7th and 8th grades and the last grade of secondary schooling. TIMSS 1999 assessed children in the 8th grade only. We focus on the most recent 8th grade data for each country (1995 in some cases if the country did not participate in 1999 or if a variable of focus was not included in the later year). When using 1995 data we used 'rescaled' data – scores derived from a psychometric model of the same form as that used for the 1999 data.
[6] The local PISA organiser in an Asian country commented to one of us that the PISA science questionnaire referred to central heating in one question. The conversion to a question on air conditioning in this instance may have been straightforward but the example serves to illustrate the general problem.

collected; TIMSS has more multiple-choice questions than PIRLS and PISA (about two-thirds of the TIMSS questions were multiple choice in 1999), while IALS has no multiple-choice questions at all. Countries vary in their traditions of multiple-choice testing in schools, so that some students will be more familiar with this form of questioning than others.[7]

The answers that a respondent gives to the questions in the surveys are summarised by the organisers into a single score for the subject concerned – maths, science, reading, different types of literacy, etc. This is usually scaled to have a mean among all persons in all participating countries of 500 and a standard deviation of 100 (the international mean in IALS is about 300). The aggregation of answers into a single score involves complex statistical modelling.[8] The basic principles in the process used in each survey are similar but the precise 'item response model' employed differs from survey to survey. Survey organisers do not report the sensitivity of results to the choice of model but our own investigations with TIMSS data suggest that this is not a trivial issue.[9] This is not an aspect of robustness that we explore in the present paper but differences between surveys in the modelling is one more reason for wanting to compare their results.

Hence age groups studied, subjects assessed, overall approach to assessment, form of questionnaire, and the method for aggregating the answers all vary from survey to survey. Other differences can be cited, including response rates. Even the basic premise that culturally-neutral questions can be successfully designed and translated into different languages can be debated, with the problems in this area probably varying from survey to survey.[10] In short, there seems ample reason for comparing results across the different surveys rather than relying on a single source.

[7] Examples of test questions from IALS, PISA and TIMSS are given in UNICEF (2002, p.10).

[8] The models come from the discipline of psychometrics. One stage of the process would be familiar to economists and sociologists using longitudinal surveys: the use of logit models for panel data (one can think of the series of answers by a respondent to the different questions as the panel element).

[9] See Brown and Micklewright (2004). The TIMSS organisers used different models in 1995 and 1999 but the 1995 data were re-modelled by the survey organisers in order to put them on the same basis as the later data. Hence results using two different models can be compared for the earlier year. We use the re-modelled 1995 data when we pool countries from the two years. See Goldstein (2003) for criticism of the modelling method used in PISA and for an illustration of an alternative.

[10] Overall country response rates in TIMSS and PISA averaged 88 percent (after replacement of non-responding schools with substitutes) and 85 percent respectively. Response in PIRLS averaged 92 percent (unweighted) and in IALS 62 percent. Variation across countries can be marked. Blum et al (2001) consider France's experience in IALS and among other things make critical comparison of the French language questionnaire used in France and that used in Switzerland. (France originally participated and then later withdrew.) See also the interchange between Prais (2003) and Adams (2003).

3. How do the English-speaking countries compare?

The six English-speaking countries – the USA, Canada, Australia, New Zealand, the UK and Ireland – all participated in TIMSS, PISA and IALS as did 12 other OECD countries. We begin our comparison by ranking these 18 countries on average score and within-country differences in scores in each survey. Both aspects of the distribution of achievement are of obvious interest. How well children and young people in any country are doing on average in absolute terms is important to know in a globalised world. But we also need to see the extent of educational inequalities within each country, inequalities that can be expected to help generate differences in incomes and other aspects of living standards in later life. And in both cases the performance of other countries is a natural choice of yardstick. Ranking countries may seem to encourage a view of national educational achievement as a beauty parade. But our main purpose in ranking is that it is a useful summary device that we use to put the surveys' results on a comparable basis.

There are eight rankings in all since eight subjects are covered in the three surveys (two in TIMSS and three each in PISA and IALS – see Table 1). We measure the average by the median and the within-country differences by the difference between 95th and 5th percentiles, P95-P5.[11] We then calculate a country's *average rank* in these eight rankings (weighting the surveys equally). This might seem an obscure way of presenting and comparing the survey results. But the average ranks have some merit as quick summary statistics. If the different subjects and surveys produced wildly differing rankings then the averaging would produce figures with little variation. A low rank in one 'league table' would likely be balanced by a high rank in another, so leaving all 18 countries clustered around an average rank of 9.5. The more the average ranks vary the more the rankings must be in agreement. Having a low or high average rank can only result from ranking consistently well or consistently badly in individual subjects.[12]

Figure 1 plots these average ranks in central tendency and dispersion against each other. They display a considerable amount of variation, implying that there must be a reasonable degree of concordance between the separate rankings. However, it is also true that

[11] As Figure 4 illustrates, scores are approximately Normal. Our results are not sensitive to our choice of measures of central tendency and dispersion.

[12] The disadvantage is that we give equal weight to an agreement between rankings on subjects within the same survey and an agreement between rankings for different surveys. One would expect the former to be stronger, and one would probably want to take less notice of it. For example, it turns out that all three within-survey correlations for PISA medians are higher than any correlation of a PISA median with those from the other two surveys. For example, PISA reading and PISA maths medians have a correlation of 0.81 while that between PISA reading and TIMSS maths is 0.46.

there is bunching in the middle of the distribution on each measure, arising either from countries consistently ranking mid-table or from an evening out of good performance on one subject and bad performance on another. The graph also allows one to see quickly whether there is a trade-off (in rank terms) between average achievement and inequality of achievement. The answer seems to be no. Countries with higher mean achievement (i.e. lower average values of ranks) also tend to have smaller within-country differences.

Where do the English-speaking countries come? We see immediately that there is substantial considerable variation among them. Canada and USA lie at the extremes of the group of six on both measures: Canada with higher mean and greater equity and the USA with lower mean and lower equity. The USA in fact comes last on equity of all 18 countries and last but two on average achievement. On the other hand Canada is (just) in third place in average rank terms for mean achievement and is in the top half of the distribution for equity. Australia is the only other English-speaking country in the top half on both measures. The UK and New Zealand are clearly more towards the USA on equity, coming in 16th and 17th places respectively on basis of their average ranks. The performance on mean achievement again puts them in the bottom half of the distribution but there is a large amount of bunching around the level of their average ranks. Ireland is right in the middle on both equity and average achievement.

Table 2 provides some insight into how these average ranks come about by looking at each of the surveys separately. It also shows the *sizes* of the differences in scores behind the rankings, although this will not mean much until we give an interpretation of the scores later on. And unlike Figure 1 it includes our fourth survey, PIRLS. The table reports the ratio of the national medians and values of P95–P5 to the average for non-English-speaking OECD countries. A value of 1.00 indicates that a country's mean achievement or within-country difference in scores in the survey in question was exactly the same as the average for OECD countries where English is not spoken. (The values are averages across the subjects each survey covers, three each for PISA and IALS, two for TIMSS and one for PIRLS.)

As far as mean achievement is concerned, the differences from the other OECD countries are typically quite modest: 10 percent or less in every case. Performance in PISA is actually better for all the English-speaking countries, save the USA (the average for the six countries is 1.04). On IALS it is worse (with all six at or below 1.00) while for TIMSS the picture is mixed. PIRLS (where the comparison group is more restricted) tells a similar story to TIMSS with all four countries close to the other OECD country mean. (The PIRLS data for Canada refer to Ontario and Quebec only.)

The differences are in general more marked for dispersion. IALS is again the survey that makes the English-speaking countries – all of them this time – stand out. The UK, New Zealand and the US are between a third to a half more unequal than the comparator group. Differences are smallest again in PISA while those in TIMSS come in between. PIRLS now clearly reinforces the general picture, with large percentage differences that are consistent with the position of the UK, New Zealand and the US in Figure 1 as the least equitable OECD countries in learning achievement.

Does high variance compensate for low mean as far as the absolute level of achievement at the top of a country's distribution is concerned? For example, do children at the top of the US distribution outperform those at the top of the Canadian distribution (the US being a high variance / low mean country and Canada the opposite)? Table 3 shows the situation for reading in PISA and PIRLS, maths in TIMSS and quantitative literacy in IALS, giving the 5th, 50th and 95th percentile scores for each of the six English-speaking countries (with ordering on the PISA scores). High achieving children in the US do *not* score above high-achieving children in Canada or in any of the other four countries, a result that is robust to choice of survey. (The small amount by which P95 in the US exceeds that of some countries in PIRLS, TIMSS and IALS is statistically insignificant.) In fact, the lack of variation in P95 in each survey is striking. Ireland and the US are a little adrift of most countries in PISA reading and the UK is somewhat ahead in PIRLS, but there are no significant differences between any of the scores in TIMSS or in IALS.[13]

By contrast, there is a lot more variation in the 5th percentiles, as can be seen by comparing the standard deviations of P5 and P95. For example, low achievers in the US fall below their counterparts in Canada by 51 points in PISA reading and by 50 points in TIMSS maths (the standard errors of the differences are 12 and 6 respectively).[14] The medians display a level of variation between those of P5 and P95 (PIRLS is an exception). Performance in the

[13] Standard errors for P5 and P95 are given in the surveys' published reports (those for TIMSS 1999 are the corrected versions on the TIMSS website) and take account of the complex survey designs. (We do not apply the Bonferroni adjustment for multiple comparisons.) We have proxied the standard errors for the UK in TIMSS and PIRLS with those for England. We do not know the standard errors in IALS but on the basis of the sizes of the differences and of the samples (much smaller than the other surveys) we are confident that all the differences insignificant. A more sophisticated comparison of the distributions for each country would be to check for 'first order stochastic dominance', i.e. whether every percentile for country X exceeds the corresponding percentile for country Y, an approach often used in the income distribution literature. Elsewhere in the paper we estimate standard errors for means and regression parameters using the 'svy' commands in STATA, allowing for clustering of students within schools.

[14] In all, 9 out of the 15 differences in P5 between countries in PISA reading are significant at the 5 percent level or less, 11 out of 15 in TIMSS maths and 5 out of 6 in PIRLS.

