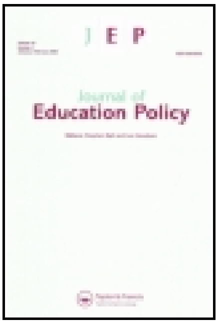
This article was downloaded by: [Institute of Education]
On: 22 September 2013, At: 03:34
Publisher: Routledge
Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Journal of Education Policy
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/tedp20

The mathematics skills of school children: how does England compare to the high-performing East Asian jurisdictions?

John Jerrim[a] & Álvaro Choi[b]

[a] Department of Quantitative Social Science, Institute of Education, University of London, London, UK.

[b] Institut d'Economia de Barcelona, University of Barcelona, Barcelona, Spain.

Published online: 11 Sep 2013.

To cite this article: John Jerrim & Álvaro Choi , Journal of Education Policy (2013): The mathematics skills of school children: how does England compare to the high-performing East Asian jurisdictions?, Journal of Education Policy, DOI: 10.1080/02680939.2013.831950

To link to this article: http://dx.doi.org/10.1080/02680939.2013.831950

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the "Content") contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms &

Journal of Education Policy, 2013
http://dx.doi.org/10.1080/02680939.2013.831950

The mathematics skills of school children: how does England compare to the high-performing East Asian jurisdictions?

John Jerrim[a]* and Álvaro Choi[b]

[a]*Department of Quantitative Social Science, Institute of Education, University of London, London, UK;* [b]*Institut d'Economia de Barcelona, University of Barcelona, Barcelona, Spain*

(*Received 17 January 2013; final version received 29 July 2013*)

The Programme for International Student Assessment (PISA) and Trends in Mathematics and Science Study are two highly respected studies of school pupils' academic achievement. English policy-makers have been disappointed with school children's performance on these tests, particularly in comparison to the strong results of young people from East Asia. In this paper, we provide new insight into the England–East Asia gap in school children's mathematics skills. We do so by considering how cross-national differences in math test scores change between ages 10 and 16. Our results suggest that, although average math test scores are higher in East Asian countries, this achievement gap does not increase between ages 10 and 16. We thus conclude that reforming the secondary school system may not be the most effective way for England to 'catch up' with the East Asian nations in the PISA math rankings. Rather, earlier intervention, during pre-school and primary school, may be needed instead.

Keywords: PISA; TIMSS; educational policy; primary education; secondary education

1. Introduction

One of the major developments in educational research over the last 20 years has been the widespread implementation of cross-national studies of pupil achievement, including the Programme for International Student Assessment (PISA), Trends in Mathematics and Science Study (TIMSS) and Progress in International Reading Literacy Study (PIRLS). These aim to produce cross-nationally comparable information on children's abilities at a particular age in at least one of three areas (reading, math and science). Regular reports are then published by the survey organisers where countries are ranked in terms of school children's test performance. This has had a major impact upon policy-makers from a number of countries, with many treating these international 'league tables' as an evaluation of their school system's success. English policy-makers have shown particular concern over England's 28th position, out of 65 countries, in the PISA 2009 mathematics assessment. Although a few northern European countries have fared rather better (e.g. Finland), it is the consistently strong performance of East Asian nations that has really caught policy-makers' attention.[1] For instance, in the PISA 2009 mathematics study, Shanghai was ranked top, Singapore 2nd, Hong Kong 3rd, Korea 4th, Taiwan 5th

---

*Corresponding author. Email: J.Jerrim@ioe.ac.uk and Japan 9th. Given the important role of human capital in economic productivity and growth (Barro 2001; Hanushek and Wößmann 2008; OECD 2010), England has looked towards the strong performance of these countries with an envious glare. Indeed, it is now widely believed that if England does not raise the academic skill of its school children, then its long-run prosperity will suffer as a result.

This has led policy-makers to consider what can be learnt from the East Asian nations to help English educational standards improve. For instance, the Secretary of State for Education Michael Gove recently stated that[2]:

> These regions and nations – from Alberta to *Singapore*, Finland to *Hong Kong*, Harlem to *South Korea* – should be our inspiration. [Emphasis our own]

With agreement from the shadow Education minister Stephen Twigg[3]:

> we must learn from high-performing nations like Japan.

Similarly, the East Asian nations have been highlighted as strong education systems in the ongoing review of England's mathematics curricula (Department for Education 2011), with an implicit suggestion that at least some of their school practices and policies hold the key to England's future educational success. Table 1 illustrates this point still further, where educational and economic inputs are compared to educational outputs across England and four comparator countries (Japan, Hong Kong, Singapore and Taiwan). Despite similar levels of GDP per

Table 1. Key characteristics of the UK, Japan, Singapore, Hong Kong and Taiwan.

	Japan	Hong Kong	Singapore	Taiwan	United Kingdom
1. GDP per capita (PPP 2005 US $000)	32.0	36.3	47.3	28.7	33.4
2. % GDP spent on education (2009)	3.8	4.8	3.1	4.1	5.4
3. Enrolment rates in pre-primary education (%)	90	97	–	29	81
4. Enrolment rate: primary education (%)	100	92	–	98	100
5. Enrolment rates: secondary education (%)	99	76	–	95	96
6. Enrolment rate: higher education (%)	59	57	–	82	59
7. Mean PISA math score (2009)	529 (94)	555 (95)	562 (104)	543 (105)	492 (87)
8. Mean PISA reading score (2009)	520 (100)	533 (84)	526 (97)	495 (86)	494 (95)
9. Mean PISA science score (2009)	539 (100)	549 (87)	542 (104)	520 (87)	514 (99)

Sources: 1. Pennworld tables (https://pwt.sas.upenn.edu/php_site/pwt_index.php).
2–6 World development indicators and Taiwan, from Ministry of Education. Data refers to 2009.
7–9 PISA survey website (http://www.oecd.org/pisa/46643496.pdf).
Notes: 7–9 Standard deviation presented in brackets.
Data is only typically available from the above sources for the UK as a while. However, PISA 2009 test scores can be provided for England and Scotland separately. These are 493 (87) and 499 (93) for math; 495 (95) and 500 (94) for reading and 515 (99) and 514 (96) for science.

capita, public expenditure on education and school enrolment rates, educational outcomes towards the end of secondary school (as measured by PISA test scores) are significantly lower in England.

It is, therefore, surprising that we do not know more about the achievement gap between England and the high-performing East Asian nations. Although insightful, studies such as PISA are often considered in isolation, providing a limited snapshot of children's abilities at one particular point in time. It would perhaps be more useful for academics and policy-makers to understand the specific point(s) in the education system that England falls behind these world leaders, and whether this is being driven by the experiences of certain sub-groups. For instance, the math skills of English and East Asian children could be roughly equal at the end of primary school, but then markedly diverge during secondary school. In this situation, reform of secondary education would perhaps be the most obvious policy response. On the other hand, it could be that most of the England–East Asia achievement gap emerges early in children's life (e.g. differences are apparent even by age 10) and that cross-national differentials do not grow much further beyond this point. Indeed, as the evidence base currently stands, one cannot rule out the possibility that English children actually catch up with their East Asian peers during secondary school. In this situation, resources and efforts for reform might be better concentrated at earlier points in children's life (e.g. before their 10th birthday). It would also suggest that analysis of studies such as PISA, which focuses upon the latter stages of secondary school, would be of little use in revealing why young people in East Asia are so much better at math than young people in England.

The aim of this paper is to thus develop a better understanding of how children's performance on internationally standardised math tests changes between ages 10 and 16, comparing the experiences of English children to those from the four aforementioned East Asian jurisdictions (Japan, Singapore, Taiwan and Hong Kong). This is, in our opinion, a vital first step towards identifying why children in East Asian countries outperform their English peers. Within this broad topic, we consider the following three specific issues.

Firstly, we illustrate how mean math test scores change with age. This is important for identifying the point(s) in the education system that English children fall behind young people in other countries (on average) and thus where efforts for school reform should be concentrated. Secondly, we investigate inequality in educational outcomes, and how the distribution of math skill changes between ages 10 and 16. Our initial focus will be upon the spread of achievement, and whether this widens or narrows in England relative to the four East Asian countries. This is followed by an assessment of whether the gap between the highest achieving children in England and highest achieving children in East Asia widens (or declines) during secondary school. This is a particularly prominent policy issue, as having a pool of very highly skilled individuals is vital for technological innovation and long-run economic growth (Bean and Brown 2005; Toner 2011). Finally, we consider an output-based measure of equality of educational opportunity, focusing upon math test score differentials between socio-economically advantaged and disadvantaged groups (a topic of much recent academic and political debate). Previous research has found that the socio-economic achievement gradient widens in England between the end of primary school and the end of secondary school (Goodman, Sibieta, and Washbook 2009; Ermisch and Del Bono 2012), but that the same is not true in other English-speaking countries (Ermisch, Jantti, and Smeeding 2012). However, there has been little work considering this issue using the TIMSS and PISA data-sets, and how England compares with the high performing East Asian jurisdictions in this respect. We make this important contribution to the existing literature.

Our results suggest that, although average math test scores are higher in East Asian countries than England, differences do not seem to increase between the end of primary and the end of secondary school. However, the gap between the highest achieving school children in East Asia and the highest achieving school children in England does seem to widen between ages 10 and 16. We also find that the vast majority of the socio-economic achievement gradient in mathematics skills in England is already apparent by age 10.

The paper now proceeds as follows. In Section 2, we describe our empirical methodology and the TIMSS and PISA data-sets. Section 3 provides estimates of change in test scores between ages 10 and 16 for England and a series of comparator countries. This is followed in Section 4 by a discussion of our findings and a series of policy recommendations.

2. Data

The aim of this paper is to examine the variation in children's math skills across countries, and how this changes between the end of primary school and the end of secondary school. Ideally, longitudinal data would be available, enabling one to track the progress of exactly the same children over time. Unfortunately, cross-nationally comparative data of this type does not exist. The next best alternative is to use repeated cross-sectional data, where samples have been collected from the same, or very similar, cohorts of school children at various points in time. From such data, one can draw inferences about the distribution of children's math skill at several ages, and thus how key points on the achievement distribution (e.g. mean, standard deviation, 10th percentile, 90th percentile, etc.) change between the end of primary school and the end of secondary school. The approach we take in this paper is to compare how these key statistics change across countries.

