CAMBRIDGE ASSESSMENT

The relationship between time in education and achievement in PISA in England

Tom Benton

Paper presented at British Educational Research Association conference, London, September 2014.

Research Division
Cambridge Assessment
1 Regent Street
Cambridge
CB2 1GG

Benton.T@cambridgeassessment.org.uk

www.cambridgeassessment.org.uk

Abstract

Current OECD publications suggest that each additional year of schooling has a substantial impact upon the performance of pupils in PISA tests. However, their figures are based on simple models that compare the performance of pupils in different school years but do not differentiate between the effect of additional schooling and the effect of pupil age. It is already well known that on average, within any given year group, the oldest pupils will outperform the youngest. This is most commonly seen amongst pupils in English schools by comparing the results of September and August-born children. Failing to account for the effect of age may lead to an overestimation of the relationship between additional schooling and performance.

This paper estimates the effect of one additional year of schooling on achievement in PISA for pupils in England. The size of this effect within English schools may indicate the extent to which the curriculum being taught in English schools aligns with the particular content of the PISA tests.

The research presented in this paper makes use of combined data available from the Programme for International Student Assessment (PISA) from 2000 and 2003. A regression discontinuity approach is used to examine the impact of additional schooling on achievement. In essence, this approach compares the performance of the youngest pupils in one year group (that is, August-born children) with the oldest pupils in the year group below (that is, September-born children) whilst accounting for the general trend in the relationship between age and achievement. Because August-born children are in a higher year group and have typically been attending school for a longer period of time, this difference provides an estimate of the likely impact of additional schooling on performance.

The analysis extends and improves upon the work of Luyten, Peschar and Coe (2008) and shows that, once the age of pupils is taken into account, there is in fact no relationship at all between additional schooling and performance in the PISA tests. This points towards a disconnect between what has historically been taught in English schools and what is being tested in PISA. This in turn raises questions about the extent to which performance in PISA can be used to assess the quality of teaching in England.

Introduction

The aim of this research is to explore the link between the extent of schooling received by pupils and their ability as measured by the Programme for International Student Assessment (PISA). More specifically the aim is to examine the impact of whether pupils have been taught up to Year 11 or only up to Year 10. It is likely that those in the former group will have received an additional year's schooling compared to those in the latter, although the flexible nature of the time at which pupils start school in reception means that this will not be universally the case.

Estimates of this nature are provided by OECD (OECD, 2010, page 169, table A1.2). On a scale of performance defined internationally to have a mean of 500 and a standard deviation of 100[1], their analysis indicates that each year of schooling is associated with an increase in performance roughly equivalent to 40 points (and 36 points in the UK). These figures are derived from the association between the year group recorded for each pupil and their estimated level of ability. However, looking at the data from PISA 2009, for pupils in the UK at least, these comparisons are problematic for the following reasons:

- Almost all pupils in England were recorded as being in Year 11. In fact, for the relevant 2009 data only 13 pupils were outside this year group.
- The vast majority of pupils recorded as being in Year 12 (248 out of 250) were in Scotland[2].
- The majority of pupils recorded as being in Year 10 (328 out of 421) were in Northern Ireland with the vast majority of the remainder (67 out of 93) being in Scotland[3].
- Thus, the comparison between different year groups is actually a comparison between (sub-groups of) different nations within the UK.
- Comparisons with pupils in year 12 in Scotland are additionally problematic as this is past the age of compulsory education in Scotland. Thus, their achievement may be influenced by (unmeasured) factors leading to the decision to remain in education rather than the impact of additional education itself[4] (Luyten, Peschar and Coe, 2008, page 340).
- Depending on how years of schooling are counted it is not necessarily true that those in year 12 in Scotland have had an additional year of schooling compared to those in year 11 in England. The labelling of year groups in Scotland starts at 1 whereas in England it starts with Reception and Year 1 is (for most pupils) the second year of schooling. Thus many pupils in Year 11 in England will have attended school for just as long as those in Year 12 in Scotland.