English speaking countries is therefore similar at the top of the distribution, less similar in the middle and quite different at the bottom.[15]

To this point we have not brought out the implications of the score differences in terms of something readily understood. The scores lack a natural metric. What does a difference of 30-50 points or of 5-10 percent really mean? Unless we know whether these numbers are big or small we cannot decide whether to be impressed by them or not. Figure 2 sheds some light by plotting the distribution of scores in Canada and the US for TIMSS maths separately for 7th and 8th graders in 1995 (only scores for 8th graders in 1999 enter the calculations for Figure 1 and Table 2). These countries have a difference in median 8th grade scores of 29 points in that year with the Canadian value about 6 percent higher than the US one. Figure 2 shows that this is roughly equivalent to the difference in medians *between* grades for either country. This is very close to the average difference in 7th and 8th grade medians for all 13 OECD countries in TIMSS 1995 of 27 points. In this sense, the median Canadian child is a year ahead of his or her US counterpart. Or, looking back at Table 2, the low-achieving US children are nearly two years behind the low-achieving Canadian children on this basis. This seems quite a big difference. The diagram also clarifies the extent of the dispersion of scores within a country. The difference between P5 and P95 in the US is twelve times the progression in median scores between the grades. But in Sweden and Canada – both low variance countries – this ratio is still as high as about eight or nine. As this illustrates, there are very large differences in achievement among children within one grade of compulsory school in *all* countries.

It is also useful to judge the levels of the scores against absolute benchmarks of competence in the subject concerned. Without these we don't know, for example, whether being at the bottom of the distribution in any country is a worry or not. Absolute benchmarks are provided by the survey organisers. Figure 3 shows the percentage of children considered by the organisers to be 'unable to do basic computations with whole numbers' (TIMSS) and 'unable to solve basic reading tasks such as locating straightforward information' (PISA). The English-speaking extremes of Canada and the USA are separated by eight points in PISA. Canada again does best in TIMSS with the back marker, New Zealand, 11 points behind. Many of the numbers seem sufficiently large to warrant concern. This implies that we are not

[15] The lack of variation in P95 does make one wonder whether scores at this level could be pushing up against an upper bound imposed by the test instrument in each survey. However, the 'item response' models used to aggregate an individual's answers to the set of questions in the test does in principle deal with this problem of censoring of ability level.

simply comparing distributions of achievement in countries where all children are comfortably above a satisfactory minimum level of competence.[16]

4. Family background and educational achievement

What differences in achievement exist between children from different family backgrounds and does each survey give a similar picture? We focus on TIMSS, PISA and PIRLS, which all cover children of compulsory school age. We first consider simple differences in mean scores for two characteristics and then turn to multivariate analysis of a wider range of factors.

a) family structure

There is a large literature on the association between child outcomes, including educational outcomes, and family structure, in particular being brought up by a single parent or step-parent (e.g. McLanahan and Sandefur (1994), Cherlin (1999), Case et al (1999) and Ginther and Pollack (2002)). We are unable to say anything concrete about causality with the achievement survey data (the literature just cited warns that we may be comparing two endogenous variables). Nevertheless, we view it as a worthwhile exercise to document the differences for all the English-speaking countries, compare these differences to those for other OECD countries, and see if they are robust to the choice of survey.

Table 4 shows the difference in mean scores between children from nuclear (both parents present) and single-parent families. The achievement of children in other family types (e.g. those with one natural and one step-parent) is investigated in regression analysis at the end of the section. Differences are almost invariably positive: children from nuclear families do better. To help interpret the numbers, it is worth bearing in mind the average progression in median scores between 7th and 8th grades in TIMSS of nearly 30 points.

The six countries are ranked in descending order of the differences for the PISA reading score. This puts the US at the top of the table but ranking on any of the six subjects would place the US in this position. The difference in scores for the US is around 50 points for each of the PISA subjects, 35-40 in TIMSS, and less than 15 in PIRLS. The difference in PIRLS scores in the other countries too is typically much less and is insignificant in the UK

[16] If we take the next benchmark up in TIMSS, the proportion of children judged 'unable to apply basic mathematical knowledge in straightforward situations' ranges from 23 percent in Canada to 44 percent in New Zealand.

and New Zealand (Canada is an exception). PIRLS is conducted on a younger age group than PISA – 10 year olds rather than 15 year-olds. This could be part of the explanation, differences by family background increasing as children get older. However, without checking TIMSS 3rd and 4th grade data, which we have not used in this paper, we can't say more.

At the other end of the range, Australia and Canada have differences in PISA reading that are similar to the average for OECD countries where English is not spoken. The TIMSS (and PIRLS) differences for Canada are notably larger than the other-OECD average, but Australia does not stand out in the same way. Irish children in single-parent families in TIMSS seem to be at no measurable disadvantage. Figure A1 in the appendix plots the differences for TIMSS maths and PISA reading for all countries in both surveys, including several from Central and Eastern Europe. The latter are among the countries with lowest values. The Scandinavian countries tend to have values at the average for TIMSS and above it for PISA.

The achievement deficit of children from single-parent families does vary substantially across the English-speaking countries. The different surveys show a not dissimilar picture, with the US being in a particularly clear position. The regression analysis later in the section shows whether these differences hold up when controlling for other factors and extends the comparison to children in other family types.

b) parental socio-economic status

A number of variables could be used to investigate the association between socio-economic background and children's achievement, a subject of enduring interest. The main PISA report (OECD 2001) placed considerable emphasis on family background, in particular the association of scores with indices constructed from principal components analysis of a range of parental characteristics, including occupational indices based on the work of Ganzeboom et al (1992). Obviously these indices cannot replicated for other surveys that have collected different family background data and we take a much simpler approach in order to make comparisons between the various data sources. This also has the merit of greater transparency with variables that are relatively easy to comprehend.

Education is one obvious parental characteristic to consider. What is the extent of intergenerational transmission from parents' education to children's learning while the children are still in compulsory schooling? The answer to this question would require more

than looking at just the raw association in the data given in Figure 4, but at least this represents a start. The graph shows the difference in mean scores between children whose mothers had at least upper secondary schooling and those whose mothers did not, comparing TIMSS maths with PISA reading. Children of less educated mothers clearly do less well everywhere. This is hardly surprising and what is of more interest is the extent of the difference, its variation between countries, and the agreement or its lack between the two surveys.

The extent of the difference in both surveys is from a minimum of around 25-30 points in Korea, Iceland and Norway (about one school year in our TIMSS metric described earlier) to over 70 points in PISA reading and 50 points or more in TIMSS maths in about half a dozen countries in each case. The latter group includes the US as only one of two countries with this size deficit in both surveys, the other being Hungary. The other English-speaking countries vary more on PISA differences than they do on TIMSS but stick out in neither case.[17]

A closer pattern of agreement between the surveys is seen in Figure 5, which shows the difference in mean achievement between children in families with more than 100 books in the household and those with less. This is information collected from the children (as is parental education) and it must be subject to considerable measurement error. Nevertheless, the reader who is sceptical (as were we) of the data may be moderately impressed by Figure A2 in the appendix which shows a close correlation between the proportions of children in PISA and TIMSS saying their families had more than 100 books (between 45 and 60 percent of children among the English-speaking countries). Figure 5 shows the English-speaking countries are pretty spread out in both surveys as far as the differences in mean achievement are concerned, with Canada at one extreme (also the minimum value for any OECD country in TIMSS) and the US at the other, with New Zealand and the UK not far behind.

c) controlling for other factors

The differences in mean achievement by family structure in Table 4 may reflect other factors, including differences in parental education. Likewise the apparent variation in achievement

[17] Between 18 percent (Canada) and 45 percent (Ireland) of children in the English-speaking countries have mothers who did not have upper secondary schooling. The position of the UK in this diagram is probably misleading. No parental education data were collected by TIMSS in England so the UK difference for TIMSS in Figure 4 is the value for Scotland alone. Comparing means for Scotland and England in PISA suggests that were parental education to have been collected in TIMSS then the difference for the UK would have been a fair bit larger.

by the latter, or by the number of books, could be proxying something else.[18] Tables 5a and 5b investigate this issue using multivariate regression for the maths scores in PISA and TIMSS. The focus on maths has the advantage of standardising for subject between the surveys along with the obvious disadvantage of the restriction to a single subject. The explanatory variables include a mix of individual and family characteristics besides family structure, mother's education, and books in the home: the child's gender, the presence of siblings and migrant status. They also include several characteristics of the child's school: location (urban/rural), student-teacher ratio and the school principal's perception of shortages. The tables report only the estimated coefficients of the family structure, mother's education and books in the home variables (those relating to other variables are given in the appendix). The equations are estimated for each of the six English-speaking countries and for France, Germany and Sweden as examples from Continental Europe.

The specification of the family structure and mother's education variables is more detailed than in the bivariate analysis. We include dummies both for children from single-parent families and for those from 'mixed' family structures e.g. one natural and one step-parent, or a natural parent with a partner to whom they are not married. (Children in nuclear families are the base category.) For parental education we include dummies for mother's education at the secondary and tertiary levels (the latter showing the additional effect of tertiary over secondary) and another where information on her education is missing altogether (which may proxy low achievement quite well given that the relevant question is asked of the child).

When compared to the raw differences in Table 4, the controls for other characteristics result in the single parent 'effect' in PISA falling by about a third to half. For example, the raw difference in Table 4 for the US is 48 points, down to 33 points once controls are added. The change in the UK is from 29 points to 22 points. In Australia the difference is no longer significant at the 5 percent level. At the same significance level one cannot reject the hypothesis in *any* of the other five English-speaking countries that the difference is the same as that in Sweden. The ceteris paribus differences for children from 'mixed' family types are in general higher for children from single-parent families (the UK and France are exceptions) although not always significantly. This conforms to the finding in the literature referred to earlier that it is children in these families that often have the worst outcomes.

[18] The possibility that on average low ability children underestimate the number of books (rendering the variable endogenous) cannot be ruled out.