To do so, we draw upon data from the following rounds of the PISA and TIMSS studies:

- The 4th grade (age 9/10) TIMSS wave from 2003.
- The 8th grade (age 13/14) TIMSS wave from 2007.
- The PISA (age 15/16) wave from 2009.

Each of these resources collects nationally representative data and has been explicitly designed to facilitate comparisons of children's cognitive skills across countries (Olson, Martin, and Mullis 2008; OECD 2011 provide further information). They also have similar sample designs, with schools firstly selected as the primary sampling unit and then either one or two classes (TIMSS) or 35 pupils (PISA) randomly chosen to participate (from within each school). In all the analysis that follows, we account for this clustering of children within schools by making the appropriate adjustment to the estimated standard errors (using either the STATA 'svy' survey command or by bootstrapping standard errors by cluster).[4] Response rates for the countries included in our analysis can be found in Appendix 2. In most of the countries considered, school response was around 80 and 90%, while pupil response typically stood at over 90%.[5] In all three studies, the survey organisers have produced a set of weights which attempt to correct for bias induced by non-response, while also scaling the sample up to the size of the national population. These weights are applied throughout the analysis.[6]

A notable feature of the three studies is that they collect data for children who were born at approximately the same time.[7] For instance, the two TIMSS studies for England refer to children who were born between September 1992 and August 1993, while those who took part in PISA 2009 were in the school year below (born between September 1993 and August 1994).[8] Consequently, one can track the performance of a very similar cohort of children at three different ages (9/10, 13/14 and 15/16). This is important if one wishes to interpret the changes observed as 'age' rather than 'cohort' effects. Although discussion shall focus on the performance of England relative to a set of leading East Asian nations, we include 13 countries that took part in each of these three studies into our analysis. This includes six from the rich western world (England, Scotland, Australia, Italy, USA and Norway), four Asian 'tiger' economies (Hong Kong, Japan, Singapore and Taiwan) and three with middle incomes (Lithuania, Russia and Slovenia).[9] Some additional commentary shall be presented regarding England's performance relative to this broader set of countries.

It is important to recognise that there are some limitations with this empirical strategy. Firstly, although each study examines children's ability in mathematics, there are some conceptual differences in the skills being measured. For instance, whereas TIMSS focuses upon children's ability to meet internationally agreed curricula, PISA examines functional ability – how well young people can use the skills in 'real life' situations. Whether this slight difference in focus is of substantive importance is, however, questionable. For example, the correlation between children's PISA math test scores and a curricula-based measure in England (key stage 3 scores) is high at over 0.80 (Micklewright and Schnepf 2006). Moreover, in Appendix 3, we also show the strong correlation ($r = 0.88$) between mean PISA and TIMSS 8th grade test scores at the country level. Nevertheless, one still cannot rule out the possibility that there are at least some subtle differences in the precise skills being measured. We also note that other studies have used the PISA, PIRLS and TIMSS data to investigate the change in children's math skills as they age. For instance, Ammermueller (2006) and Waldinger (2007) use these data to investigate whether the impact of family background on children's test scores increases between primary and secondary education. Similarly, Hanushek and Wößmann (2006) and Jakubowski (2010) use PISA, PIRLS and TIMSS data to investigate whether school tracking leads to greater inequality in educational achievement (though producing conflicting results). These previous studies do not jeopardise our original contribution, as they do not explicitly consider the academic performance of English school children relative to their East Asian peers.[10] However, they do illustrate how the PISA, PIRLS and TIMSS data have been used to investigate cross-country changes in children's test scores over time.

Secondly, there are some differences between the surveys in the test score metric generated. In all three studies, children's responses to the test questions are combined into a set of possible overall test scores via an item-response model.[11] Five 'plausible values' are then created for each child; these are five separate estimates of children's ability in mathematics. The intuition behind this process is that children's true ability cannot be observed, and must be estimated from their answers on the test. This results in a measure of children's achievement that has a mean of 500 and standard deviation of 100 in all three studies.[12] However, each of the surveys contains a different pool of countries upon which these achievement scores are based. For instance, while PISA includes all members of the OECD, the two TIMSS studies do not. Consequently, although the test metric across the three surveys appears to be on the same scale, figures are not actually directly comparable (e.g. a mean score of 500 in PISA is not the same as a mean score of 500 in TIMSS).

To overcome this problem, all test score data are transformed (within each survey) into international *z*-scores. In other words, scores have been normalised at the pupil level, so that in each survey the mean is 0 and the standard deviation is 1 across the 13 countries considered. This is a standard method for obtaining comparable units of measurement for variables that are on different scales and is similar to the approach taken by Brown et al. (2007) in their comparison of the PISA and TIMSS data-sets. One implication of this is that estimates refer to English pupils' test performance relative to that of children in the 13 other countries. Thus, our focus is upon how England's performance relative to other countries changes between primary and secondary school. Terms like 'relative decline' shall, therefore, be used as international *z*-scores are comparative measures.

Similar difficulties arise when one considers the availability and comparability of children's background characteristics. For instance, the TIMSS studies contain very little information on pupils' socio-economic status. This poses a problem for estimating the socio-economic gradient in mathematics achievement, and whether this gradient steepens as children age. We, therefore, turn to what many consider to be the best available proxy for family background that is contained within each of the three data-sets and measured in a comparable way – the number of books in the family home.[13] Sociologists (e.g. Evans et al. 2010) have argued that this reflects the scholarly culture of a household, and is thus a measure of the educational environment in which a child is being raised. On the other hand, various economists have argued that books in the home are 'the single most important predictor of student performance in most countries' (Wößmann 2008) and that there is evidence that this is a cross-nationally comparable proxy for socio-economic position (Schütz, Ursprung, and Wößmann 2008; Hanushek and Wößmann 2010).[14] It has been widely used in this manner by various academics in the analyses of the PISA, PIRLS and TIMSS data-sets (Waldinger 2007; Wößmann 2008; Schütz, Ursprung, and Wößmann 2008; Ammermueller and Pischke 2009; Machin 2009; Evans et al. 2010; Jakubowski 2010; Hermann and Horn 2011; Brunello, Weber, and Weiss 2012) including investigations of how the socio-economic gradient changes with age across countries (Ammermueller 2006; Jerrim and Micklewright 2012a). Using the PISA 2009 data, we have found modest statistically significant correlations between books in the home and various measures of socio-economic status (treating all categorical data as continuous, linear variables). This was 0.37 in the case of maternal education, 0.38 for paternal education, 0.37 for social class and 0.45 for a measure of cultural possessions. Moreover, Jerrim and Micklewright (2012b) investigate possible measurement error in children's reports of books in the home, by comparing their responses to those of their parents. A cross-tabulation of child and parent reports of books in the home for England using the PIRLS data, drawn directly from Jerrim and Micklewright 2012b, can be found in Appendix 4. Although there is a positive association between the two sets of reports, agreement is far from complete. We thus proceed with caution, acknowledging books in the home to be an imperfect proxy for socio-economic status, though one which has been widely used in the data sources under our investigation.

In each data-set, we use this variable in a series of OLS regression models to estimate how inequality of educational opportunity varies across countries. This takes the form:

$$\mathrm{A}_{ijk} = \alpha + \beta_1 \cdot \mathrm{SES}_i + \beta_2 \cdot \mathrm{Sex}_i + \beta_3 \cdot I_i + \beta_4 \cdot \mathrm{SES}_i \times I_i + \varepsilon_{ij} \quad \forall k \tag{1}$$

where A = Children's score on the TIMSS or PISA math test; Sex = A binary indicator of the child's gender (0 = female, 1 = male); I = Whether the child is a first or second generation immigrant (0 = Native, 1 = Immigrant); SES = A set of four dummy variables reflecting the number of books in the family home (Reference: less than 25 books); i = child i. j = child j; k = country k.

This specification follows the existing literature on international comparisons of socio-economic achievement gradients (e.g. Schütz, Ursprung, and Wößmann 2008; Wößmann 2008; Jerrim 2012). Socio-economic status (as measured by books in the home) is the covariate of interest, with controls included for gender and whether the child was a first- or second-generation immigrant. As argued by Wößmann (2008), other characteristics (e.g. type of school attended) are intentionally not controlled, so that the SES parameter proxies all the channels by which family background influences children's test performance.[15] The estimated coefficients will thus proxy the *cumulative* impact of family background on children's test performance, including their experiences during the first years of life (which Cunha, Heckman, and Lochner 2006, amongst others, have stressed are extremely important). During this paper, we focus upon test score differences between the most advantaged (more than 200 books) and least advantaged (less than 25 books) groups. Our primary interest is: (a) how does this socio-economic achievement gradient vary across countries and (b) how does the gradient change as children move from the end of primary school to the end of secondary school.

Given the data difficulties described above, our analysis shall proceed with some caution. Specifically, our strategy is to treat the TIMSS 4th grade survey as a broad indicator of children's math skills towards the end of primary school (when children are aged 9/10) with the TIMSS 8th grade and PISA 2009 studies as two separate indicators of math skills towards the end of secondary school. Our intention is thus to look for evidence of robust changes in math achievement (at the country level) that hold whether either TIMSS 2007 (8th grade) or PISA 2009 is used as the secondary school follow-up survey.

3. Results

3.1. Average test scores

In Table 2, countries are ranked by mean test scores at ages 9/10, 13/14 and 15/16. The countries of interest are highlighted in shades of light (England) or dark (East Asia) grey. At each point, England sits in the middle of the cross-country ranking, with average test scores roughly in-line with those achieved by children from the USA. Indeed, on no occasion can one reject the null hypothesis that average test scores in England are significantly different from zero at the 5% level. In other words, England's performance is always roughly in-line with the cross-national average (within this pool of 13 countries). A particularly notable feature of Table 2

Table 2. Average math test scores at ages 9/10, 13/14 and 15/16 (international *z*-scores).

