CAMBRIDGE ASSESSMENT

Can we monitor standards over time in one subject using data from another? Why a high correlation isn't enough

Tom Benton

Cambridge Assessment Research Report
11th May 2015

University of Cambridge
Local Examinations Syndicate

Author contact details:
Tom Benton
ARD Research Division
Cambridge Assessment
1 Regent Street
Cambridge
CB2 1GG
Benton.t@cambridgeassessment.org.uk

http://www.cambridgeassessment.org.uk/

Cambridge Assessment is the brand name of the University of Cambridge Local Examinations Syndicate, a department of the University of Cambridge. Cambridge Assessment is a not-for-profit organisation.

How to cite this publication:

Benton, T. (2015). *Can we monitor standards over time in one subject using data from another? Why a high correlation isn't enough*. Cambridge Assessment Research Report. Cambridge, UK: Cambridge Assessment.

Table of contents

Introduction

Maintaining examination standards over time has been a major topic of debate for a long time. One recently proposed solution to this issue is the introduction of the National Reference Test (NRT). Although there will be many important technical issues to deal with, the principle is easy to understand – use the same Mathematics and English assessments with representative samples of pupils to provide useful information about how performance in Mathematics and English evolves over time and thus to inform decisions about how GCSE pass rates should change. However, some of those related to the examinations industry would like to go beyond this. For example, Ofqual Chief Executive Glenys Stacey was quoted in the TES as saying that the NRT would need to look only at [changing performance in] the two core subjects because they are "good predictors of general achievement"[1]. That is, once we know about how performance is changing in Mathematics and English we can simply infer that performance in other subjects must be moving in the same direction. Thus, if the NRT suggests that the percentage gaining grade C or above in both Mathematics and English should rise by 2 percentage points, this should also be taken as evidence that pass rates in other subjects should rise by a similar amount. A similar method to this is already applied in Hong Kong where secure monitoring tests are used to maintain standards in core subjects and then, in turn, results in these subjects are used to set and maintain standards in optional subjects (Hong Kong Examinations and Assessment Authority [HKEAA], 2011) [2].

The logic behind this idea is fairly simple. We certainly know from existing data that achievement in the core subjects is a strong predictor of achievement in other subjects. For example, Benton and Sutch (2013) show that the correlations of Mathematics and English GCSE grades with the average grades achieved in all other GCSE subjects are 0.83 and 0.85 respectively. Other (unpublished) analysis of the same data showed that, in 2011, across 52 GCSE subjects the correlations between the average grade achieved in English and Mathematics and the grade achieved in the individual subjects ranged between 0.49 and 0.85 with a median (across subjects) of 0.73. In other words it is certainly true that, within an individual year, achievement in English and Mathematics is a reasonable predictor of how well candidates will achieve in other GCSE subjects. Thus, whenever a candidate is better at English and Mathematics we also expect them to do better in other subjects. For this same reason, so the logic would go, if we can demonstrate that a whole cohort of pupils is better at English and Mathematics, surely we can also infer that they would be better at other subjects.

Putting this more succinctly, the logic chain goes something like this:

1. Constructs A and B are strongly related. That is, pupils who perform well in one construct also tend to perform well in the other.
2. We know between Year X and Year Y, pupils have improved their performance in construct A.
3. Therefore, we can assume that they have also improved their performance in construct B.

There are of course some simple theoretical reasons why the logic above may not follow. For example, suppose the government requested that all schools spend an additional hour each week teaching Mathematics. Then we would expect (and certainly hope) to see an improvement in the average Mathematics performance of secondary school pupils. However, there would be no reason to expect any concomitant improvement in performance standards in (say) French.

---

[1] Published in the TES magazine on 20 February, 2015. https://www.tes.co.uk/article.aspx?storyCode=11006451. More recently, having awarded the contract for the development of the National Reference Test to NFER, Glenys Stacey gave a more cautious comment saying "We will act cautiously as we build our understanding of the information that the tests will provide and how this is used in GCSE awarding." (https://news.tes.co.uk/b/news/2015/03/30/schools-quot-nervous-quot-about-new-standards-test-heads-warn.aspx).

[2] This is subtly different from the suggestion of cohort referencing by Stringer (2012). Although Stringer's suggestion involves using data in some (compulsory) subjects to inform grade boundaries in others, the starting point is that the grade distribution in core subjects will be "locked" over time so that there is no attempt to measure changes in performance standards.

Indeed, it may even be that the additional focus on Mathematics could lead to a *decrease* in the level of attainment in other subjects as these are given less attention to make room in the timetable.