This seems a fairly clear set of results but those for TIMSS are rather different – a useful reminder of the need not to get fixated on one survey alone. Here the controls have less impact in reducing the apparent deficit of single-parent children. There is little or no change in Australia and Canada and only moderate change in New Zealand and the UK. The Irish single parent 'effect' actually increases with respect to Table 5, becoming significant. Only the US sees a substantial abatement. And in contrast with the PISA results, all six English-speaking countries have a difference that is significantly larger at the 5 percent level than that in Sweden. However, the 'mixed' family children again have deficits that are bigger than those for single-parent children, and significantly so in all but Canada, Germany and the US.

A mother with upper secondary education is worth about 20 PISA maths points in all nine countries, except in the US and Germany where the 'effect' is greater, countries that are both outliers in the crude PISA reading differences in Figure 4. A mother with tertiary education is associated with a more variable increment in achievement – none in Ireland, France and Sweden but 25-30 points in Australia, the USA and Germany. (The way we have specified the model means this is the additional impact over and above that for secondary education.) Children who cannot report their mother's education are tend to be lower achievers (the US, Germany and Ireland are exceptions). The pattern for books in the home follows that for reading in Figure 5, with larger differences for NZ, the UK and the US. The TIMSS results display more variability on both education and books.[19]

To summarise the results, we calculate the predicted scores for children from two types of family (Table 6). One child (Type A) is from a nuclear family, has a mother with tertiary education and he or she estimate that there are more than 100 books in their home. The other (Type B) is from a single-parent family, has a mother who did not complete secondary education and believes that there are fewer than 100 books in the home. We report both the predicted scores and the differences in these scores between the two family types, ranking on the latter basis for PISA. We also show what the difference represents in terms of the national standard deviation.

The US, the UK and New Zealand occupy the top three places in PISA with differences between the family types of 100+ points. And they are in 3 of the top 4 places in TIMSS. Canada and Ireland are in the bottom two places in both cases. Only Germany among the three Continental European comparators is up at the 100+ level, and only for PISA. The predicted difference between two children of these family backgrounds in the US is one and a

[19] The mother's education missing variable is effectively an England dummy in the case of the UK since information on parental education was not collected in that part of the UK.

half standard deviations in PISA. This is getting on for twice as much as in Sweden where the figure is less than in all the English-speaking countries (also the case in France). Canada stands out for the opposite reason in the TIMSS data: smaller differences than the other five countries.

The absolute levels of the scores for both family types show considerable variation across the English-speaking countries (the standard deviations can be compared with those for P95 and P5 in Table 3). This is true of both surveys but the variation is less for TIMSS. The UK stands out in PISA for the high predicted scores for both types of child while the US and Ireland stand out for their low scores for Type B child, well beneath those of the other four countries. However, neither result is found in TIMSS. Here the UK is little different from the rest; the Type A child in the US is up at the Canadian level while Canada stands out for the higher achievement of the Type B child.

5. Achievement and Public Expenditures

It is tempting to view our regressions in the previous section as providing estimates of educational production functions. The equations contain some school input variables (reported on in the appendix) and one might think to compare their impact from country to country and hence learn something about how expenditures on education are transformed into children's learning and how this varies across the countries we consider. And in the ideal world, if there were well-determined effects of the school input variables then these could be interacted with the family background variables to try to see if the level of school inputs matters more for children from disadvantaged backgrounds.

We have resisted the temptation to go down this route for several reasons. First, our quest to compare results across surveys (the focus of the paper) means that the range of school characteristics that we can include is very restricted since each survey tends to collect rather different information in this area. Second, there are well-known concerns over the possible endogeneity of some potential regressors. For example, consider class size (not included in our equations but present in both PISA and TIMSS); this is an input measure of natural interest with a direct inverse link to expenditures. One might expect class size to have a negative effect on learning. But achievement may be lower for children in smaller classes because schools devote more resources to low ability students, sorting them into smaller

groups.[20] Third, countries may differ in their methods of trying to produce educational outputs from a given set of inputs (school, family and others, both observed and unobserved). Looking at the ceteris paribus effect of an observed school input, X, in each country may give the impression that X is used poorly in some countries (small estimated coefficient) and well in others (large coefficient) whereas it may just be the case that the way the inputs are combined varies across countries.

School inputs can be measured in monetary terms and their aggregate then compared across countries. Figure 6 plots the average median score in the five subjects covered by TIMSS and PISA (which all have a similar scale) against an estimate of the cumulative expenditure (both public and private) on a child's education up until age 15 measured in purchasing power parity US Dollars. These estimates on the horizontal axis are taken from the first PISA report from the OECD where they were used as the explanatory variable in a bivariate regression of average mean achievement in the three PISA subjects: 'as expenditure per student on educational institutions increases, so also does a country's mean performance, expenditure per student explaining 17 percent of the variation between countries in mean performance' (OECD 2001: 93). However, we have reproduced this result and find that it is entirely dependent on the inclusion in the regression of an outlier, Mexico. (Mexico is not in Figure 6 since it did not participate in TIMSS). With Mexico excluded, the regression collapses (the coefficient on the expenditures variable halves and the t-statistic falls from 2.5 to 1.4). Including the two TIMSS subjects as in Figure 6 changes nothing – the expenditure variable remains insignificant (we use the medians from each survey rather than means but this has no impact on the results). It is also insignificant if we use the average of the P5 values as the dependent variable, to see if expenditures have an association with achievement at the bottom of each national distribution. And nor is there any change to this picture (for either the median or P5 achievement) if we use an estimate of only the public sector's expenditures rather than the sum of the public and private totals.[21]

[20] When we include class size in the equations (whether in addition to or instead of the school's student-teacher ratio) it typically has a positive association with achievement (as found with other data sources, see e.g. Vignoles et al 2000), which is consistent with the reverse causation described in the text. Similarly, the estimated effect of the student-teacher ratio (which we do include) is hard to interpret since educational jurisdictions may put more resources (i.e. teachers) into schools where students are weak. More generally, the 'does money matter?' debate has raged long and hard in research on education (e.g. Burtless 1995, Hanushek 1995).

[21] We estimate the public sector expenditures by applying to the Figure 6 data the percentage of total expenditure in each country on primary, secondary and post-secondary non-tertiary education in 1999 that is public, taken from OECD and UNESCO Institute for Statistics (2002: 185). This figure averages over 90 percent (including in the US).

This lack of any relationship between the national averages or P5 values and total expenditures is not surprising. Not only may a simple bivariate regression using aggregate data be far too blunt a tool to reveal much but one also has to think about what the differences in expenditure actually represent. Total public and private expenditures vary across the English-speaking countries from about $35,000 (in PPP terms) in Ireland to over $70,000 in the US. What does the additional money buy in the US? Teacher compensation represents the lion's share of expenditures in both countries (the figure is higher in Ireland). If average teacher salaries in the US are say 50 percent higher than in Ireland does that mean that the US teachers are on average 50 percent more productive? [22] It seems unlikely.

Public expenditures on education of children come primarily via schools. It is schools that receive public money and convert it into staff time, infrastructure or other inputs. And each school spends in ways that may often benefit all their children, rather than a select few. With this in mind it is useful to look at how total variation of achievement scores within countries is split between that between schools and that within schools. Figure 7 shows this split for PISA reading. Both the variation within schools (vertical axis) and between schools (horizontal axis) are expressed relative to the average variation within the OECD as a whole. The horizontal and vertical dashed lines showing the average OECD split: about two-thirds within schools and one-third between schools. Countries outside the diagonal line have more total variation in achievement than the OECD average, i.e. more dispersed scores than on average is found elsewhere.

The English-speaking countries all have less variation between schools than the OECD average and more within schools – they all come in the north-west quadrant. The US is only just in this position; it is outside the diagonal line solely because of its greater than average within-school variation. (This may come as some surprise to those who argue that the size of the US coupled with its very decentralised school system must lead to large between school differences.) Alternatively, viewed another way, the US has the same extent of within-school variation as Iceland and Sweden but differs from them in overall dispersion because of higher between school variation.[23] A high share of within school variation suggests that public policy to tackle educational disadvantage needs to operate through families as much as through schools. (We have yet to check the within-between split in TIMSS or PIRLS, where

[22] 56 percent of total expenditure (public and private) in primary, secondary and post-secondary non-tertiary education in the US in 1999 went on teacher compensation and 80 percent in Ireland. Student-teacher ratios in lower secondary schooling in the two countries are virtually the same, 16.3 and 15.9 (15.8 and 21.5 in primary schools) (OECD and UNESCO Institute for Statistics 2002: 190 and 203).

[23] The countries with high values of between-school variation are often those with state school systems that separate secondary school children by ability, e.g. Germany, Austria and Hungary.

the figures should be different in part because these surveys take a single class within a school rather than a random sample of students.)

6. Conclusions

Results of the international surveys of learning achievement of young people need to be compared with each other and this has been the main purpose of the paper. We have drawn on four different surveys and have considered the position of the English speaking countries as the substantive focus.

The surveys show the English-speaking countries to have achievement levels that are typically around the average when compared to other OECD countries and, more obviously, that display above average spread. The situation does vary from survey to survey but taken together the different sources seem to provide a reasonably clear picture, especially on the within-country differences, where New Zealand, the UK and the US stand out as high variance countries. Canada tends to be an exception to the rule – with higher average and lower variance – and is the most obvious example of the variation that exists among the English-speaking six. Performance at the top of the distribution displays little variation across the six countries, while performance at the bottom varies a lot. It is important to note that all countries (both those that speak English and those that do not) have large differences within them for a single age group or grade.

Differences within countries are strongly associated with family background. This is scarcely surprising but we have quantified their extent across surveys and across countries. Countries with larger differences are the ones with greater dispersion of scores overall – the US, the UK and New Zealand. Children from single parent and 'mixed' family types do worse than those from nuclear families in all English speaking countries with few exceptions in terms of subject and survey. Those countries that do best in containing the difference manage to restrict it to about what is found on average in the rest of the OECD.

Differences between countries in average levels of achievement have no relationship with summary measures of real national expenditure on education in the 18 countries considered in this paper, a result robust to choice of measure and survey (the same is true of levels at the bottom of the distribution).