	Age 9/10			Age 13/14			Age 15/16	
	Mean	SE		Mean	SE		Mean	SE
Singapore	0.820*	0.062	Taiwan	0.904*	0.050	Singapore	0.729*	0.056
Hong Kong	0.570*	0.037	Singapore	0.844*	0.062	Hong Kong	0.644*	0.053
Japan	0.446*	0.023	Hong Kong	0.599*	0.078	Taiwan	0.521*	0.059
Taiwan	0.435*	0.026	Japan	0.571*	0.040	Japan	0.371*	0.057
Lithuania	0.064	0.041	England	–0.072	0.064	Australia	0.215*	0.032
Russia	0.037	0.051	Russia	–0.103*	0.041	Slovenia	0.070	0.060
England	0.035	0.053	USA	–0.130*	0.035	Norway	0.032	0.025
USA	–0.128*	0.034	Lithuania	–0.166*	0.038	Scotland	0.004	0.047
Italy	–0.326*	0.047	Slovenia	–0.219*	0.025	England	–0.016	0.041
Australia	–0.375*	0.052	Australia	–0.278*	0.057	USA	–0.077	0.046
Scotland	–0.473*	0.042	Scotland	–0.379*	0.054	Italy	–0.121*	0.029
Slovenia	–0.623*	0.033	Italy	–0.477*	0.035	Lithuania	–0.193*	0.042
Norway	–0.959*	0.031	Norway	–0.596*	0.023	Russia	–0.285*	0.038

Notes: (1) *indicates where average test scores are statistically different from 0 at the 5% level. This illustrates whether average math test scores are significantly different from the 13 country cross-national average. All standard errors take into account the clustering of children within schools.

(2) Age 9/10 refers to TIMSS 2003 4th grade data, age 13/14 is TIMSS 2007 8th grade and age 15/16 PISA 2009.

(3) All figures presented are international *z*-scores.

is that the East Asian nations are consistently at the top of the international rankings, with a sizeable gap between this group and all other countries included in the analysis. For instance, even when children are in primary school (age 9/10), there is a big difference (almost 0.4 of an international standard deviation) between the lowest performing East Asian country (Taiwan) and the highest performing other country included in the sample (Lithuania). Thus, a substantial and statistically significant cross-national achievement gap has emerged long before the start of secondary school.

England is clearly quite some distance behind the leading East Asian nations (in terms of pupils' average math achievement) before children reach their 10th birthday. But, do English children fall further behind during secondary school? The answer to this question can be found in Table 3. This provides the change in average test scores between ages 10 and 14 (left hand columns) and 10 and 16 (right hand columns) across the 13 countries. The column labelled 'Sig Diff to 0' indicates whether there is a statistically significant change in a country's performance relative to the cross-national average between the two ages (based upon a two sample *t*-test with independence between surveys). On the other hand, the column labelled 'Sig Diff to Eng' illustrates whether there is a significant improvement or decline in average test scores relative to the change observed within England (based upon a two sample t-test assuming independence between countries within each of the surveys). This has similarities to a classic difference-in-difference test, where change in one 'treatment group' over time (e.g. English secondary schooling and culture) is compared to the change in other 'treatment groups' (e.g. various form of East Asian secondary schooling and culture).

Starting with England, notice that the change in mean test scores between both ages 10 and 14 (−0.107) and 10 and 16 (−0.051) are small and statistically indistinguishable from 0 at conventional thresholds. Thus, there is little evidence that the math skills of English children either improve or deteriorate (relative to young people in our pool of 13 countries) between the end of primary school and the end of secondary school (on average). This is in contrast to some countries (Norway and Slovenia) where average test scores clearly increase, while in others (Lithuania and Russia) there is a marked decline. Yet, there is also little to suggest that English pupils fall further behind children in the leading East Asian nations. For instance, notice that the change in mean test scores between ages 10 and 16 in England is not significantly different to that in any of the East Asian countries. A similar result also seems to hold for Scotland (if anything, children in this country may actually catch up with their East Asian peers). This point is further emphasised in Figure 1, which plots mean test scores for the countries of interest at the three ages. Although the gap between England and the four East Asian countries is always large (often half an international standard deviation or more), there is no consistent evidence that the gap widens or declines during the primary (age 9/10) to secondary (age 13/14 or 15/16) transition.

A clear implication for policy-makers is that it is not during secondary school that the leading East Asian countries pull away from England in terms of school pupils' math skills. Rather, the causal factor(s) behind these countries' strong performance seemingly occurs much earlier in life (i.e. before the age of 9/10) and this relative advantage is then maintained. Consequently, reforming the secondary school system may not be the most effective way for England to 'catch up' with such countries in the PISA rankings. Earlier intervention (e.g. during pre-school and primary school)

Table 3. Change in average math test scores between primary and secondary school.

	Change 10–14				Change 10–16			
	Change	SE	Sig Diff to 0	Sig Diff Eng	Change	SE	Sig Diff to 0	Sig Diff Eng
Norway	0.363	0.038	***	***	0.991	0.039	***	***
Slovenia	0.404	0.042	***	***	0.693	0.068	***	***
Australia	0.097	0.078	–	*	0.590	0.062	***	***
Italy	−0.151	0.058	***	–	0.205	0.055	***	***
Scotland	0.094	0.068	–	*	0.477	0.063	***	***
Taiwan	0.469	0.056	***	***	0.086	0.064	–	–
Hong Kong	0.030	0.086	–	–	0.074	0.065	–	–
USA	−0.002	0.049	–	–	0.051	0.058	–	–
England	−0.107	0.083	–	–	−0.051	0.067	–	–
Japan	0.125	0.046	***	**	−0.075	0.061	–	–
Singapore	0.024	0.088	–	–	−0.091	0.084	–	–
Lithuania	−0.230	0.056	***	–	−0.257	0.058	***	***
Russia	−0.140	0.066	**	–	−0.322	0.063	***	***

Notes: (1) *, ** and *** indicate statistical significance at the 10, 5 and 1% level. 'Sig Diff to 0' illustrates whether the change in average math test scores are significantly different from the change for the 13 country cross-national average (using a two sample *t*-test with independent surveys). 'Sig Diff Eng' illustrates whether the change in average math test scores are significantly different from the change seen in England (using a two sample *t*-test assuming independence between countries). All standard errors take into account the clustering of children within schools.

(2) The left-hand columns refer to the change in average math test scores between age 10 (TIMSS 2003 data) and age 14 (TIMSS 2007 data). The right-hand columns refer to the change in average math test scores between age 10 (TIMSS 2003 data) and age 16 (PISA 2009 data).

(3) All figures presented are international *z*-scores.

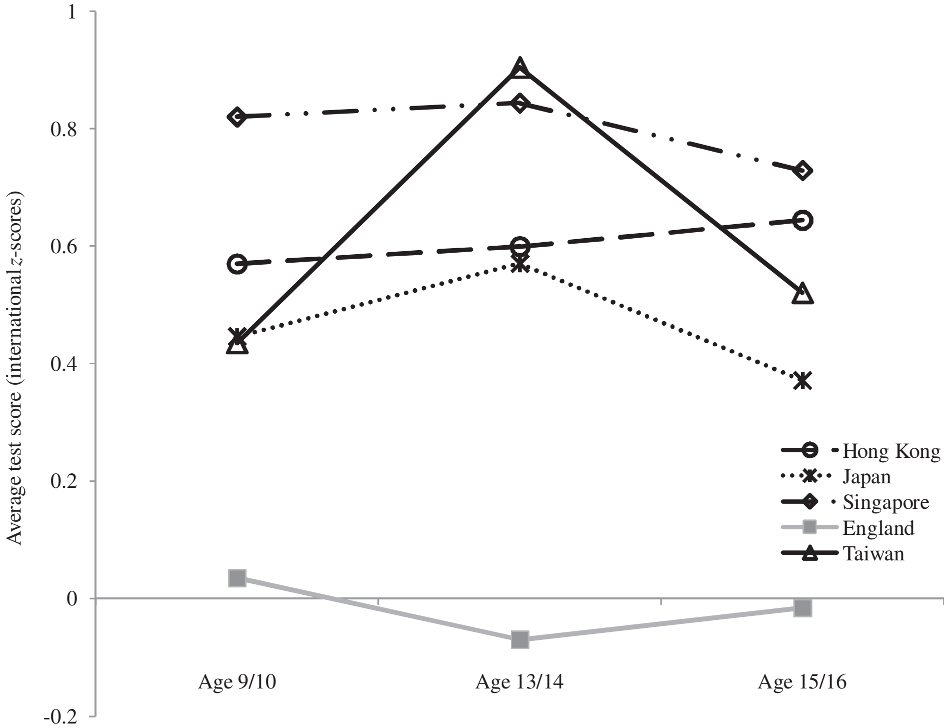


Figure 1. Average math test scores at ages 9/10, 13/14 and 15/16 – England compared to a selection of East Asian countries.
Notes: (1) Age 9/10 refers to TIMSS 2003 4th grade data, age 13/14 refers to TIMSS 2007 8th grade data and age 15/16 refers to PISA 2009. (2) All figures presented in terms of international *z*-scores.

may be needed instead. Moreover, it seems unlikely that analysis of data-sets that focus upon the latter stages of secondary school (like PISA) will be able to explain why average math performance is so much higher in East Asia than England.

3.2. Inequality in educational outcomes

Although England's relative performance in terms of pupils' average math test scores may not change significantly between primary and secondary school, it is possible that the distribution of achievement could alter as children age. Evidence on this matter can be found in Figure 2. This plots the standard deviation of children's math test scores at ages 9/10, 13/14 and 15/16 (note that Ferreira and Gignoux 2011 consider several possible measures of inequality in educational outcomes and conclude that the standard deviation is the most appropriate when analysing the international achievement datasets). England is highlighted using a light grey line with square markers.