Examining the code used to undertake the analysis[5] it also clear that weights are applied at pupil level rather than school level. This means that the cross-national data will be weighted as if schools in England contained more pupils than those in Wales, Northern Ireland or Scotland rather than as if the sampled schools in England represent a much larger number of schools than those in Wales, Northern Ireland or Scotland. This is likely to lead to incorrect relative weighting of the four countries in this analysis. Indeed the OECD are aware of this issue themselves and have stated that "technical discussion is currently under way regarding the use of separate weights at the different levels" (OECD, 2009, page 229, note 2).

Finally it should be noted that none of the OECD's estimates disaggregate the effect of age from the effect of additional schooling. Thus, although it is typically shown that students in

[1] For OECD countries participating in PISA 2000.
[2] With 1 pupil in England and 1 in Wales recorded as being in this group.
[3] With 12 pupils in England and 14 in Wales recorded as being in this group.
[4] The group of year 10 pupils in Scotland will also be non-randomly assigned to this group since this is likely to be the result of an earlier decision to hold pupils back from starting education for a year.
[5] Provided by the macro "PROC_MIXED_PV.sas", available in OECD (2009, page 306).

higher school years tend to perform better than those in lower school years, it cannot be ascertained whether this is because they are older or because of the impact of additional schooling.

It should be noted that the OECD's report makes very little of the individual figures for the UK on this topic[6] so these observations are not intended to pose a major challenge to any of the OECD's conclusions. The important point is to note that the figure from the OECD report suggesting that each additional year of schooling is associated with an increase of 40 points is worthy of further scrutiny.

One previous paper examining this topic in England is that of Luyten, Peschar and Coe (2008). In order to disaggregate the impact of age from the impact of additional schooling, their analysis was based upon a regression discontinuity approach. This approach relies upon a fairly deterministic assignment of pupils to year groups dependent upon the month in which they were born. Specifically, it uses the fact that for pupils born in a particular calendar year, those born from January to August will (usually) be placed in a higher year group than those born between September and December. The analysis then examines the trend of the relationship between month of birth and performance and looks for a jump (or "discontinuity") in this trend between those born in August and those born in September. Their analysis, based on data from PISA 2000, identified a "remarkably modest" effect of 0.12 standard deviations (roughly equivalent to 12 points on the PISA scale) for each additional year of schooling. However, this paper also had a number of weaknesses:

- Although the paper claims to be examining data for England it is clear that data from Northern Ireland has been included as well. For example, page 325 of the paper states that 6,327 pupils were included in analysis. However, as is made clear Micklewright and Schnepf (2006, page 12) only 4,120 pupils in England participated in PISA 2000.
- The above point is crucial as the break between year groups for pupils in Northern Ireland occurs between pupils born in June and those born in July rather than between August and September. That is, in Northern Ireland, pupils born in July and August are typically amongst the oldest in a year group rather than the youngest.
- Thus, the analysis excluded pupils in Northern Ireland born in July and August and in Year 10 (the most common year group for these pupils). Instead the analysis included a small number of pupils born in these months who were in Year 11[7]. These are likely to be have been pupils with accelerated progress through school.
- Ability measures in PISA are provided not as a single figure but as multiple *plausible values*. These values are intended to force researchers to attend to not only the estimated ability level of pupil but also the size of measurement error. The analysis by Luyten, Peschar and Coe treated these separate plausible values as if they were multiple indicators of ability. This is inappropriate and can lead to incorrect estimates of population characteristics (see Mislevy, 1993).

Further exploration of these issues, including an attempt to recreate the figures from the analysis, also indicated that it is unlikely that the above issues were dealt with by weighting the data[8]. Thus, it is worth re-examining the link between time in schooling and achievement in England.

[6] Indeed the figure in question for the UK is not individually referred to at all.
[7] The figure of 6,327 pupils to include in analysis (as stated in the paper by Luyten et al.) can be achieved from the publicly available data for PISA 2000 by restricting data to all pupils in England and Northern Ireland born in 1984 and in Year 10 if born in September to December and in Year 11 if born in January to August.
[8] With similar issues to those being discussed by OECD likely to be at the heart of this as both approaches make use of multilevel modelling. Specifically, it is unlikely that different weights were applied at different levels.