Do any of these differences in achievement that we have found among English speaking countries and between them and other countries really matter in a fundamental sense – in terms of future learning, labour market success, or other later life outcomes? There is

certainly evidence that outcomes within any country are linked to the measures covered by the international surveys. For example, Blau and Kahn (2001) relate individuals' wages recorded in IALS to their test scores and conclude that a part of higher wage inequality in the US is explained by the higher dispersion of scores than elsewhere. But what of the differences between countries in absolute levels of learning? Clearly there are many other things that differ between countries' economies and societies. Difference in children's learning levels (such as the consistent shortfall of the US low-achievers below Canadian low-achievers) do not therefore necessarily translate directly into an equivalent difference in real living standards in later life. But if we believe the data do contain useful information on learning then it is hard to believe that these differences are of no import at all. It would therefore be interesting to try to relate the results on achievement that we have found in this paper to other cross-national results on outcomes (e.g. see Smeeding 2002 for comparisons of the real living standards of families with children).

References

R J Adams (2003) 'Response to 'Cautions on OECD's recent educational survey (PISA)'', *Oxford Review of Education*, Vol. 29 no 3.

F. Blau and L. Kahn (2001), 'Do Cognitive Test Scores Explain Higher US Wage Inequality?' NBER Working Paper 8210.

A. Blum, H. Goldstein and F. Guerin-Pace (2001), 'An Analysis of International Comparisons of Adult Literacy', *Assessment in Education*, vol. 8, no. 2.

G. Brown and J. Micklewright (2004), 'Using International Surveys of Achievement and Literacy: A View from the Outside', UNESCO Institute for Statistics working paper, Montreal, forthcoming.

G. Burtless ed. (1996), *Does Money Matter: The Effect of School Resources on Student Achievement and Adult Success*, Washington DC.

A. Case, I-F. Lin and S. McLanahan (1999), 'Household Resource Allocation in Stepfamilies: Darwin Reflects on the Plight of Cinderella', *American Economic Review*, Vol. 89, No. 2, pp. 234-238.

A. Cherlin (1999), 'Going to Extremes: Family Structure, Children's Well-Being, and Social Science', *Demography*, Vol. 36, No. 4, pp. 421-428.

H.B.G. Ganzeboom, P. De Graaf and D.J. Treiman (with J. De Leeuw) (1992), 'A standard international socio-economic index of occupational status', *Social Science Research*, Vol. 21 (1), pp. 1-56.

D. Ginther and R. Pollack (2002), 'Does Family Structure Affect Children's Educational Outcomes?', *Federal Reserve Bank of Atlanta Working Paper*, 2000-13a.

H. Goldstein (2003), 'International comparisons of student attainment: some issues arising from the PISA debate'. Institute of Education, University of London (www.ioe.ac.uk/hgpersonal/)

E. Hanushek (1995), 'Interpreting Recent Research on Schooling in Developing Countries' *World Bank Research Observer*, vol 10 no 2, pp.227-46

S. McLanahan and G. Sandefur (1994), *Growing Up with a Single Parent: What Hurts, What Helps*, Harvard University Press, Cambridge.

I. Mullis, M. Martin, E. Gonzales and A. Kennedy (2003), *PIRLS 2001 International Report*, Boston College.

I. Mullis, M. Martin, E. Gonzalez, K. Gregory, R. Garden, K. O'Connor, S. Chrostowski, T. Smith (2000), *TIMSS 1999 International Mathematics Report*, Boston College.

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

OECD and Statistics Canada (2000), *Literacy in the Information Age – Final Report of the International Adult Literacy Survey*, OECD, Paris.

OECD and UNESCO Institute for Statistics (2002) *Financing Education – Investments and Returns*, OECD, Paris.

OECD and UNESCO Institute for Statistics (2003) *Literacy Skills for the World of Tomorrow - Further results from PISA 2000*, OECD, Paris

S. Prais (2003), 'Cautions on OECD's recent educational survey (PISA)', *Oxford Review of Education*, Vol. 29, pp. 139-163.

T. Smeeding (2002), 'Real Standards of Living and Public Support for Children: A Cross-National Comparison', Luxembourg Income Study Working Paper 345.

UNICEF (2002) *A League Table of Educational Disadvantage in Rich Nations*, Innocenti Report Card 4.

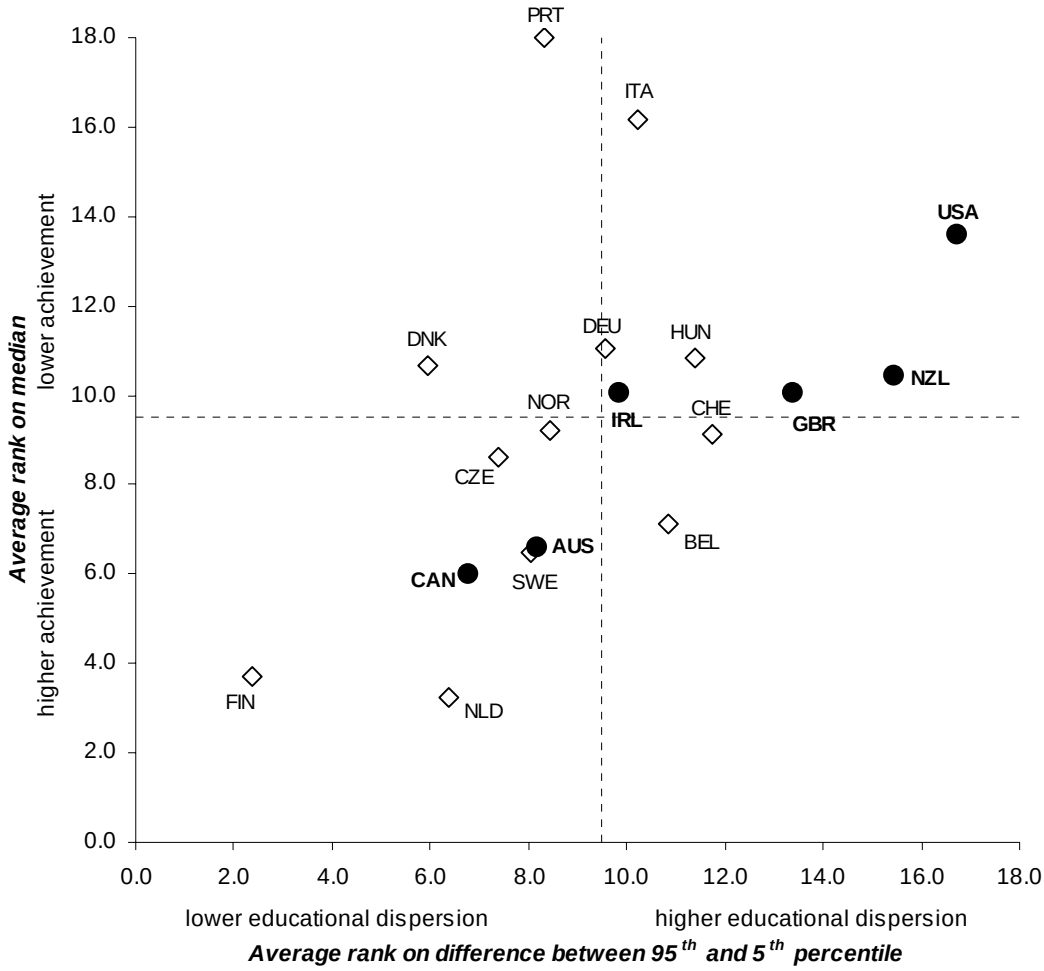
A. Vignoles, R. Levacic, J. Walker, S. Machin, and D. Reynolds (2000), 'The Relationship between Resource Allocation and Pupil Attainment: A Review', Centre for the Economics of Education, London School of Economics, Discussion Paper 2

Table 1: Cross-national survey data on achievement used in this paper

Survey	Age group	Subjects covered	Sample size per country (av.)
Trends in International Maths and Science Study (TIMSS), 1995 and 1999	14 (grades 7 and 8)	maths and science	3,800 at each grade
Programme of International Student Assessment (PISA), 2000	15	reading, maths and science	5,700
International Adult Literacy Survey (IALS), 1994-98	16-24	document, prose and quantitative literacy	700
Progress in International Reading Literacy Study (PIRLS), 2001	10	reading	4,300

Notes: TIMSS and PIRLS are organised by the International Study Center, Boston College, USA. PISA is organised by OECD. IALS was organised by OECD and Statistics Canada.

Figure 1: Average rank in educational achievement and educational inequalities for 8 different measures (PISA, TIMSS, IALS)



Note: the graph shows the average rank on median and difference between 95th and 5th percentiles in 3 surveys (TIMSS, PISA, IALS) with 8 different subjects for 18 countries. The higher the median and the lower the dispersion the smaller in number the rank. Gridlines show the average for all countries (9.5). The surveys are equally weighted in the averaging. IALS Belgium refers to only Flanders. GBR refers to England and Scotland for TIMSS and IALS.

Table 2: Ratio of median and P95–P5 to the average for non-English speaking OECD countries

		PISA	TIMSS	IALS	PIRLS
Median	Canada	1.06	1.03	1.00	1.01
	Australia	1.05	1.04	0.98	-
	UK	1.05	1.00	0.96	1.02
	New Zealand	1.06	0.98	0.96	0.99
	Ireland	1.02	1.01	0.95	-
	USA	0.99	0.99	0.91	1.01
	Average	*1.04*	*1.01*	*0.96*	*1.01*
P95 – P5	Canada	0.93	0.97	1.15	1.07
	Australia	0.99	1.06	1.07	-
	Ireland	0.93	1.11	1.18	-
	UK	1.01	1.12	1.33	1.30
	New Zealand	1.06	1.16	1.37	1.38
	United States	1.06	1.18	1.52	1.23
	Average	*1.00*	*1.10*	*1.27*	*1.25*