At age 9/10, inequality in mathematics achievement stands at roughly 1.1 international standard deviations in England. This is notably higher than in the East Asian nations, with the standard deviation being only 0.9 in Japan and less than 0.8 in Hong Kong and Taiwan. Yet, this situation seems to reverse towards the end of secondary school; whereas inequality in mathematics achievement falls in England (to 0.95 of an international standard deviation by age 16), it increases in a number

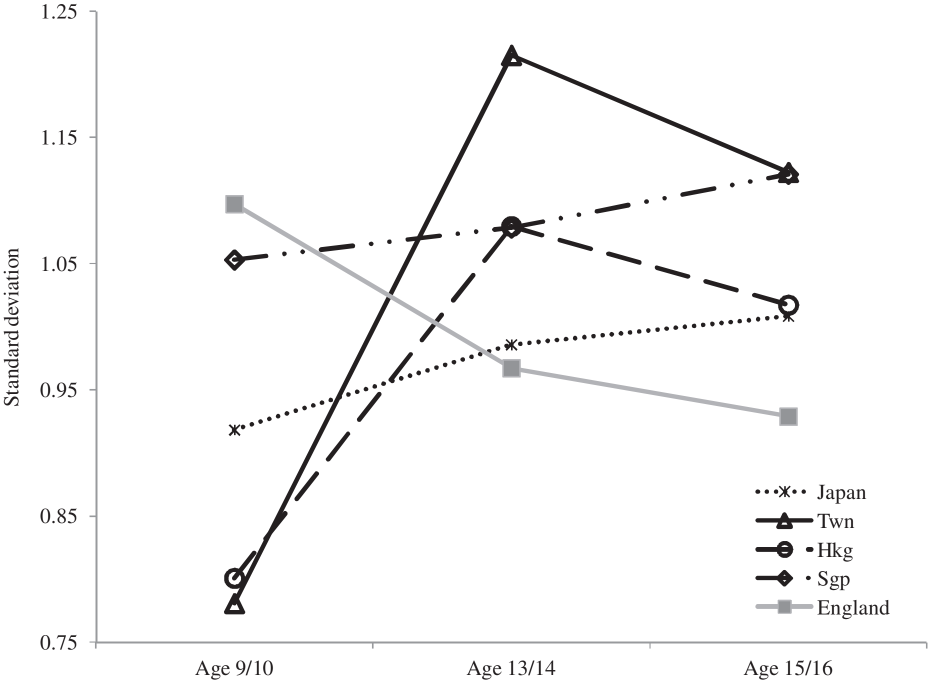


Figure 2. Standard deviation of mathematics test scores (inequality in math outcomes). Notes: (1) Age 9/10 refers to TIMSS 2003 4th grade data, age 13/14 refers to TIMSS 2007 8th grade data and age 15/16 refers to PISA 2009. (2) All figures presented in terms of international *z*-scores.

of East Asian countries (e.g. it is up from 0.80 at age 10 to 1.02 at age 16 in Hong Kong). Thus, although there is little change in average test scores between ages 10 and 16, the same does not appear to be true with regards to educational inequality. In particular, whereas mathematics achievement seems to become more equal in England during secondary school, in the East Asian countries it becomes more dispersed.[16]

What is behind this apparent change in educational inequality? Table 4 panel A presents the 10th percentile of the achievement distribution at the three ages. This reflects the math skills of the lowest achieving pupils within each of the 13 nations. Figure 3 illustrates how the 10th percentile changes between primary and secondary school for England and East Asian nations. The left-hand side refers to the age 10–14 comparison and the right-hand side the age 10–16 comparison. The thin black line running through the centre of the bars represents the estimated 90% confidence interval. Interestingly, there is some evidence of an increase in P10 within England, particularly for the age 10 to age 16 comparison. In other words, the low achievers in England manage to improve relative to low achievers in other countries. The opposite is true, however, in Singapore, Hong Kong and Taiwan, where P10 declines (e.g. in Hong Kong P10 declines from −0.48 at age 10 to −0.72 at age 16). Consequently, one can see that between primary school and the end of secondary school, the gap between the lowest achieving children in England and the lowest achieving children in East Asian countries is reduced. This is consistent with government policy in England during this period, when a number of initiatives

Table 4. The estimated 10th and 90th percentile of the math test score distribution at ages 9/10, 13/14 and 15/16 (international z-scores).

	Age 9/10			Age 13/14			Age 15/16	
	P10	SE		P10	SE		P10	SE
(a) Test scores at the 10th percentile								
Hong Kong	−0.48	0.048	Singapore	−0.66	0.093	Hong Kong	−0.72	0.076
Singapore	−0.56	0.111	Japan	−0.70	0.041	Singapore	−0.79	0.051
Taiwan	−0.57	0.034	Taiwan	−0.84	0.078	Japan	−0.92	0.076
Japan	−0.76	0.029	Hong Kong	−0.93	0.198	Taiwan	−0.94	0.050
Lithuania	−1.21	0.071	USA	−1.28	0.048	Australia	−1.08	0.034
Russia	−1.22	0.062	Slovenia	−1.32	0.031	Norway	−1.14	0.032
USA	−1.39	0.044	Lithuania	−1.36	0.042	England	−1.23	0.044
England	−1.41	0.052	Russia	−1.39	0.050	Slovenia	−1.24	0.043
Italy	−1.69	0.053	England	−1.38	0.074	Scotland	−1.27	0.049
Australia	−1.73	0.076	Australia	−1.48	0.084	USA	−1.33	0.056
Scotland	−1.74	0.060	Italy	−1.61	0.056	Italy	−1.40	0.040
Slovenia	−1.93	0.037	Norway	−1.61	0.031	Lithuania	−1.41	0.040
Norway	−2.30	0.049	Scotland	−1.63	0.052	Russia	−1.44	0.037

	Age 9/10			Age 13/14			Age 15/16	
	P90	SE		P90	SE		P90	SE
(b) Test scores at the 90th percentile (international z-scores)								
Singapore	2.08	0.064	Taiwan	2.33	0.054	Singapore	2.12	0.074
Hong Kong	1.56	0.045	Singapore	2.14	0.067	Hong Kong	1.90	0.050
Japan	1.58	0.043	Hong Kong	1.85	0.061	Taiwan	1.94	0.064
Taiwan	1.41	0.027	Japan	1.80	0.067	Japan	1.67	0.060
England	1.41	0.048	England	1.15	0.072	Australia	1.51	0.040
Russia	1.29	0.081	Russia	1.12	0.043	Slovenia	1.41	0.078
Lithuania	1.21	0.044	USA	1.02	0.041	Scotland	1.29	0.060
USA	1.07	0.031	Lithuania	1.02	0.049	Norway	1.21	0.033
Italy	0.94	0.063	Australia	0.91	0.096	England	1.20	0.052

(*Continued*)

Table 4. (*Continued*).

	Age 9/10			Age 13/14			Age 15/16	
	P90	SE		P90	SE		P90	SE
(b) Test scores at the 90th percentile (international z-scores)								
Australia	0.86	0.059	Slovenia	0.86	0.032	USA	1.20	0.050
Scotland	0.74	0.052	Scotland	0.81	0.052	Italy	1.14	0.026
Slovenia	0.56	0.051	Italy	0.62	0.044	Lithuania	1.01	0.057
Norway	0.31	0.028	Norway	0.37	0.019	Russia	0.87	0.059

Notes: (1) Age 9/10 refers to TIMSS 2003 4th grade data, age 13/14 is TIMSS 2007 8th grade and age 15/16 PISA 2009.
(2) All figures presented are international *z*-scores.
(3) All standard errors take into account the clustering of children within schools. See Appendix 1 for further details on calculation of standard errors.

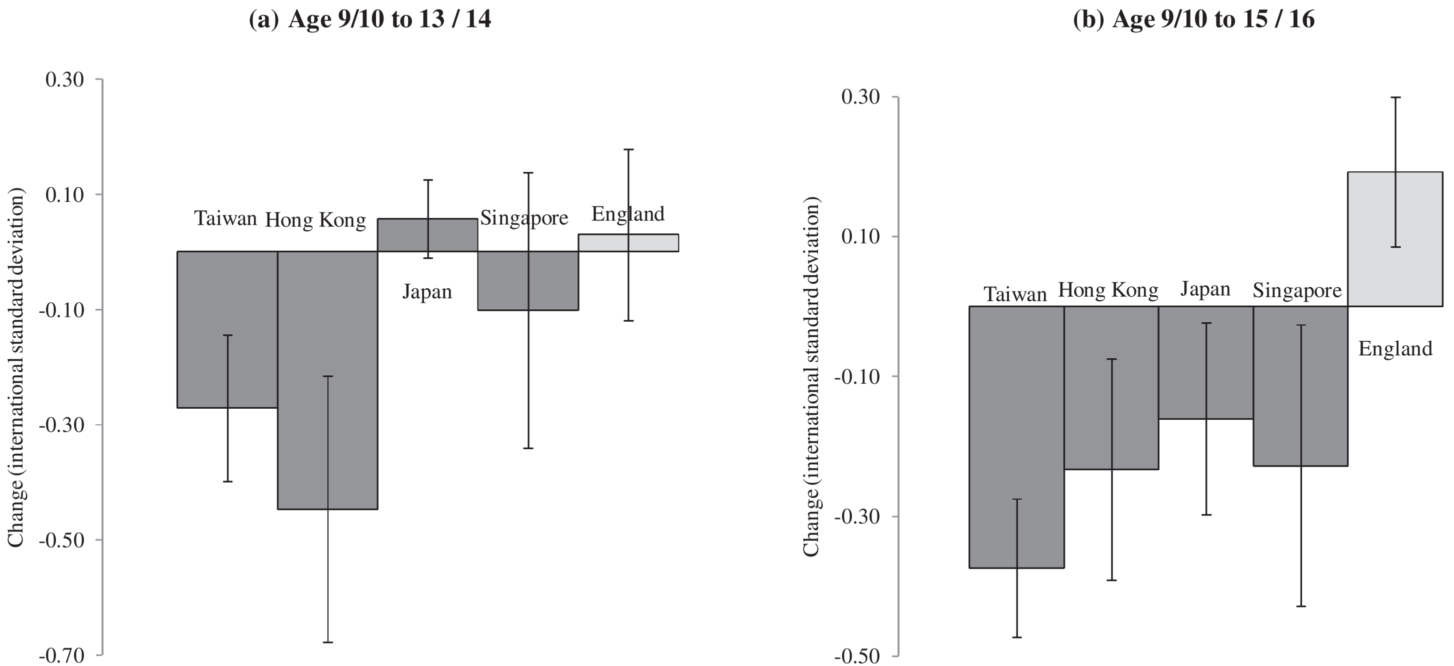


Figure 3. Change in the 10th percentile of the math test distribution between the end of primary school and secondary school.