The aim of this report is to set aside the type of hypothetical scenario described above and instead focus purely on the extent to which empirical data supports the logic chain we have described. Specifically, we will look at data measuring whether performance has risen or fallen in different (yet highly correlated) subjects and explore the extent to which an improvement over time in one subject tends to go hand in hand with improvements in another.

Data

The analysis in this report draws on data from the Programme for International Student Assessment (PISA) from 2009 and 2012. Specifically, we will examine the extent to which improvements in Mathematics, Reading and Science between 2009 and 2012 necessarily go alongside each other. For example, this includes exploring whether improvements in Reading relate to improvements in Mathematics. The correlations between students' abilities in the subjects used for analysis are given in Table 1 below (which is itself reproduced from Table 12.6 in Organisation for Economic Co-operation and Development [OECD], 2014b).

Table 1: Latent correlations between scales (N=485,490)

	Mathematics	Reading	Science
Mathematics	1.00		
Reading	0.86	1.00	
Science	0.90	0.88	1.00

As can be seen, the various scales are highly correlated. The lowest correlation (Mathematics with Reading) is still 0.86 and the maximum correlation of Mathematics with Science is somewhat higher at 0.90. It is worth noting that all of the correlations in this table are at least as high as those seen between English and Mathematics achievement (combined) and grades achieved in other GCSE subjects. Indeed, the majority of GCSE subjects display correlations with English and Mathematics achievement that are somewhat lower than this[3].

The correlations above are at the student level when calculated across all participating countries[4]. Another mechanism to examine the relationship between the different subjects is to examine how the pass rates in one subject relate to pass rates in another subject at the level of whole countries. That is, to what extent does a country's pass rate in one subject tell us their likely pass rate in another? This was examined by looking at the percentage of students achieving PISA Level 3 or above and PISA Level 5 or above in each country. These Levels were chosen as Level 3 appears closest in standard to a GCSE grade C and Level 5 appears closest to a GCSE grade A. For example, comparing reported results from PISA 2012 to GCSE results in June 2014 as published by the Joint Council for Qualifications (JCQ)[5]:

- 55 per cent of students in the UK achieved Level 3 or above in PISA Mathematics compared to 62 per cent who achieved GCSE Mathematics grade C or above.
- 12 per cent of students in the UK achieved Level 5 or above in PISA Mathematics compared to 15 per cent who achieved GCSE Mathematics grade A or above.
- 60 per cent of students in the UK achieved Level 3 or above in PISA Reading compared to 62 per cent who achieved GCSE English grade C or above.
- 9 per cent of students in the UK achieved Level 5 or above in PISA Reading compared to 14 per cent who achieved GCSE English grade A or above.

[3] This may be partially explained by the fact that the correlations in Table 1 are estimated after removing the impact of measurement error, whereas the estimates for GCSE do not remove this influence. However, given the length of GCSE tests and the fact that all GCSEs comprise of multiple assessments this is likely to only account for a small proportion of the difference.
[4] However, correlations calculated at student level based purely upon data within the UK are very similar to those in Table 1.
[5] See http://www.jcq.org.uk/examination-results/gcses.

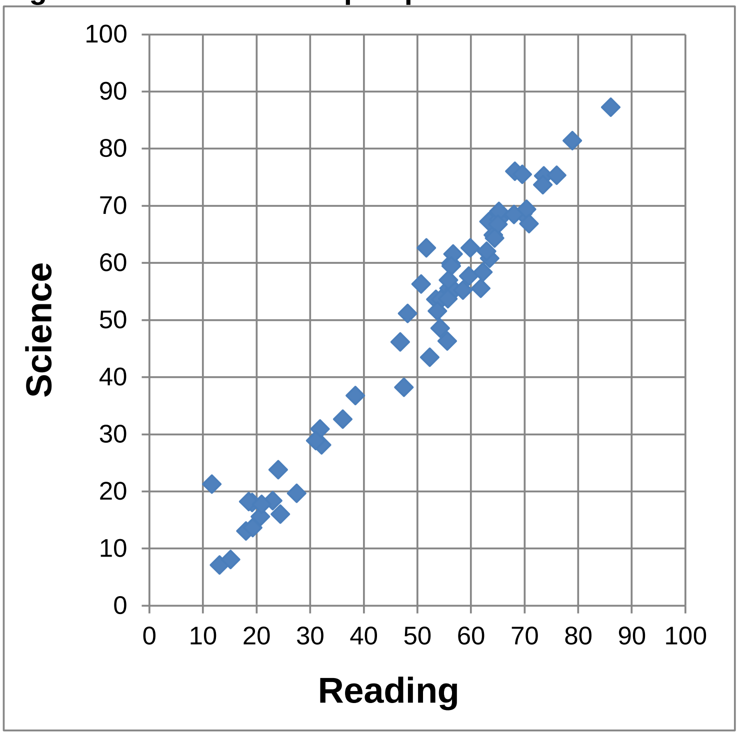
The correlations in pass rates at Level 3 and Level 5 across 60 countries with available data for both PISA 2009 and PISA 2012[6] are shown in Table 2. As can be seen, these correlations are very high. This is particularly true at Level 3 where the lowest correlation is still as high as 0.97 (Reading with Mathematics). The correlations in pass rates are slightly lower at Level 5 (probably due to the decrease in the range of pass rates) but nonetheless remain high; ranging from 0.85 to 0.95.