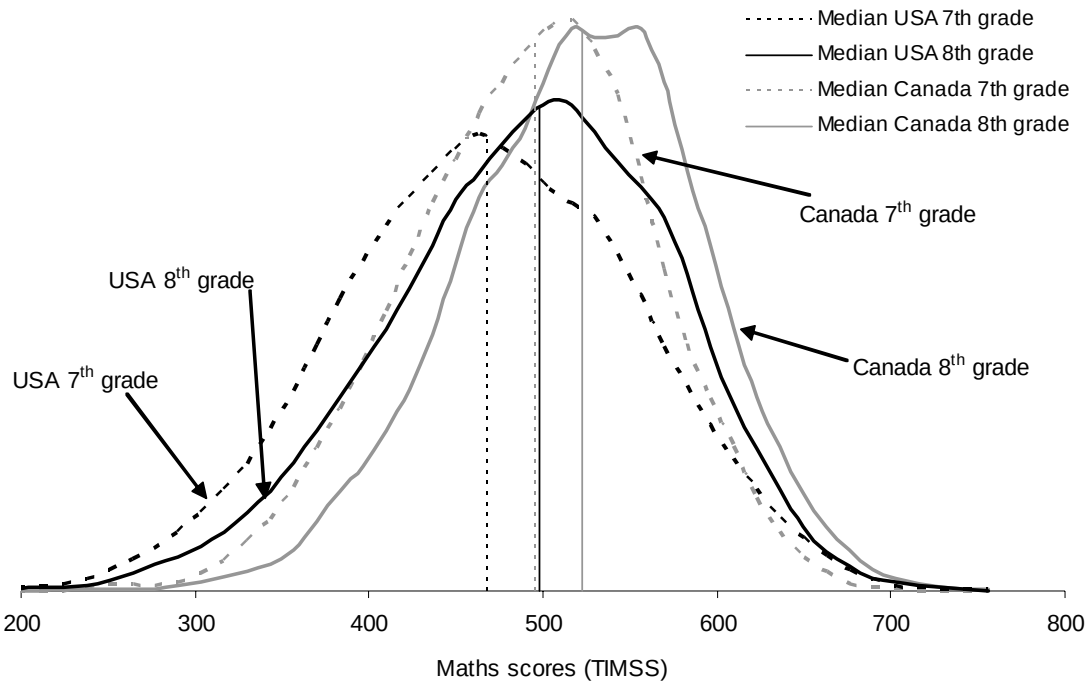
Note: Other OECD countries for PIRLS are: Czech Republic, Germany, Hungary, Italy, Netherlands, Norway and Sweden. In addition to these, the other OECD countries for PISA, TIMSS and IALS are Belgium, Denmark, Finland, Portugal and Switzerland. The figures for each survey are an average value for the subjects it assessed. Countries are ordered by their average position in PISA, TIMSS and IALS. Data for the UK refer to England and Scotland for TIMSS, IALS and PIRLS.

Table 3: P5, P50 and P95 for English-speaking countries in PISA and PIRLS reading, TIMSS maths and IALS quantitative literacy

		PISA reading	TIMSS maths	IALS quant.	PIRLS reading
P95	New Zealand	693	632	359	668
	Australia	685	648	357	
	UK	682	632	353	683
	Canada	681	646	360	658
	Ireland	669	643	357	
	USA	669	642	355	663
	Std dev.	*9.2*	*7.2*	*2.6*	*11.0*
P50	Canada	540	533	290	547
	New Zealand	538	493	276	537
	Australia	534	529	286	
	Ireland	533	525	280	
	UK	527	496	274	557
	USA	511	504	262	551
	Std. dev.	*10.6*	*17.8*	*9.9*	*8.2*
P5	Canada	371	406	194	419
	Ireland	360	371	167	
	Australia	354	387	190	
	UK	352	360	142	393
	New Zealand	337	341	158	360
	USA	320	356	126	389
	Std dev.	*18.0*	*23.5*	*26.7*	*24.2*

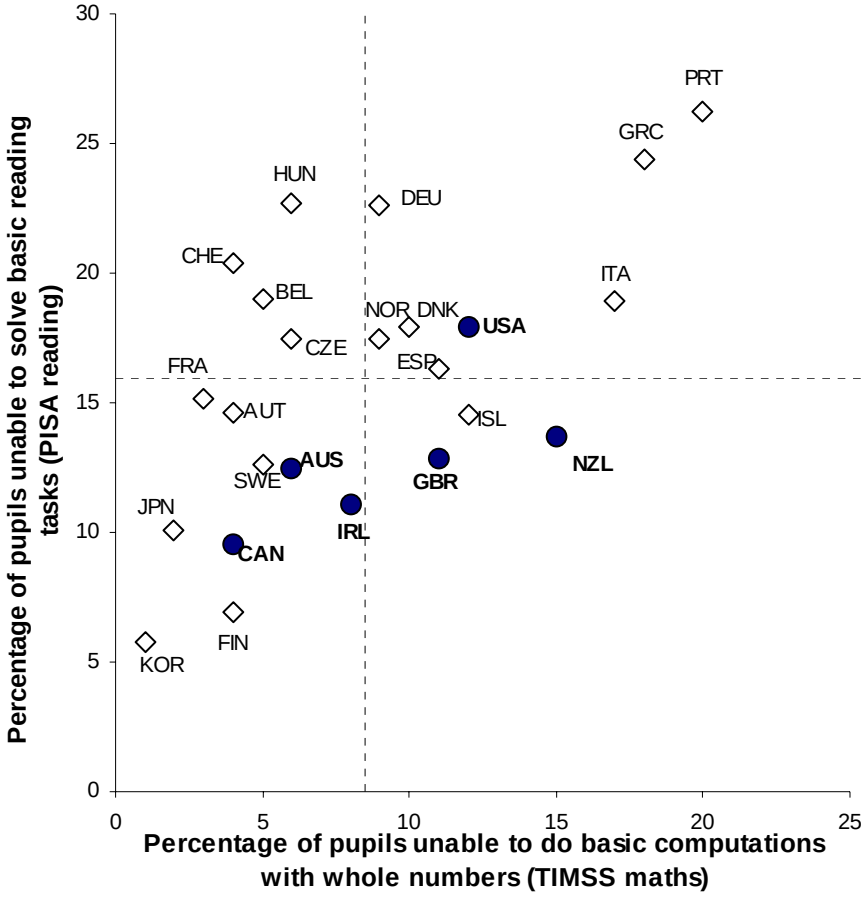
Note: countries are ordered on the PISA scores. Data for the UK refer to England and Scotland for TIMSS, IALS and PIRLS.

Figure 2: Maths achievement in 7th and 8th grade in USA and Canada (TIMSS 1995)



Note: Kernel density estimation using the average of plausible values.

Figure 3: Percentages of pupils judged "unable to do basic computations with whole numbers" (TIMSS maths 8th grade) and "unable to solve basic reading tasks" (PISA reading)



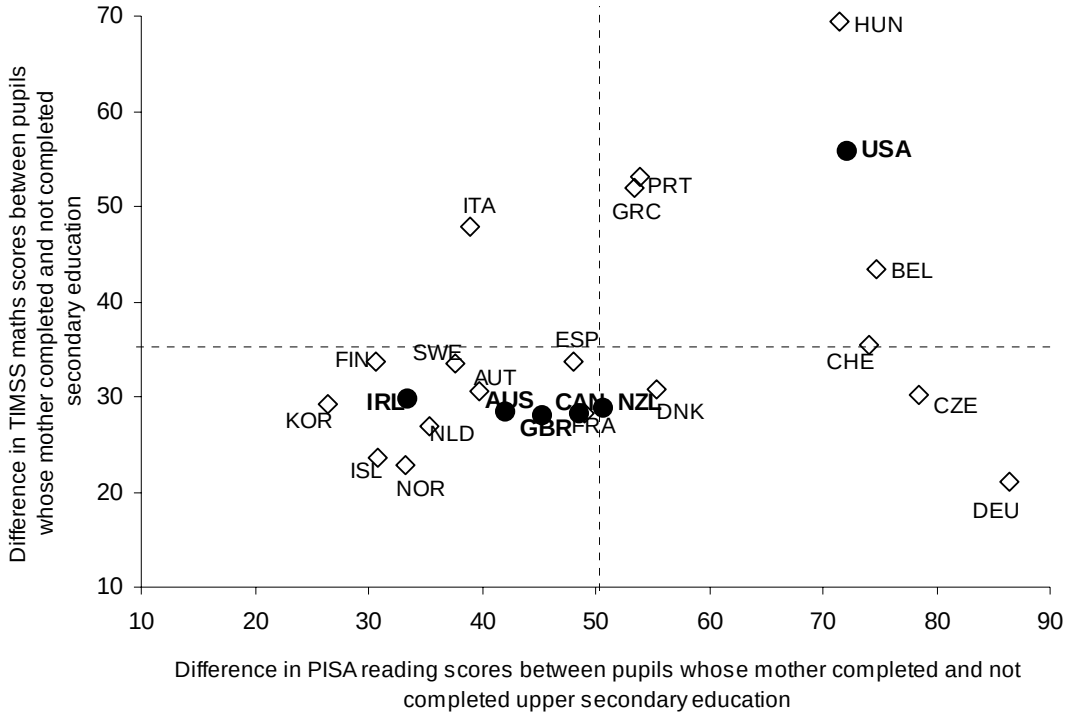
Note: Grid lines refer to OECD average. Pupils who are considered by the organisers to be 'unable to do basic computations with whole numbers' are those who score below the lower quarter international benchmark in TIMSS maths. Pupils who are 'unable to solve basic reading tasks' such as locating straightforward information are those who score at or below PISA reading literacy level 1.

Table 4: Difference in scores between children from nuclear and single parent families in PISA, TIMSS (8th grade) and PIRLS

	PISA			TIMSS		PIRLS
	Reading	Math	Science	Math	Science	Reading
USA	46.0	48.3	50.3	34.5	38.5	13.9
UK	32.3	28.7	31.7	22.6	21.1	7.6*
New Zealand	28.0	30.2	16.9	23.8	25.1	1.4*
Ireland	23.1	24.2	20.3	8.6*	-0.8*	-
Canada	14.2	19.6	16.1	22.3	19.4	13.4
Australia	13.7	14.2	8.2*	15.1	9.4	-
Average ESC	**26.2**	**27.5**	**23.9**	**21.1**	**18.8**	**9.1**
Other OECD	**13.6**	**18.4**	**12.1**	**13.3**	**9.7**	**6.2**

Note: Grey fields denote that the average difference in achievement is significantly higher (5 percent level) than that of other OECD countries. Other OECD countries are for PISA and TIMSS: Austria, Belgium Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Iceland, Italy, Korea, Netherlands, Norway, Spain, Sweden and Switzerland. For PIRLS other OECD countries are Czech Republic, France, Germany, Greece, Hungary, Iceland, Italy, Netherlands, Norway and Sweden. For TIMSS and PISA children in single parent families live with one of the following: mother, father, female guardian or male guardian; children in nuclear families live with their father and mother. For PIRLS single parent and nuclear are proxied by children in single adult families and children in all other families. Asterisks indicate that differences are not significant at a 1 percent level. For TIMSS, data for New Zealand, Canada, Australia and Ireland are for 1995, data for UK refer to Scotland 1995 and England 1999, and USA to 1999. (About two thirds of other OECD countries in the TIMSS calculation refer to 1995 data.)