Method
The analysis in this paper retains the spirit of the approach employed by Luyten, Peschar and Coe in that it also relies upon a regression discontinuity design. However, to avoid the issues relating to the need to apply different weights at pupil and school level, the analysis in this paper uses ordinary least-squares regression rather than multilevel modelling. The standard errors derived from the regression are adjusted to account for the fact that pupils are clustered within schools[9].

The analysis also extends upon the work of Luyten et al. (2008) in the following ways:

- It examines achievement in all of Reading, Mathematics and Science rather than just Reading.
- It includes data from PISA 2003 in addition to data from PISA 2000.

Note that it is not possible to conduct an analysis of this type for any of PISA 2006, 2009 and 2012. The reason for this is that the timing of fieldwork was moved from March to November so that after 2003, all PISA samples in England were based upon a single year group (year 11) rather than those born in a single calendar year. This means that it is not possible to apply the regression discontinuity approach to these more recent studies.

In order to apply the regression discontinuity technique, data was restricted to pupils in England born within the calendar year of interest (1984 for PISA 2000 and 1987 for PISA 2003) where those born from January to August were recorded as being in year 11 and those born from September to December were recorded as being in year 10. In total 4,012 pupils from PISA 2000 and 3,625 from PISA 2003 were included in analysis[10]. For PISA 2000, estimates of ability in Mathematics and Science were not supplied for all pupils so that in this year, for these subjects, 2,231 and 2,223 pupils respectively were available for analysis.

For each subject of interest, separate regression lines were fitted examining the relationship between month of birth and ability for those born between January and August and those born from September to December. The standard errors of the fitted regression lines were calculated at each month. Using these standard errors it was possible to construct 84 per cent intervals for the expected performance in each month. Using the logic of Cumming (2009), we can infer that there is a significant difference in performance of pupils born in August (in year 11) from those born in September (in Year 10) if there is no overlap in the confidence intervals for expected performance in these two months of birth.

For the purposes of brevity, only the analysis of the combined data from PISA 2000 and PISA 2003 is displayed. It should be noted that separate analyses of the data from each year largely leads to the same conclusions. One exception to thiswill be discussed briefly within the results section.

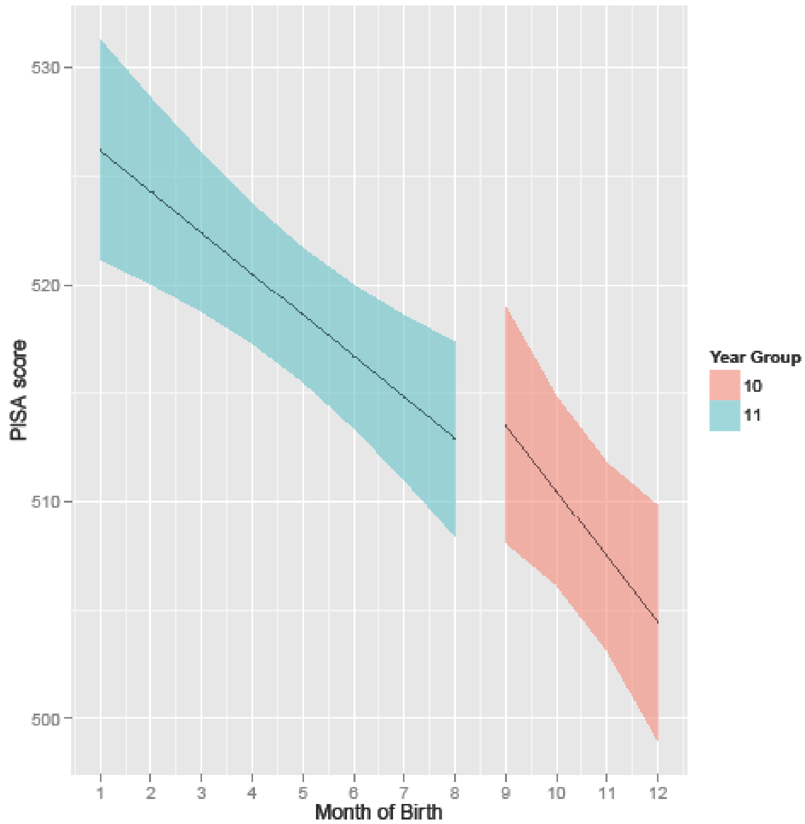
Results
The results of analysis are shown in figures 1 to 3. In both subjects, in common with other research, the results show that within each year group, the oldest pupils tend to outperform the youngest. However, more interestingly, for both Reading and Mathematics, the results actually imply that the expected performance of August-born pupils in Year 11 is actually *below* that of September born pupils in Year 10. In other words, having nearly completed Year 11[11] and having received the associated additional instruction does not appear to be