Notes: (1) The left-hand panel refers to the change in the 10th percentile of math achievement between age 9/10 (TIMSS 2003 4th grade) and age 13/14 (TIMSS 2007 8th grade). The right-hand panel provides analogous figures for the change between age 9/10 (TIMSS 2003 4th grade) and age 15/16 (PISA 2009).

(2) All figures presented in terms of international *z*-scores.

(3) The thin black line running through the centre of each bar is the estimated 90% confidence interval. All standard errors take into account the clustering of children within schools. See Appendix 1 for further details on methodology.

attempted to raise the basic skills of low-achieving groups. However, it should be noted that, despite this progress, a significant gap remains between the lowest achievers in England and the lowest achievers in East Asia, even at age 16.[17]

Does the same hold true for the highest achieving children? In Table 4 panel b, we provide analogous results for the 90th percentile of the math achievement distribution (i.e. the test performance of the highest achieving children within each of the countries). Figure 4 then compares the change in the 90th percentile between the end of primary school and the end of secondary school. Unfortunately, it seems that England does lose some ground relative to its international competitors (and particularly the East Asian nations) in this respect. The bars in both the left- and right-hand panel of Figure 4 are negative for England, with the estimated 90% confidence interval not crossing zero. The implication is that the test score gap between the highest achieving children in England and the highest achieving children in the 12 other countries we consider increases.[18] On the other hand, the opposite is true in several of the East Asian countries – the highest achieving pupils tend to further extend their lead. For instance, Table 4(b) reveals that the 90th percentile in Hong Kong moves from 1.56 standard deviations above the cross-country mean at age 10 to 1.90 standard deviations at age 16. Pulling these results together, Figure 4 suggests that the gap between the highest achieving children in England and the highest achieving children in East Asia increases between the end of primary school and the end of secondary school.

3.3. *Inequality of educational opportunity*

Finally, we turn to the issue of inequality of educational opportunity, defined as the difference in math test scores between high (more than 200 books) and low (25 or fewer books) socio-economic groups. Table 5 provides estimates at the three ages. It becomes immediately apparent that England has a particularly large socio-economic achievement gradient when measured in this way. For instance, at age 9/10, children from advantaged backgrounds score (on average) 0.87 standard deviations more on the TIMSS math test than children from disadvantaged backgrounds. This is bigger than any other country included in the analysis at the 5% level (with the exception of Singapore). Moreover, no country has a significantly bigger socio-economic achievement gap than England at either age 13/14 or age 15/16. It is also interesting to note that there is no common pattern across the East Asian countries, with quite large socio-economic differences occurring regularly in some (e.g. Singapore, Taiwan) but not in others (e.g. Hong Kong).

Does the socio-economic test score gradient increase between ages 10 and 16 in England? Evidence on this issue can be found in Figure 5. This plots the socio-economic test score gap at the three ages. Children from advantaged backgrounds do indeed extend their lead over their disadvantaged peers in England, as has been found in previous research (Goodman, Sibieta, and Washbook 2009; Ermisch, Jantti, and Smeeding 2012). Although this increase of 0.26 of a standard deviation (from 0.87 at age 10 to 1.13 at age 16) is statistically significant ($t = 2.5$, $p = 0.01$)[19] and of reasonable magnitude, the vast majority of socio-economic inequality in educational achievement is nevertheless apparent by age 10. Moreover, Figure 5 would seem to suggest that the socio-economic gradient also increases in the four East Asian countries by roughly the same (Singapore, Japan) or even greater (Hong Kong, Taiwan) amounts.

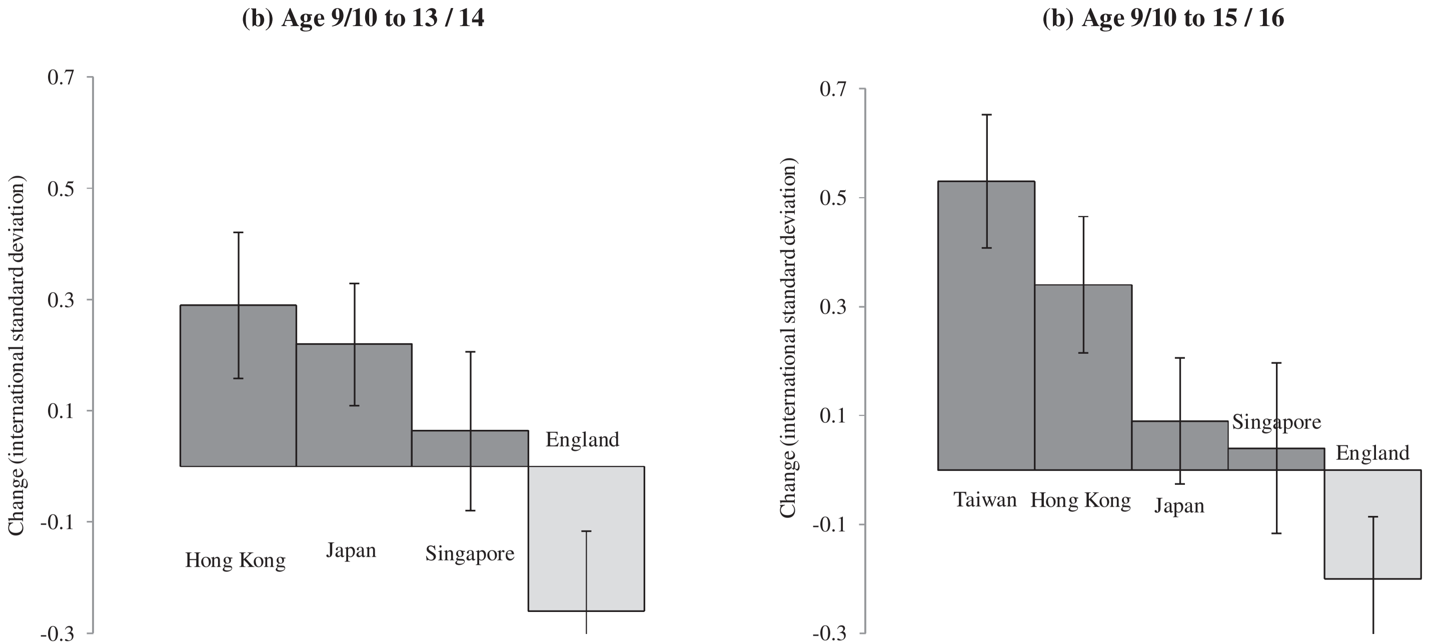


Figure 4. Change in the 90th percentile of the math test distribution between the end of primary school and secondary school.
Notes: (1) The left-hand panel refers to the change in the 90th percentile of math achievement between age 9/10 (TIMSS 2003 4th grade) and age 13/14 (TIMSS 2007 8th grade). The right-hand panel provides analogous figures for the change between age 9/10 (TIMSS 2003 4th grade) and age 15/16 (PISA 2009).
(2) All figures presented in terms of international *z*-scores. The thin black line running through the centre of each bar is the estimated 90% confidence interval. All standard errors take into account the clustering of children within schools. See Appendix 1 for further details on methodology.
(3) Results for Taiwan have been excluded from the left hand panel for clarity of presentation. The 90th percentile is estimated to increase by 0.9 of a standard deviation between age 10 and age 14 in this country (see Table 4 panel b for further details).

Table 5. Socio-economic differences in children's math test scores at age 10, 14 and 16 (international z-scores).

	Age 10		Age 14		Age 16	
	Beta	SE	Beta	SE	Beta	SE
Singapore	0.882	0.079	0.939	0.093	0.999	0.082
England	0.871	0.087	1.141	0.082	1.126	0.052
USA	0.680*	0.044	0.782*	0.043	1.051	0.057
Australia	0.612*	0.095	0.853*	0.079	1.063	0.041
Scotland	0.582*	0.080	0.963	0.075	1.211	0.078
Japan	0.566*	0.062	0.722*	0.070	0.692*	0.057
Taiwan	0.533*	0.039	1.046	0.063	1.047	0.066
Norway	0.513*	0.074	0.648*	0.038	1.007	0.045
Lithuania	0.489*	0.075	0.798*	0.067	0.948*	0.061
Russia	0.422*	0.093	0.677*	0.074	0.700*	0.063
Hong Kong	0.335*	0.088	0.652*	0.116	0.865*	0.083
Slovenia	0.266*	0.084	0.686*	0.051	1.130	0.096
Italy	0.115*	0.077	0.637*	0.057	0.913*	0.045

Notes: (1) Authors' calculations based upon the regression model presented in Section 2.
(2) Figures refer to the difference in average test scores between children with few (0–25) vs. children with many (more than 200) books.
(3) All figures presented in terms of international z-scores.
(4) *indicates where socio-economic gradient significantly different to England at the 5% level. All standard errors take into account the clustering of children within schools.

Thus, although we have replicated previous findings of an increasing socio-economic achievement gradient between ages 10 and 16 in England, we have also presented evidence that suggests the same holds true in the leading East Asian nations. Moreover, as discussed in Section 2, one cannot rule out the possibility that the increasing socio-economic achievement gradient found in all countries may be the result of the 9/10 year olds in the TIMSS 2003 data struggling to provide a good estimate of the number of books in the family home. Nevertheless, as inequality in educational achievement is already large before children finish primary school, this further suggests that public investment into increasing opportunities for young people from disadvantaged homes may be best placed in the early years (Cunha, Heckman, and Lochner 2006).