Figure 4: Difference in PISA reading and TIMSS maths scores between pupils whose mother completed and not completed (upper) secondary education



Note: GBR refers solely to Scotland for TIMSS.

Figure 5: Difference in reading (PISA) and math scores (TIMSS) between pupils with up to 100 and with more than 100 books at home

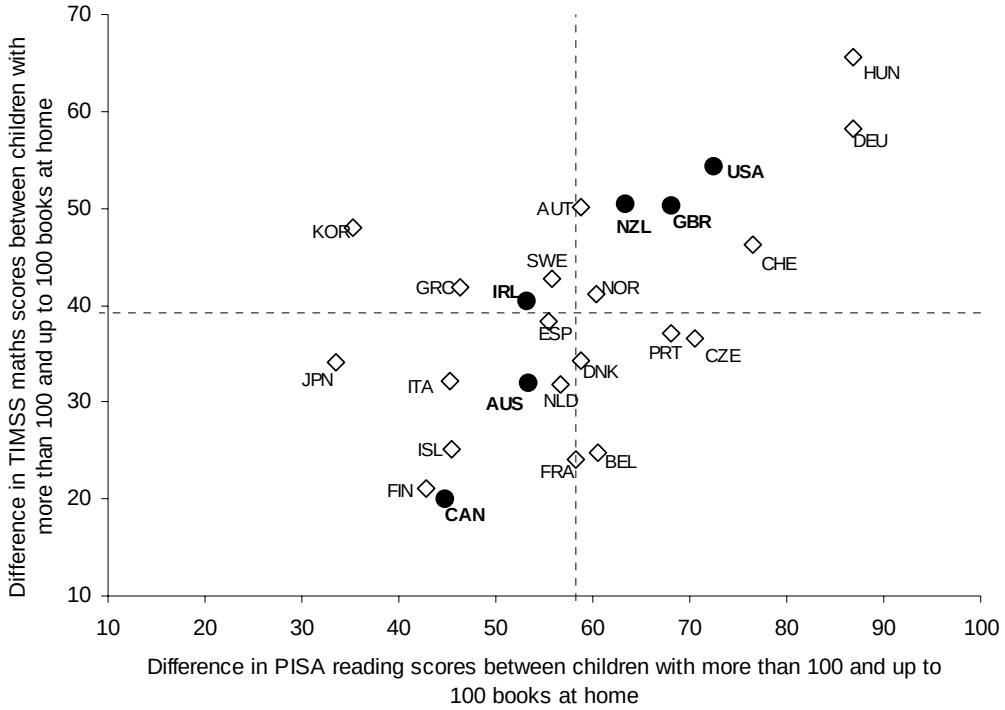


Table 5a: OLS regression coefficients for family background in PISA maths

PISA	Australia	Canada	Ireland	New Zealand	UK	USA	France	Germany	Sweden
Single parent family	-6.8	-14.7	-13.9	-18.2	-21.7	-32.5	-16.6	-12.9	-17.0
	(5.1)	(2.5)	(5.0)	(5.0)	(3.6)	(6.1)	(4.4)	(4.7)	(5.1)
Other family type	-12.1	-22.9	-27.8	-31.1	-18.6	-46.7	-12.8	-30.8	-22.1
	(5.7)	(2.5)	(7.4)	(5.3)	(5.4)	(4.9)	(5.2)	(6.0)	(4.9)
Mother secondary education	15.0	19.5	24.2	20.7	20.7	44.1	24.1	38.3	20.4
	(5.3)	(2.7)	(3.6)	(5.7)	(4.5)	(8.1)	(3.9)	(5.5)	(5.2)
Mother tertiary education	29.2	17.5	-0.3	23.4	12.3	24.3	3.5	28.6	0.9
	(4.9)	(2.2)	(4.5)	(4.3)	(4.2)	(5.5)	(4.2)	(5.4)	(3.7)
Data education missing	-23.2	-30.5	-16.1	-11.6	-20.8	-7.8	-35.9	13.3	-32.1
	(8.6)	(7.4)	(17.1)	(6.4)	(7.1)	(10.0)	(8.1)	(7.4)	(9.3)
Books in household > 100	35.5	26.2	39.8	44.6	45.7	46.9	33.5	50.6	36.2
	(4.2)	(1.9)	(3.7)	(4.5)	(4.0)	(4.8)	(3.7)	(4.2)	(4.0)

Table 5b: OLS regression coefficients for family background in TIMSS maths

TIMSS	Australia	Canada	Ireland	New Zealand	UK	USA	France	Germany	Sweden
Single parent family	-18.5	-18.4	-15.4	-15.7	-16.7	-24.5	-9.9	-13.4	-5.2
	(2.7)	(2.4)	(3.6)	(3.2)	(3.2)	(3.0)	(2.6)	(3.4)	(2.8)
Other family type	-26.5	-17.8	-30.2	-35.1	-28.4	-30.8	-17.7	-22.9	-18.0
	(3.5)	(2.6)	(6.5)	(3.8)	(3.9)	(3.3)	(2.4)	(3.5)	(2.3)
Mother secondary education	6.3	19.7	18.1	0.0	7.4	22.2	24.4	7.9	18.1
	(2.8)	(3.0)	(2.5)	(3.4)	(3.3)	(3.8)	(3.0)	(3.0)	(2.8)
Mother tertiary education	28.9	6.7	2.2	26.3	29.2	21.5	10.7	15.1	6.1
	(3.1)	(2.4)	(3.5)	(3.4)	(6.3)	(3.2)	(3.1)	(4.6)	(2.6)
Data education missing	-15.5	-2.4	-17.8	-13.5	-4.3	3.7	-3.0	-20.8	-0.9
	(2.7)	(3.4)	(3.0)	(3.2)	(4.4)	(4.0)	(2.5)	(3.1)	(2.8)
Books in household > 100	37.3	9.2	33.1	44.3	45.8	37.2	19.4	43.6	36.7
	(2.8)	(2.5)	(2.4)	(3.1)	(4.3)	(2.3)	(2.6)	(3.4)	(1.9)

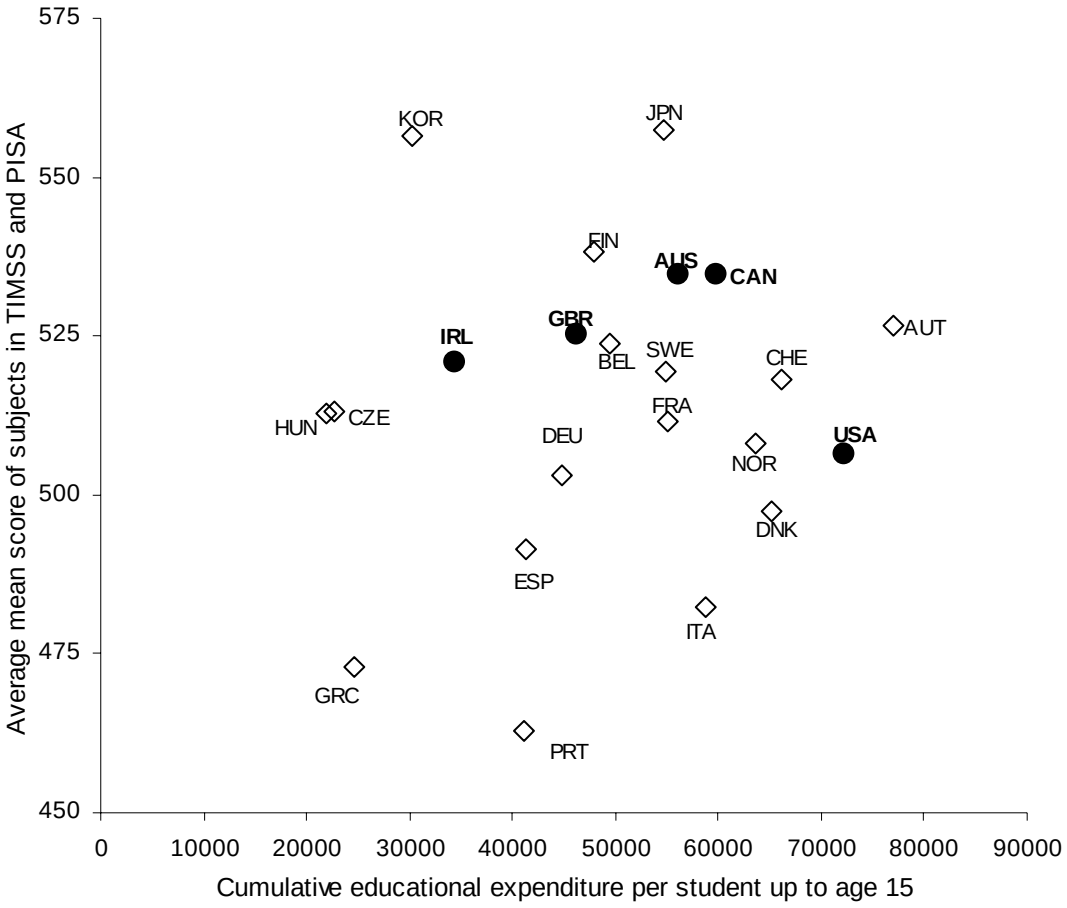
Note: Both 7th and 8th grade children are included in the TIMSS regressions. Estimated standard errors are in brackets and allow for clustering of children within schools. Results of the other explanatory variables used in the model are given in the Tables A1 and A3 in the appendix. For PISA and TIMSS summary statistics see Tables A2 and A4. For TIMSS, data for New Zealand, Canada, Australia and Ireland are for 1995, data for the UK refer to Scotland 1995 and England 1999, and USA to 1999.