[9] This is achieved via the method of *balanced repeated replication*, which is itself facilitated by the existing replicate weights provided within the PISA data sets.
[10] With approximately one twelfth of pupils born in each calendar month within each year.
[11] The tests were conducted in March.

associated with any increase in performance in either Reading or Mathematics. As can be seen from the clear overlap of the confidence intervals for the two year groups at the point of the discontinuity, the difference between year groups is not statistically significant. Specifically, the analysis suggests that, for August-born pupils the effect of being Year 11 on Reading performance is -3.7 points with a standard error of 4.8[12]. For Mathematics, the equivalent figures are -4.2 with a standard error of 5.5

For Science, the effect is in the reverse direction, however, it remains far from statistically significant. Specifically, the effect of being in year 11 is estimated to be +4.7 for August-born pupils with a standard error of 5.1. Thus, the data provides no evidence that the additional material taught to pupils in Year 11 has any effect on their performance in PISA tests.

Figure 1: Relationship between expected Reading performance and month of birth for each year group with 84 per cent confidence intervals.



[12] Note that, any estimated effect needs to approximately twice as large as its standard error in order to be considered statistically significant at the 5 per cent level.

Figure 2: Relationship between expected Maths performance and month of birth for each year group with 84 per cent confidence intervals.

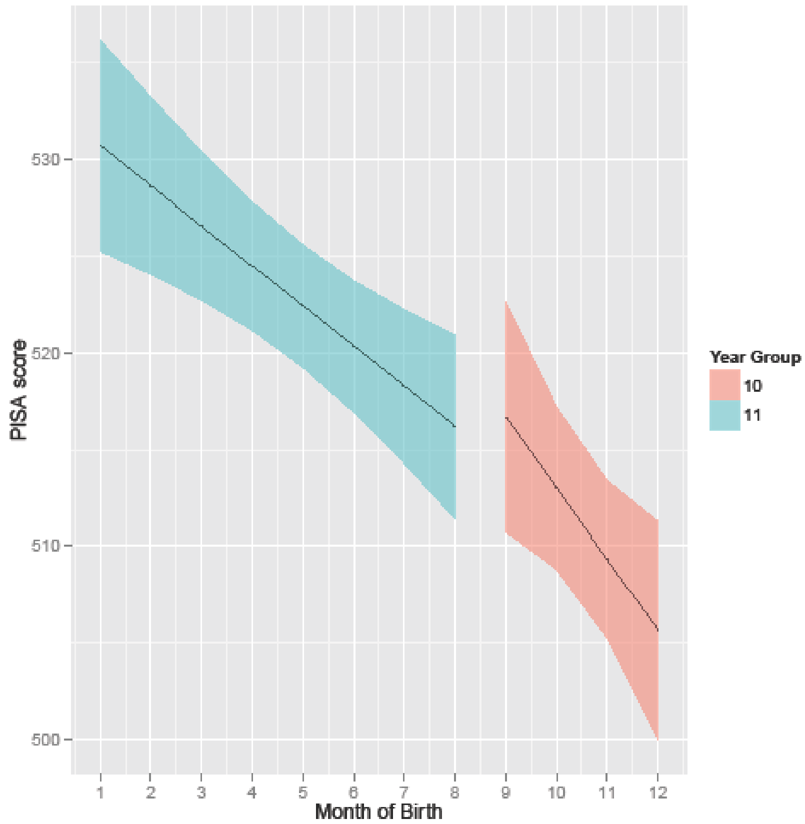
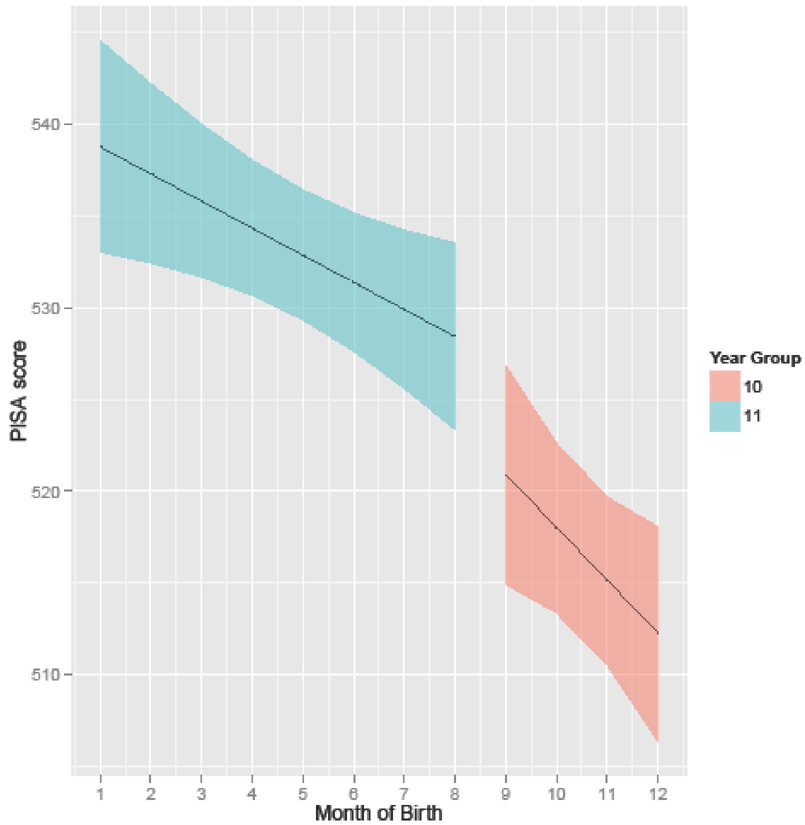