4. Discussion, policy recommendations and conclusions

The PISA and TIMSS are two highly respected studies of school pupils' academic achievement. Policy-makers have shown great interest in their findings – particularly, the dominance of East Asian countries towards the top of the PISA and TIMSS rankings. The Secretary of State for Education, Michael Gove, and his shadow, Stephen Twigg, have both suggested that England must learn lessons from these high-performing jurisdictions, including policies that could be successfully implemented in this country. We have provided some guidance on this issue by attempting to identify the age at which children in England are overtaken by their peers in East Asia (in terms of average mathematics test scores), and thus where efforts to reform the schooling system should be concentrated. Although our statistical analysis has not attempted to establish cause and effect, our findings and survey of the wider literature do lead us to the following three policy suggestions.

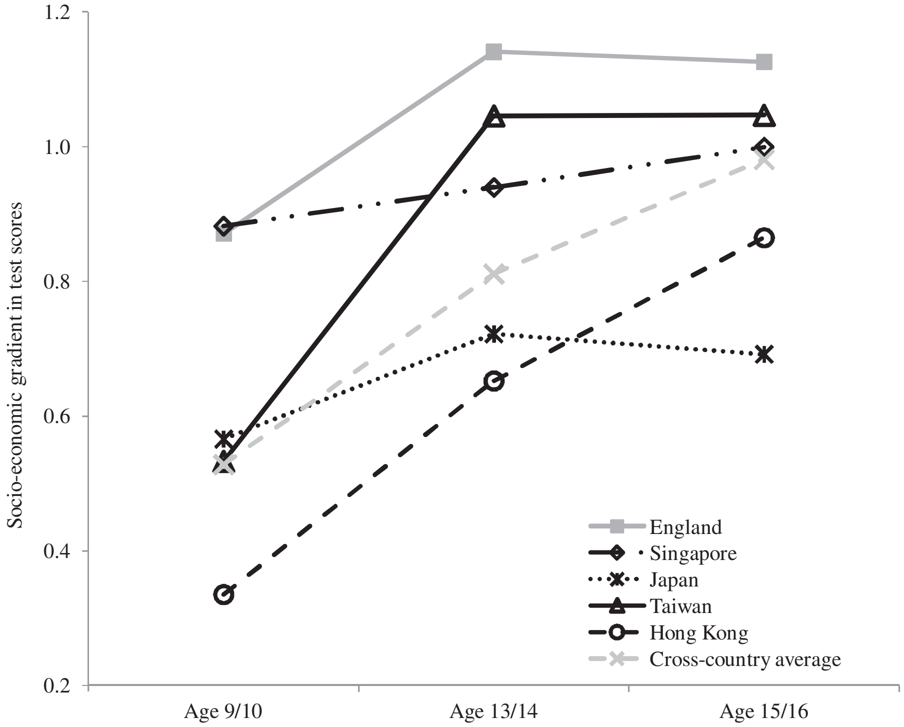


Figure 5. Socio-economic inequality in math test scores at ages 9/10, 13/14 and 15/16.
Notes: (1) Estimates refer to differences between children from households with few books (0–25) to those with many books (more than 200 books.
(2) Age 9/10 refers to TIMSS 2003 4th grade data, age 13/14 refers to TIMSS 2007 8th grade data and age 15/16 refers to PISA 2009.
(3) All figures presented in terms of international *z*-scores.

Firstly, policy-makers might concentrate on reforming mathematics education in the early primary and pre-school years. This paper has shown how there is a large gap in math achievement between England and leading East Asian nations even at age 10, but also that this gap does not appreciably widen during secondary school. Thus, despite major policy focus on secondary schools, there is little evidence that these institutions are responsible for England's disappointing position in the PISA and TIMSS rankings. What policies from East Asian countries could England adopt to boost math skills before the end of primary school? Unfortunately, the answer does not seem to be straightforward. One might suggest that there is a need for government to provide more (and higher quality) pre-school care, as there is evidence that this has a positive impact upon children's later academic achievement (Cunha, Heckman, and Lochner 2006). However, pre-school enrolment rates are already higher in England than Japan, Singapore, Taiwan and Hong Kong (recall Table 1). Moreover, although we are unable to compare pre-school quality, it is interesting to note that the OECD has recently suggested that certain East Asian nations should learn lessons from the UK in this respect (Taguma et al. 2012). Investment in education also seems unlikely to be the cause, as the percentage of GDP per capita spent on education has been consistently lower in the East Asian countries than the UK during the 1999–2009 periods (World Bank 2012). Primary school class sizes also tend to be larger in East Asia and instructional hours lower (OECD 2011). However, one factor that does notably differ is the quality and status of teachers. For instance, teachers in East Asia tend to be high academic performers (OECD 2011), and have a duty to study and research as well as teach (Jensen et al. 2012). Moreover, they receive high earnings both in comparative international terms and relative to other professional groups. Although establishing the causal impact of this higher pay and status is beyond the scope of this paper, we do suggest that raising the prestige of teaching (particularly at the primary school level) could be an important lever upon which English policy-makers may draw. Consequently, further research should be devoted to understanding the impact of different school-level educational resources.

Our second policy suggestion is that further investment in the skills of children from disadvantaged backgrounds is needed, again with a focus on the primary and pre-school years. Section 3.3 illustrates that the socio-economic gradient in math test scores seems to be steeper in England than East Asian countries. While this gap may widen slightly in England during secondary school, socio-economic differences in academic achievement are largely in place by age 10. Although some caution is required when interpreting this result, given the limitations of the data available (Jerrim and Micklewright 2012b), we note that our findings (and subsequent policy recommendation) are consistent with a host of other academic research (e.g. Cunha, Heckman, and Lochner 2006; Heckman 2007; Schütz, Ursprung, and Wößmann 2008; Jerrim 2012). As primary education is free or nearly free in England and most East Asian countries, alternative explanations for the large socio-economic achievement gradient in England should be sought. One possibility is that ability grouping in primary school mathematics classes is relatively common in England, but not East Asia (Boaler, Altendorff, and Kent 2011; OECD 2012).[20] As Gamoran (2004) and OECD (2012) note, there is little evidence that such streaming improves average performance, but may exacerbate test score differences between advantaged and disadvantaged groups. Similarly, between school selection processes are weaker in East Asian countries than England (OECD 2012), meaning that disadvantaged children are likely to have better access to quality educational resources. Reducing the segregation of pupils in England, both within and between primary schools, may thus make an important contribution to narrowing the socio-economic achievement gap in mathematics.

Finally, although we continue to suggest that policy-makers should focus on the earlier stages of young people's educational career, some important changes may be needed to improve aspects of mathematics provision during secondary school. A pressing issue is to ensure that the curriculum stretches the best young mathematicians enough, and that they are motivated (and incentivised) to fully develop their already accumulated academic skill. Evidence presented in this paper has suggested that the gap between the highest achieving children in England and the highest achieving children in East Asia widens between ages 10 and 16 (at least in mathematics). One possible explanation for this finding is the widespread use of private tuition by East Asian families for both remedial and enrichment purposes (Ono 2007; Sohn et al. 2010). This helps to boost the performance of all pupils, including those already performing well at school. In comparison, private tutoring in England is mainly undertaken by a relatively small selection of children from affluent backgrounds, often for remedial purposes. While a large proportion of East Asian families are willing to personally finance such activities through the private sector, the same is unlikely to hold true in the foreseeable future within England. Consequently, the state may need to intervene. Gifted and talented schemes, a shift of school and pupil incentives away from reaching floor targets (e.g. a C grade in GCSE mathematics) and enhanced tuition for children who excel in school are all possible policy responses.

These suggestions do, however, come with important caveats. Firstly, although it is true that most of East Asia's modern educational systems '*were strongly and deliberately modelled after the Western educational rubric* (Jeynes 2008, 900)', the identification of successful policies in some countries does not necessarily ensure the success of their implementation in others. Even when policies and teaching methods have been proven to be effective in East Asia, culture and context potentially limit the extent to which such initiatives can be successfully transferred to other countries (Cowen 2006). Secondly, it is worth underlining that cultural and social factors might be behind these countries' strong PISA and TIMSS test performance. In East Asian cultures, education has historically been considered a highly valued good and the main legitimate method for social mobility. This can be seen not only in the East Asian teachers' high salaries, but also by the heavy investment of families in private tutoring services. Family and social commitment to education is also reflected in the large number of weekly hours East Asian students spend in self-study activities and, as Zhu and Leung (2011) argue, the great impact extrinsic motivation has on their mathematics test performance (much more so than their Western peers). Consequently, the implementation of some of the characteristics of the East Asian educational model may imply the need for a cultural shift towards greater belief in the value of education amongst all and the importance of a hard work ethic. Indeed, it is important for academics and policy-makers to recognise that East Asian children vastly out-perform their English peers even when they have been through the English schooling system.[21] This is perhaps the clearest indication that it is actually what happens outside of school that is driving these countries superior PISA and TIMSS math test performance. We recognise, of course, that such cultural shifts cannot be expected to take place in England in the short run, as it is notoriously difficult to modify people's attitudes and beliefs. Similarly, although such policies can lead to higher academic performance, they have well-known side effects, such as the pressure which students (physical and psychological) and parents (financial) must put up with (Bray 2003). Yet, in an increasingly competitive world, such a cultural shift may be necessary to ensure England's future prosperity and long-run economic success.

Notes

1. Finland has only routinely taken part in the PISA study and not the other international assessments (e.g. PIRLS or TIMSS). On the other hand, a number of leading Asian economies (e.g. Hong Kong, Singapore) have participated in PISA, PIRLS and TIMSS for a number of years. It is the East Asian countries consistently strong performance (throughout various studies and numerous survey waves) that is perhaps most impressive.
2. See http://www.education.gov.uk/inthenews/inthenews/a0070008/secretary-of-state-comments-on-pisa-study-of-school-systems.
3. See http://www.bbc.co.uk/news/education-18057883.
4. When considering mean test scores (Section 3.1) and socio-economic gradients (Section 3.3), we use the 'svy' STATA survey command. When investigating percentiles of the test distribution (Section 3.2), estimates are bootstrapped by cluster (schools) using

50 replications to calculate approximate standard errors (see Appendix 1 for further details).