Table 6: Difference in predicted scores between children from two family backgrounds (points and number of SDs)

	PISA maths				TIMSS maths			
	Points			SDs	Points			SDs
	(A)	(B)	Diff.		(A)	(B)	Diff.	
USA	581	433	148	1.51	574	468	105	1.20
New Zealand	619	512	107	1.08	546	460	86	0.96
UK	655	555	100	1.09	560	461	99	1.19
Australia	594	507	87	0.96	535	444	91	0.97
Canada	586	508	78	0.92	576	522	54	0.66
Ireland	512	435	78	0.92	511	443	69	0.83
France	653	575	78	0.87	576	512	64	0.99
Germany	614	484	130	1.27	602	522	80	1.01
Sweden	534	460	75	0.80	541	475	66	0.94

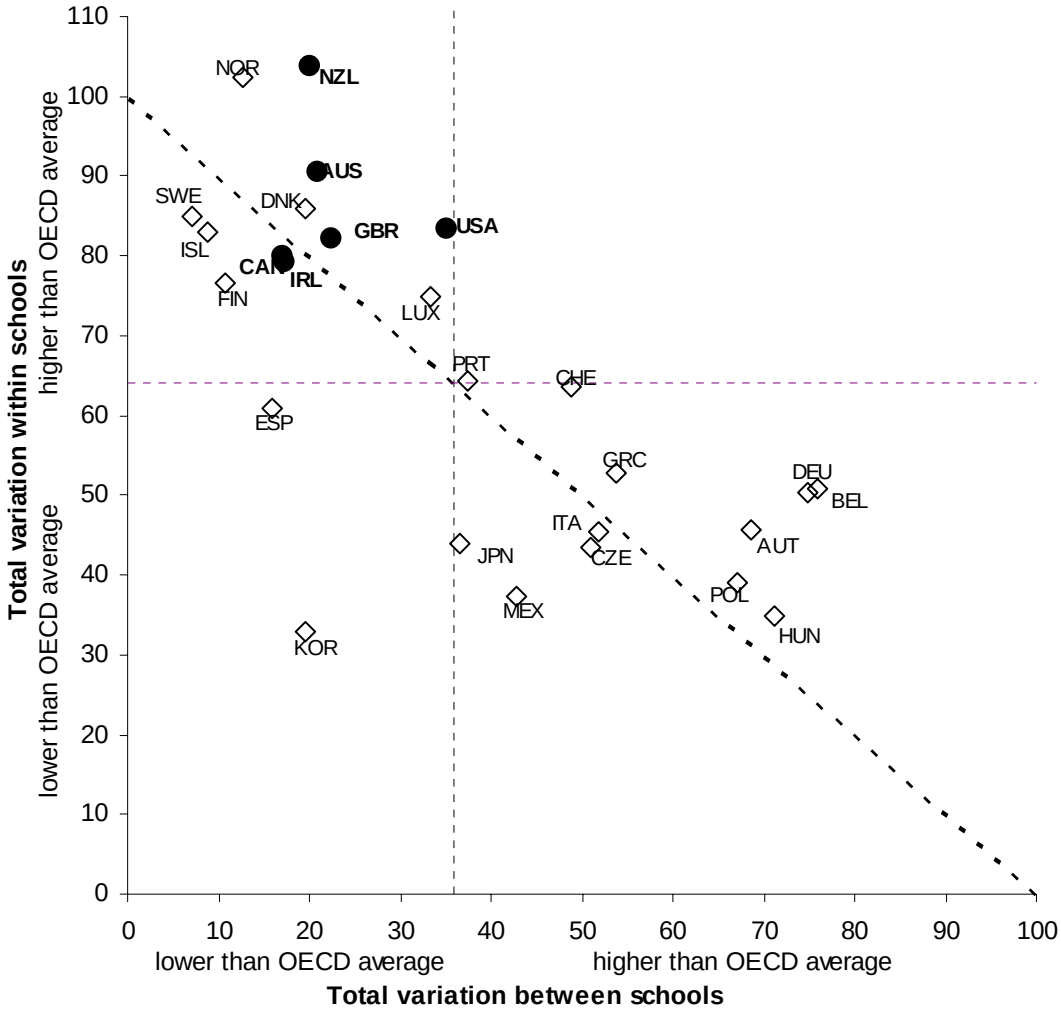
Note: the two family backgrounds are (A) nuclear family, mother has tertiary education, and more than 100 books at home and (B) single-parent family, mother did not complete secondary education, and up to 100 books at home. For English speaking countries the standard deviations of predicted scores for group A children are 47.4 and for group B children 48.1 in PISA and 25.0 (group A) and 29.0 (group B) in TIMSS. SDs shows what the difference in scores between the two family types represents in terms of the national standard deviation. National standard deviation for UK refers solely to England 1999.

Figure 6: Educational expenditures and average mean score in five subjects of TIMSS (8th grade) and PISA



Note: information on educational expenditures (which include both public and private) from OECD (2001: 264)

Figure 7: Variation within and between schools for PISA reading

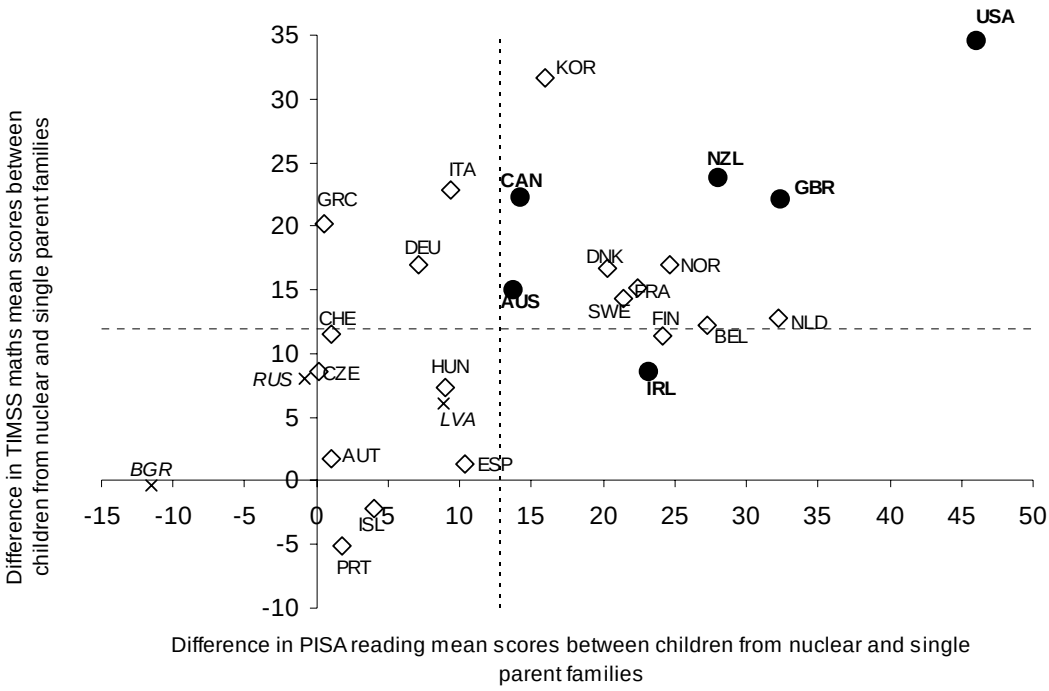


Source: OECD and UNESCO Institute for Statistics (2003: 357)

Note: Variation expressed as a percentage of average variation in reading scores across OECD countries. Countries below the diagonal line display a lower total variation than the OECD average. The vertical and horizontal dashed lines show the OECD average within and between school variations expressed as a percentage of average total OECD variation.

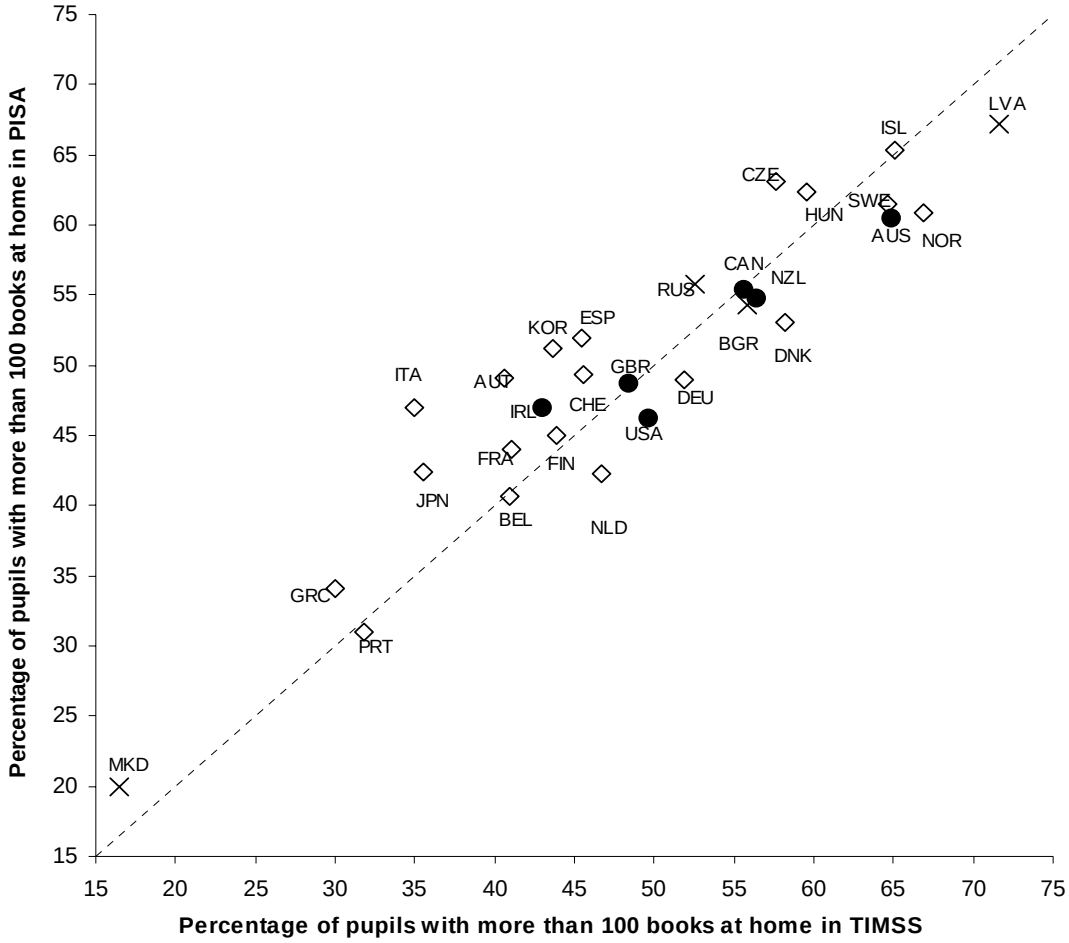
Appendix

Figure A1: Difference in scores between children from nuclear and single parent families in PISA reading and TIMSS maths



Source: See Table 4.