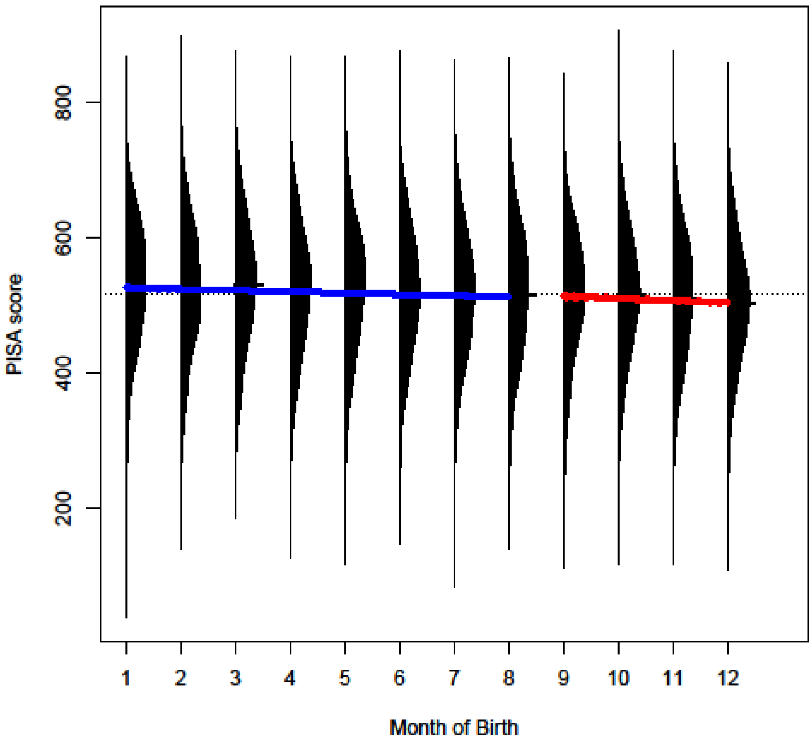
Figure 3: Relationship between expected Science performance and month of birth for each year group with 84 per cent confidence intervals.