5. The school response rate we refer to is after replacement schools have been included.
6. Our experimentations with the data suggest that substantive conclusions remain intact whether the weights are applied or not.
7. The TIMSS studies collect information from children within the same school 'grade' (i.e. the same school year group), while in PISA children are all the same age (i.e. between 15 years 3 months and 16 years and 2 months old).
8. The target population in TIMSS is children within the same *school* year. In contrast, the PISA target population is children born within the same *calendar* year. This difference in target population forms part of the explanation why English children in the PISA 2009 sample were in the school year below those in the TIMSS 2007 (8th grade) and TIMSS 2003 (4th grade) samples.
9. For Scotland, we drop roughly 10% of observations when using the PISA 2009 data. These are children in 'grade 12' (S5). We have tested the robustness of our results to this sample selection, and find little change in the substantive conclusions drawn.
10. Indeed, the former head of the PISA study Andreas Schleicher has also anecdotally made such comparisons. See http://www.tes.co.uk/article.aspx?storycode=6307101.
11. A one-parameter Rasch model PISA is used to generate test scores in PISA while a three-parameter item scaling procedure is used in TIMSS.
12. In view of the large volume of data we are analysing, we use the first plausible value only throughout our analysis. OECD (2009, 129) note that 'analysing one plausible value instead of five plausible values provides unbiased population estimates'.
13. In a background questionnaire, children in PISA and TIMSS are asked about the number of books there are in their household, and instructed to tick the corresponding category.
14. For instance, Hanushek and Wößmann (2010) state 'Schütz, Ursprung, and Wößmann 2008 corroborate the cross-country validity of the books-at-home variable by showing that the association between household income and books at home does not vary significantly between the six countries for which both income and books measures are available in the PIRLS data-set'.
15. We recognise that there are many important factors influencing children's educational development, including early education, parental education, material support, attitudes and aspirations. One would, therefore, ideally include multiple variables into the analysis to capture these various effects. Unfortunately, the PISA and TIMSS data-set contain limited comparable information, and thus books in the home is used as the best available proxy to reflect the combined influence of such factors. A similar approach has been used by other authors when using these data (e.g. Ammermueller 2006; Wößmann 2008).
16. Here, we refer to inequality in educational outcomes (the spread of achievement) and not equality of opportunity (how achievement differs between socio-economic groups). The latter shall be the focus of the following sub-section.
17. This can be seen in the right-hand column of Table 4 panel A.
18. It is important to once again stress that in this paper, we are referring to relative differences between countries. Thus, although English children's maths ability will clearly improve between the end of primary school and the end of secondary school, this may be at a slower rate than their East Asian peers (and hence be in relative decline).
19. Statistical significance based upon a two-sample *t*-test with independent samples.
20. Hallam and Parsons (2012) show that one in six UK children are being taught in ability streams at age 7.
21. In 2011, 78.5% of Chinese children achieved five or more A*–C grades including math and English. This compares to a national average of 58.2%. See http://www.education.gov.uk/rsgateway/DB/SFR/s001057/index.shtml.

Notes on contributors

John Jerrim is a lecturer in Economics and Social Statistics at the Institute of Education, University of London. His research interests include the economics of education, access to higher education, intergenerational mobility, cross-national comparisons and educational inequalities. He has worked extensively with the OECD Programme for International

Student Assessment (PISA) data, with this research reported widely in the British media. He was the recipient of an ESRC Research Scholarship 2006–2010 and awarded the prize as the 'most promising PhD student in the quantitative social sciences' at the University of Southampton. In October 2011, he was awarded a prestigious ESRC post-doctoral fellowship to continue his research into the educational and labour market expectations of adolescents and young adults. Since then he has won the inaugural ESRC Early Career Outstanding Impact award and has just received an ESRC grant to study cross-national comparisons of educational attainment and social mobility.

Álvaro Choi is a lecturer in the Economics of Education at the Institut d'Economia de Barcelona.

References

Ammermueller, A. 2006. "Educational Opportunities and the Role of Institutions." ZEW Discussion Paper 05-44, 1–50.

Ammermueller, A., and J. Pischke. 2009. "Peer Effects in European Primary Schools: Evidence from the Progress in International Reading Literacy Study." *Journal of Labor Economics* 27 (3): 315–348.

Barro, J. 2001. "Human Capital and Growth." *American Economic Review* 91 (2): 12–17.

Bean, F., and S. Brown. 2005. "A Canary in the Mineshaft? International Graduate Enrollments in Science and Engineering in the United States." Accessed May 30, 2012. http://www.utexas.edu/cola/centers/european_studies/_files/pdf/immigration-policy-conference/bean.pdf

Boaler, J., L. Altendorff, and J. Kent. 2011. "Mathematics and Science Inequalities in the United Kingdom: When Elitism, Sexism and Culture Collide." *Oxford Review of Education* 37 (4): 457–484.

Bray, M. 2003. *Adverse Effects of Private Supplementary Tutoring: Dimensions, Implications and Government Responses*. Paris: International Institute for Educational Planning.

Brown, G., J. Micklewright, S. Schnepf, and R. Waldmann. 2007. "International Surveys of Educational Achievement: How Robust are the Findings?" *Journal of the Royal Statistical Society Series A* 170 (3): 623–646.

Brunello, G., G. Weber, and C. Weiss. 2012. "Books are Forever: Early Life Conditions, Education and Lifetime Income." IZA Discussion Paper 6386.

Cowen, R. 2006. "Acting Comparatively upon the Educational World: Puzzles and Possibilities." *Oxford Review of Education* 32 (5): 561–573.

Cunha, F., J. Heckman, and L. Lochner. 2006. "Interpreting the Evidence on Life Cycle Skill Formation." In *Handbook of the Economics of Education*, edited by E. Hanushek and F. Welch. Amsterdam: Holland North.

Department for Education. 2011. "Review of the National Curriculum in England: What Can We Learn from the English, Mathematics and Science Curricula of High-Performing Jurisdictions?" Accessed March 10, 2012. https://www.education.gov.uk/publications/standard/publicationDetail/Page1/DFE-RR178

Ermisch, J., and E. Del Bono. 2012. "Inequality in Achievements during Adolescence." In *Inequality from Childhood to Adulthood: A Cross-National Perspective on the Transmission of Advantage*, edited by J. Ermisch, M. Jantti, and T. Smeeding. New York: Russell Sage Foundation.

Ermisch, J., M. Jantti, and T. Smeeding. 2012. *Inequality from Childhood to Adulthood: A Cross-National Perspective on the Transmission of Advantage*. New York: Russell Sage Foundation.

Evans, M., J. Kelley, J. Sikora, and D. Treiman. 2010. "Family Scholarly Culture and Educational Success: Books and Schooling in 27 Nations." *Research in Social Stratification and Mobility* 28 (2): 171–197.

Ferreira, F., and G. Gignoux. 2011. "The Measurement of Educational Inequality: Achievement and Opportunity." IZA Discussion Paper 6161.

Gamoran, A. 2004. "Classroom Organization and Instructional Quality." In *Can Unlike Students Learn Together? Grade Retention, Tracking and Grouping*, edited by H. Walberg, A. Reynolds, and M. Wang. Greenwich: Information Age.

Goodman, A., L. Sibieta, and E. Washbook. 2009. "Inequalities in Educational Outcomes among Children Aged 3 to 16." Accessed February 6, 2012. http://sta.geo.useconnect.co.uk/pdf/Inequalities%20in%20education%20outcomes%20among%20children.pdf

Hallam, S., and S. Parsons. 2012. "Prevalence of Streaming in UK Primary Schools: Evidence from the Millennium Cohort Study." *British Educational Research Journal* 39 (3): 514–544.

Hanushek, E., and L. Wößmann. 2010. "The Economics of International Differences in Educational Achievement." IZA Discussion Paper 4925.

Hanushek, E., and L. Wößmann. 2006. "Does Educational Tracking Affect Performance and Inequality? Differences in Differences Evidence across Countries." *Economic Journal* 116 (510): C63–C76.

Hanushek, E., and L. Wößmann. 2008. "The Role of Cognitive Skills in Economic Development." *Journal of Economic Literature* 46 (3): 607–668.

Heckman, J. 2007. "The Economics, Technology, and Neuroscience of Human Capability Formation." *Proceedings of the National Academy of Sciences* 104 (3): 13250–13255.

Hermann, Z., and D. Horn. 2011. "How are Inequality of Opportunity and Mean Student Performance are Related? A Quantile Regression Approach Using PISA Data." IEHAS Discussion Papers 1124.

Jakubowski, M. 2010. "Effects of Tracking on Achievement Growth: Exploring Difference-in-Differences Approach to PIRLS, TIMSS and PISA Data." In *Quality and Inequality of Education*, edited by J. Dronkers. Netherlands: Springer.

Jensen, B., A. Hunter, J. Sonnemann, and T. Burns. 2012. "Catching Up: Learning from the Best School Systems in East Asia." Accessed November 11, 2011. http://grattan.edu.au/publications/reports/post/catching-up-learning-from-the-best-school-systems-in-east-asia/

Jerrim, J. 2012. "The Socio-Economic Gradient in Teenagers' Literacy Skills: How Does England Compare to Other Countries?" *Fiscal Studies* 33 (2): 159–184.

Jerrim, J., and J. Micklewright. 2012a. "Parental Socio-Economic Status and Children's Cognitive Achievement at Ages 9 and 15: How do the Links Vary Across Countries?" In *Inequality from Childhood to Adulthood: A Cross-National Perspective on the Transmission of Advantage*, edited by J. Ermisch, M. Jantti, and R. Erikson, New York: Russell Sage Foundation.

Jerrim, J., and J. Micklewright. 2012b. "Socioeconomic Gradients in Children's Cognitive Skills: Are Cross-Country Comparisons Robust to Who Reports Family Background?" DoQSS Working Papers 12–06.

Jeynes, W. 2008. "What We Should and Should Not Learn from the Japanese and Other East Asian Education Systems." *Educational Policy* 22 (6): 900–927.

Machin, S. 2009. "Inequality and Education." In *The Oxford Handbook of Economic Inequality*, edited by W. Salverda, B. Nolan, and T. Smeeding. Oxford: Oxford University Press.