Figure A2: Percentage of children reporting more than 100 books at home in PISA and TIMSS



Note: correlation coefficient is 0.93.

Table A1: PISA maths OLS regression results

	Australia	Canada	Ireland	New Zealand	UK	USA	France	Germany	Sweden
Female pupil	-14.8	-11.2	-11.8	-3.5	-12.6	-9.4	-15.9	-22.3	-11.9
	(4.9)	(1.8)	(4.2)	(4.8)	(4.4)	(4.9)	(3.4)	(3.7)	(3.3)
Sibling	-19.5	-14.6	-18.7	-25.4	-24.3	1.1	-3.4	2.9	1.5
	(7.9)	(3.2)	(7.6)	(9.5)	(6.2)	(8.1)	(5.5)	(4.5)	(7.7)
First generation migrant	0.4	-3.7	21.0	-26.4	-22.7	-13.0	-30.5	-43.5	-38.3
	(6.0)	(3.7)	(15.2)	(10.2)	(9.0)	(14.1)	(6.7)	(6.8)	(8.4)
Non-native	-7.8	-12.7	71.1	-7.3	-32.9	-13.8	-52.9	-56.4	-51.0
	(7.1)	(4.8)	(13.5)	(6.7)	(15.2)	(7.7)	(12.2)	(9.1)	(9.8)
Rural location	-16.7		-4.6	-9.3	2.3	-19.3	-26.3	-13.2	-8.1
	(6.4)		(4.6)	(6.3)	(5.4)	(8.1)	(7.9)	(8.1)	(4.2)
Data location missing			17.2		-3.7	3.9	4.8	-17.5	8.1
			(4.5)		(9.0)	(12.1)	(10.6)	(11.7)	(15.9)
Student/teacher ratio	0.9		3.7	0.4	-2.8	-0.2	-5.7	-1.2	1.4
	(1.4)		(1.4)	(1.4)	(1.3)	(1.0)	(1.2)	(0.7)	(0.9)
Lack of school inputs	-5.9	-5.8	-4.0	-5.4	-1.1	-13.1	-18.6	-12.4	-2.5
	(3.9)	(2.3)	(2.7)	(5.5)	(2.6)	(9.5)	(5.7)	(4.2)	(2.9)
Constant	514.2	522.5	448.5	530.4	576.3	465.4	591.8	496.5	476.5
	(22.3)	(4.3)	(20.8)	(19.0)	(23.9)	(18.6)	(17.3)	(16.4)	(14.3)
Observations	2785	16041	2080	1958	4971	1894	2428	2689	2419
R-squared	0.13	0.10	0.15	0.17	0.18	0.27	0.25	0.28	0.14

Note: standard errors in parentheses (adjusted for clustering of children within schools)

Table A2: PISA Maths summary statistics

	Australia	Canada	Ireland	New Zealand	UK	USA	France	Germany	Sweden
Maths score	533	533	503	537	529	493	517	490	510
Gender	0.46	0.50	0.51	0.49	0.50	0.51	0.52	0.51	0.50
Sibling	0.96	0.94	0.97	0.95	0.94	0.95	0.92	0.88	0.96
Single parent family	0.16	0.15	0.13	0.21	0.20	0.20	0.14	0.14	0.16
Other family type	0.11	0.13	0.04	0.14	0.13	0.25	0.12	0.11	0.14
First generation immigrant	0.11	0.10	0.01	0.06	0.06	0.07	0.09	0.05	0.04
Non-native	0.12	0.10	0.01	0.13	0.02	0.05	0.02	0.10	0.06
Mother secondary education	0.68	0.82	0.55	0.65	0.72	0.77	0.61	0.64	0.76
Mother tertiary education	0.29	0.48	0.24	0.39	0.33	0.30	0.28	0.12	0.43
Data education missing	0.04	0.03	0.03	0.21	0.10	0.12	0.07	0.14	0.06
Books in household	0.60	0.56	0.46	0.56	0.49	0.46	0.44	0.50	0.62
Rural location	0.14	0.00	0.59	0.24	0.27	0.27	0.26	0.31	0.48
Data location missing	0.00	1.00	0.01	0.00	0.08	0.23	0.11	0.11	0.02
Student/teacher ratio in school	13.69		15.05	12.50	16.06	14.97	12.50	17.84	12.64
Lack of school inputs	0.08	0.13	0.52	0.12	0.55	0.08	0.14	0.52	0.31
Sample size	2859	16490	2128	2048	5195	2136	2597	2830	2464

Table A3: TIMSS maths regressions

	Australia	Canada	Ireland	New Zealand	UK	USA	France	Germany	Sweden
Female pupil	5.8	-0.3	-6.9	-3.0	-16.4	-10.5	-5.7	-7.3	-1.5
	(3.0)	(2.1)	(5.5)	(3.0)	(4.6)	(2.1)	(1.8)	(3.0)	(1.6)
Sibling	3.4	4.4	1.5	5.6	-4.4	0.7	0.8	0.4	5.4
	(3.0)	(2.6)	(4.0)	(2.9)	(4.2)	(2.5)	(2.6)	(2.6)	(2.2)
First generation	-11.0	-8.9	14.3	-24.3	7.6	-9.7		-24.5	-12.1
Migrant	(4.6)	(4.2)	(11.1)	(4.8)	(6.2)	(5.7)		(8.0)	(5.4)
Non-native	7.1	-15.8	19.8	3.3	7.4	-27.3		-30.0	-37.7
	(5.1)	(6.1)	(11.1)	(5.4)	(9.6)	(7.2)		(7.1)	(5.8)
Rural location	-12.6	-2.9	2.7	-2.9	22.1	-10.5	9.0	-11.1	-9.6
	(10.2)	(6.9)	(7.7)	(6.5)	(10.5)	(6.0)	(5.7)	(11.9)	(4.6)
Data location	19.2	11.4	-0.5	-13.1	-5.4	-0.1	7.3	-12.9	-6.4
Missing	(9.4)	(11.3)	(14.0)	(11.0)	(8.7)	(8.2)	(10.2)	(9.0)	(6.3)
Student/teacher	2.8	-3.3	3.9	0.2	2.6	-1.1	-0.6	-3.8	3.4
Ratio	(2.4)	(1.0)	(1.1)	(0.7)	(3.2)	(1.3)	(2.5)	(2.3)	(1.5)
Lack of school	2.6	-2.6	-0.9	-1.2	-6.2	-4.8	-1.3	-5.2	-3.9
Inputs	(3.7)	(2.4)	(2.1)	(3.3)	(4.6)	(2.7)	(2.5)	(5.3)	(1.6)
Grade	-26.7	-28.7	-20.4	-35.2	-37.3		-34.7	-15.1	-52.2
	(3.3)	(2.9)	(4.6)	(5.8)	(4.6)		(3.3)	(2.6)	(3.5)
Constant	462.7	540.2	457.9	475.6	477.5	492.9	521.7	535.8	480.0
	(22.7)	(14.0)	(16.2)	(10.4)	(24.0)	(13.3)	(21.5)	(19.6)	(9.9)
Observations	12368	15823	6028	6627	8110	8217	5715	5257	8560
R-squared	0.18	0.12	0.17	0.19	0.17	0.20	0.19	0.22	0.25

Note: standard errors in parentheses (adjusted for clustering of children within schools). Data for New Zealand, Canada, Australia and Ireland are for 1995, data on UK refer to Scotland 1995 and England 1999, and USA to 1999.

Table A4: TIMSS maths summary statistics

	Australia	Canada	Ireland	New Zealand	UK	USA	France	Germany	Sweden
Maths score	505	506	508	484	493	502	510	492	510
Gender	0.51	0.50	0.52	0.47	0.49	0.50	0.50	0.51	0.49
Sibling	0.89	0.83	0.94	0.88	0.88	0.84	0.86	0.77	0.83
Single parent family	0.16	0.17	0.09	0.16	0.16	0.19	0.11	0.14	0.12
Other family type	0.10	0.12	0.04	0.12	0.13	0.20	0.11	0.10	0.12
First generation migrant	0.10	0.10	0.01	0.06	0.08	0.08		0.05	0.04
Non-native	0.09	0.08	0.01	0.08	0.03	0.06		0.07	0.05
Mother secondary education	0.47	0.61	0.50	0.52	0.06	0.73	0.29	0.27	0.47
Mother tertiary education	0.22	0.35	0.12	0.20	0.02	0.46	0.13	0.05	0.20
Data education missing	0.21	0.21	0.19	0.32	0.91	0.19	0.51	0.34	0.41
Books in household	0.67	0.58	0.44	0.66	0.48	0.50	0.40	0.50	0.65
Rural location	0.14	0.15	0.23	0.17	0.12	0.20	0.32	0.12	0.19
Data location missing	0.15	0.15	0.04	0.05	0.18	0.18	0.10	0.40	0.08
Student/teacher ratio	8.55	10.45	11.56	11.65	8.26	9.13	7.69	8.58	6.27
Lack of school input	0.44	0.54	1.17	0.54	0.58	0.51	1.00	0.42	0.69
Grade	0.51	0.50	0.50	0.49	0.09	0.00	0.51	0.51	0.33
Sample size	12,852	16,581	6,203	6,867	8,736	9,072	6,014	5,763	8,855

Note: Data for New Zealand, Canada, Australia and Ireland are for 1995, data for UK refer to Scotland 1995 and England 1999, and USA to 1999.