Note that the result for Science shown in Figure 3 differs if the data from PISA 2000 is analysed on its own (that is, without the data being combined with that from PISA 2003). If this is done, the analysis estimates that the effect of being in Year 11 is +17.1 points for an August-born pupil with a standard error of 8.1; a result that is just statistically significant at the 5 per cent level. However, it should be noted that this result is by no means repeated for the data from PISA 2003 (which yields a *negative* estimated effect of -5.9 points and standard error of 7.9). Overall, therefore, there is still no consistent evidence of additional schooling being associated with increased performance in PISA.

Figures 4 to 6 show the effects above within the context of the full range of ability of pupils in England (unweighted distributions[13,14]). As can be seen, when viewed in this way, both the effect of age, and the effect of which grade pupils are part of is very small. This shows that the effect of both of the variables of interest is far less important than other (possibly unmeasured) factors in driving differences in performance between different pupils. These charts also confirm the good fit of the linear models used for analysis to the data[15] in that the fitted lines closely match the midpoints of the score distributions for each month of birth.

Figure 4: (Unweighted) Distribution of Reading performance by month of birth



[13] These charts were created using the R package *beanplot* (Kampstra, 2008). This package does not allow for analysis of weighted data. However, for the purposes of visualisation, the charts based on unweighted data are sufficient.
[14] Distributions are created using all five plausible values for each student.
[15] This was also confirmed by examining the impact of adding further quadratic terms to the regression models. These additional terms were not statistically significant and, furthermore, did not alter the overall conclusions about the impact of being in Year 11 rather than Year 10.

Figure 5: (Unweighted) Distribution of Mathematics performance by month of birth

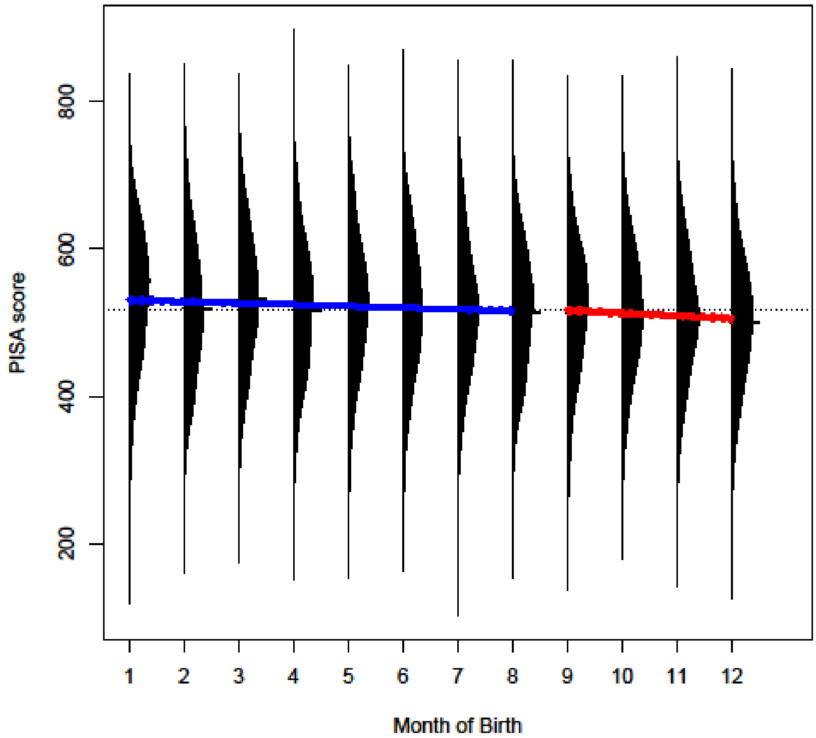
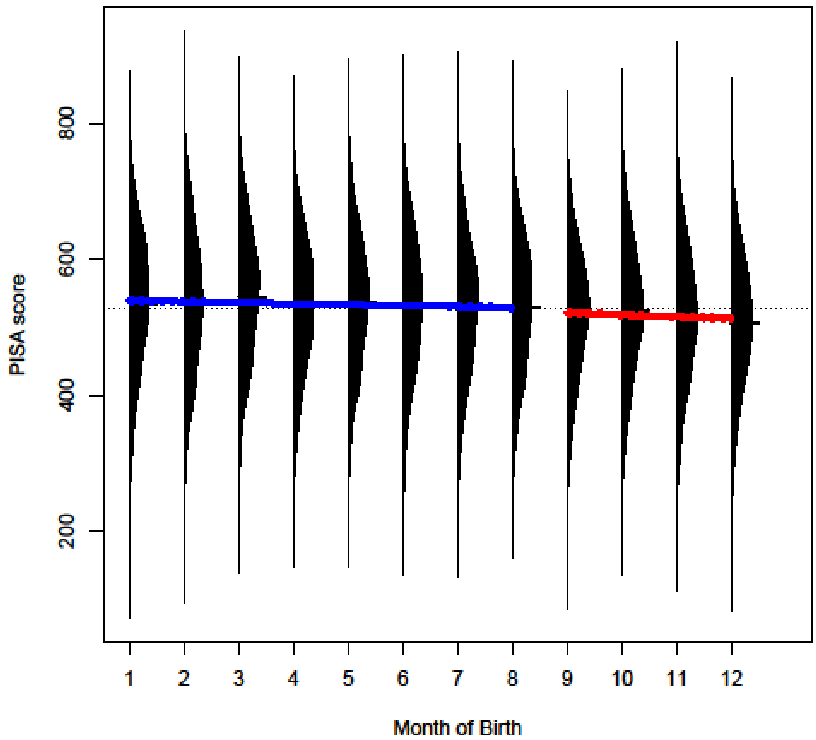