Micklewright, J., and S. Schnepf. 2006. "Response Bias in England in PISA 2000 and 2003." Accessed June 23, 2011. https://www.education.gov.uk/publications/eOrderingDownload/RR771.pdf.

OECD. 2009. *PISA Data Analysis Manual: SPSS Second Edition*. 2nd ed. Paris: OECD.

OECD. 2010. *The High Cost of Low Educational Performance*. Paris: OECD.

OECD. 2011. *Education at a Glance 2011*. Paris: OECD.

OECD. 2012. *Equity and Quality in Education: Supporting Disadvantaged Students and Schools*. Paris: OECD.

Olson, J., M. Martin, and I. Mullis. 2008. *TIMSS 2007 Technical Report*. Boston, MA: Boston College.

Ono, H. 2007. "Does Examination Hell Pay Off? A Cost–Benefit Analysis of 'ronin' and College Education in Japan" *Economics of Education Review* 26 (3): 271–284.

Schütz, G., H. Ursprung, and L. Wößmann. 2008. "Education Policy and Equality of Opportunity." *Kyklos* 61 (2): 279–308.

Sohn, H., D. Lee, S. Jang, and T. K. Kim. 2010. "Longitudinal Relationship Among Private Tutoring, Student-Parent Conversation, and Student Achievement." *KEDI Journal of Educational Policy* 7 (1): 23–41.

Taguma, M., I. Litjens, J. H. Kim, and K. Malowiecki. 2012. *Quality Matters in Early Childhood Education and Care: Korea*. Paris: OECD.

Toner, P. 2011. “Workforce Skills and Innovation: An Overview of Major Themes in the Literature.” OECD Education Working Papers Number 55.

Waldinger, F., 2007. “Does Ability Tracking Exacerbate The Role of Family Background for Students’ Test Scores?” Accessed October 19, 2011. http://www2.warwick.ac.uk/fac/soc/economics/staff/academic/waldinger/research/ability_tracking.pdf.

Wößmann, L. 2008. “How Equal are Educational Opportunities? Family Background and Student Achievement in Europe and the United States.” *Zeitschrift für Betriebswirtschaft* 78 (1): 45–70.

World Bank. 2012. *World Development Indicators*. Washington, DC: World Bank.

Zhu, Y., and F. Leung. 2011. “Motivation and Achievement: Is There an East Asian Model?” *International Journal of Science and Mathematics Education* 9 (5): 1189–1212.

Appendix 1. Estimates at the 10th and 90th percentile – estimates and standard errors

Estimates of 10th and 90th percentile of the test distribution

In Table 4, we present estimates of the 10th and 90th percentile of the test scores distribution at three ages (10, 14 and 16) along with the associated standard errors. These can be produced in STATA by simply running a quantile regression at the given percentile (e.g. P90) with no covariates included. An example of the STATA code used for England at age 10 (TIMSS 2003 4th grade data) is given below:

```
qreg Z_Math_Score if Country==826, q(0.90)
```

where ‘*Z_Math_Score*’ = the internationally standardised math *z*-score.

The above provides an accurate estimate of the 90th percentile of the test distribution for England. However, standard errors are likely to be underestimated as the clustering of children within schools has not been taken into account. We, therefore, bootstrap by cluster (schools) all of the estimates produced for the 10th and 90th percentile (using 50 bootstrap replications). An example of the coding used can be found below:

```
program q90
  qreg Z_Math_Score if Country==826, q(0.90)
end
  bs, cluster(School_ID): q90
```

Estimates of change for the 10th and 90th percentile of the test distribution

In Figures 3 and 4, we present estimates of how the 10th and 90th percentile change as children age (either between ages 10 and 14 or between ages 10 and 16). One way to calculate these changes is to simply subtract the relevant values given in Table 4. For instance, the change in the 90th percentile for England between ages 10 and 14 is given by the calculation (1.15–1.41 = −0.26).

An alternative is to the pool the TIMSS 2003 4th grade (age 10) and TIMSS 2007 8th grade (age14) data into a single file (*post* standardisation) and to estimate a quantile regression model with an ‘age’ dummy variable as a covariate. An example of the STATA code used for England is as follows:

```
xi:qreg Z_Math_Score i.Age if Country==826, q(0.90)
```

where the age dummy takes a value of 1 to indicate TIMSS 2007 (age 14) data. The coefficient for the age dummy variable is thus equal to the change in the 90th percentile between ages 10 and 14 (this can be confirmed for England – where we get a value of −0.26 as explained above). However, standard errors are likely to be underestimated as the clustering of children within schools has not been taken into account. We, therefore, bootstrap by cluster (schools) all of the estimates produced for the 10th and 90th percentile (using 50 bootstrap replications). An example of the coding used can be found below:

```
program q90
  xi:qreg Z_Math_Score i.Age if Country==826, q(0.90)
end
  bs, cluster(School_ID): q90
```

Moreover, note that the above model can be extended to include an ‘Age’ by ‘Country’ interaction term:

```
xi:qreg Z_Math_Score i.Age*i.Country, q(0.90)
```

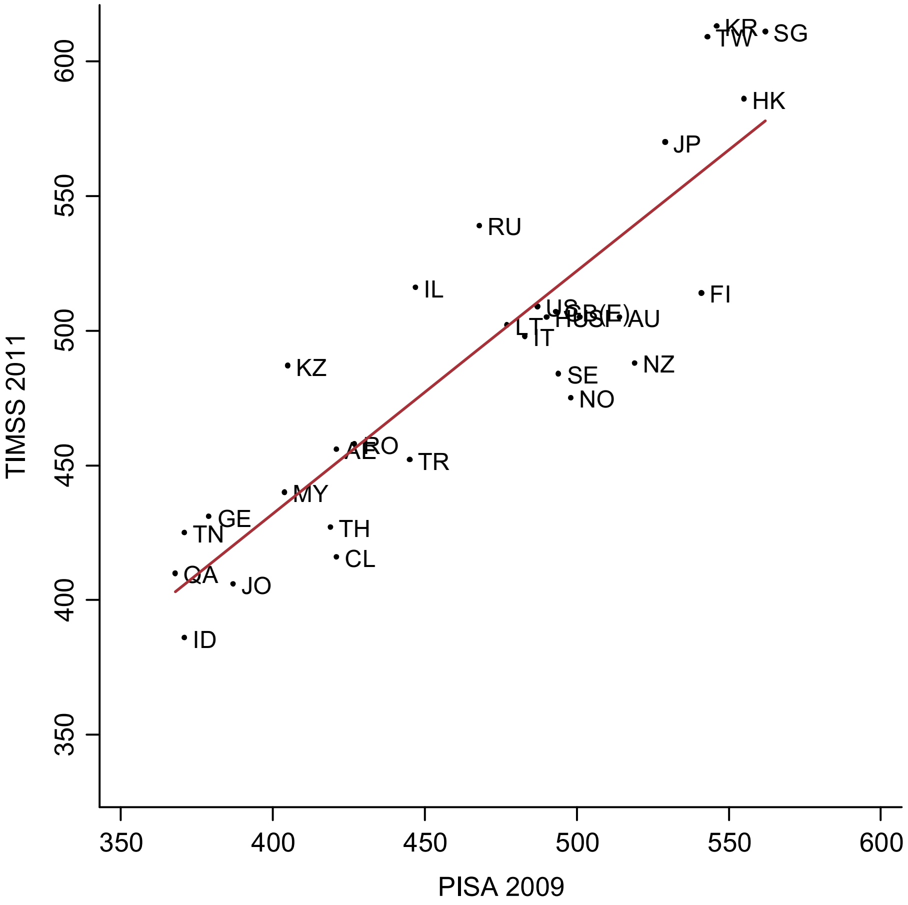
If England is set to the baseline country, then the Age∗Country interaction will indicate whether the change in the 90th percentile between the ages 10 and 14 is statistically significant relative to this country. In other words, is the change in the 90th percentile significantly greater in England than other countries?

Appendix 2. Response rates across countries and surveys

	4th grade TIMSS (2003)		8th grade TIMSS (2007)		PISA (2009)	
	School	Pupil	School	Pupil	School	Pupil
Singapore	100	98	100	96	98	91
Japan	100	97	97	97	95	95
Taiwan	100	99	100	100	100	95
Italy	100	97	100	97	99	92
Russia	100	97	100	98	100	97
Slovenia	99	92	99	95	98	91
Lithuania	96	92	99	94	100	93
Norway	93	95	93	95	97	90
Australia	90	94	100	95	99	86
Hong Kong	88	95	79	96	97	93
Scotland	83	92	86	90	89	84
England	82	93	86	93	87	87
United States	82	96	83	95	78	87

Notes: (1) School response rates refer to after replacement schools have been included.
(2) See pages 355 and 357 of http://timss.bc.edu/PDF/t03_download/T03_M_AppA.pdf for TIMSS 2003. See http://timss.bc.edu/timss2007/PDF/T07_M_IR_AppendixA.pdf pages 389 and 391 for TIMSS 2007. For PISA 2009 see http://www.oecd.org/pisa/pisaproducts/pisa2009/50036771.pdf pages 165–168.

Appendix 3. The cross-country correlation between average PISA 2009 and TIMSS 2011 (8th grade) test scores



Notes: (1) Data refers to country average (mean) test scores in the PISA 2009 and TIMSS 2011 assessments. These are presented in terms of the PISA/TIMSS test metric – meaning, figures are not directly comparable (i.e. a score of 500 on PISA does not mean the same thing as a score of 500 on TIMSS).
(2) Pearson correlation coefficient = 0.88, spearman's rank = 0.86.
(3) Official two letter country codes used (http://www.iso.org/iso/country_names_and_code_elements). GB(E) refers to England.

Appendix 4. Parent and child reports of books in the home for England (row %)

Number of books		Child report						
		0–10	11–25	26–100	101–200	>200	Total	*n*
Parent report	0–10	21	30	34	10	5	100	198
	11–25	17	28	37	12	6	100	468
	26–100	9	13	42	19	17	100	1,133
	101–200	3	7	30	29	31	100	892
	>200	2	4	20	20	54	100	888

Notes: The data are the pooled PIRLS 2001 and 2006 samples. Unweighted data.