Figure 6: (Unweighted) Distribution of Science performance by month of birth



Discussion

The analysis in this paper shows no evidence of any relationship at all between additional schooling and performance in the PISA tests. More specifically, after accounting for the impact of age, it reveals that pupils who had been taught a Year 11 curriculum did not perform any better in these international tests than those who had only been taught up to Year 10. Despite the changes and extensions to the analysis performed by Luyten, Peschar and Coe, the conclusion of the analysis here differs only slightly from theirs. Their paper concluded that “a remarkably modest effect on reading performance was found” whereas we go a little further to say that there is in fact no evidence of any effect on performance at all.

The results here suggest a disconnect between what has historically been taught in English schools (at least in Year 11) and what is being tested in PISA. This in turn raises questions about the extent to which performance in PISA can be used to assess the quality of teaching in England. Specifically, it raises the question of whether any supposedly “poor performance” or “stagnation” in these international tests can be attributed to the quality of teaching in England if the curriculum being taught doesn’t seem to align well with the content of the PISA tests.

One weakness of this analysis is that it is based on data which is now quite old – from 2000 and 2003. Whilst the change in the nature of the samples collected for PISA (that is, sampling from pupils in a single year group rather than those pupils born in a given calendar year) is probably helpful in encouraging participation in international tests, it is unfortunate that it is not possible to explore the issues examined in this paper from more recent data. However, even without this further evidence, the analysis here reiterates the need for users of international rankings to be mindful of what has been measured and the extent to which this may (or may not) link to what is taught in schools.

References

Cumming, G. (2009) Inference by eye: Reading the overlap of independent confidence intervals. *Statistics in Medicine*, 28, 205-220.

Kampstra, P. (2008) Beanplot: A Boxplot Alternative for Visual Comparison of Distributions. *Journal of Statistical Software: Code Snippets*, 28(1), 1-9. Available online: http://www.jstatsoft.org/v28/c01.

Luyten, H., Peschar, J., and Coe, R. (2008) Effects of schooling on reading performance, reading engagement, and reading activities of 15-Year-Olds in England. *American Educational Research Journal*, 45(2), 319-342.

Micklewright, J., and Schnepf, S.V. (2006) *Response Bias in England in PISA 2000 and 2003*. DfES Research Report RR771. DfES: London.

Mislevy, R.J. (1993) Should “multiple imputations” be treated as “multiple indicators”? *Psychometrika*, 58(1), 79-85.

OECD (2009) *PISA Data Analysis Manual: SAS® Second Edition*. OECD: Paris.

OECD (2010) *PISA 2009 Results: What students know and can do – student performance in reading, mathematics and science (Volume I). OECD: Paris.* Available online: http://dx.doi.org/10.1787/9789264091450-en.