Table 2: Correlations in pass rates at various Levels in 2009 and 2012 (N=60)

	PISA 2009			PISA 2012		
	Mathematics	Reading	Science	Mathematics	Reading	Science
Pass rate at Level 3						
Mathematics	1.00			1.00		
Reading	0.97	1.00		0.97	1.00	
Science	0.98	0.98	1.00	0.98	0.98	1.00
Pass rate at Level 5						
Mathematics	1.00			1.00		
Reading	0.85	1.00		0.89	1.00	
Science	0.91	0.94	1.00	0.88	0.95	1.00

An example of one of the strongest of these relationships is given in Figure 1. This shows the relationship between the pass rates at Level 3 for Science and Reading in 2012. As can be seen, a country's pass rate in one of the subjects is very informative about their likely pass rate in the other. However, having said this, there are already hints that the error surrounding any such inferences may be quite large. For example, amongst countries with a pass rate in Reading of roughly 50 per cent, the pass rate in Science ranges all the way from 38 per cent to 63 per cent. However, in part this variation may be due to measurement error and the fact that both scales are measured with only small numbers of items. This fact will be considered further in the next section.

Figure 1: The relationship in pass rates at Level 3 between Reading and Science in 2012



6 Note that three countries (Malaysia, Costa Rica and United Arab Emirates) that took the PISA 2009 assessments in 2010 are not included in this analysis.

Analysis of relationships in changes

Having established that the various subjects are strongly correlated with one another *within* an individual year, we can continue with the main purpose of analysis – examining the strength of the relationship in changes *between* years. Table 3 examines this in terms of the correlations between changes in the pass rates in each subscale. At both Level 3 and Level 5, the correlations remain high but noticeably lower than in Table 2. For example, whereas the correlation in absolute pass rates at Level 3 between Reading and Science was around 0.98 in each year, the correlation in the change in these pass rates over time is much smaller at only 0.80. Most strikingly, the change in the Level 3 pass rate for Reading displays a correlation of only 0.66 with the change in the Level 3 pass rate for Mathematics. This is despite a correlation of over 0.85 between the two measures at individual Level and of 0.97 between countries' pass rates in the two subjects within a given year. Another way to think about this is to note that the square of the correlation indicates the proportion of variance in one variable that can be attributed to another. Applying this thinking here would indicate that only just over 40 per cent of the variance in changes in the Reading pass rate can be explained by changes in the Science pass rate[7].

Table 3: Correlations in changes in pass rates between 2009 and 2012 (N=60)

	Correlations in changes in pass rate between 2009 and 2012		
	Mathematics	Reading	Science
Change in Pass rate at Level 3			
Mathematics	1.00		
Reading	0.66	1.00	
Science	0.83	0.80	1.00
Pass rate at Level 5			
Mathematics	1.00		
Reading	0.78	1.00	
Science	0.68	0.66	1.00

Figures 2 and 3 give further details on the relationship between changes in pass rates for Mathematics and Science at Level 3 and Level 5 respectively. Figure 2 shows that at Level 3 there is a fairly strong relationship between the changes in pass rates – as we would expect given the correlation of 0.83 in Table 3. However, this chart also shows that the change in one pass rate is not a very accurate predictor of the likely change in the other. For example:

- In Chile, the Level 3 pass rate for Mathematics rose by 1.4 percentage points (from 21.7 to 23.1 per cent) whereas the pass rate for Science fell by 1.6 percentage points (from 32.6 to 30.9 per cent[8]).
- In Finland, the Level 3 pass rate for Mathematics fell by 9.4 percentage points (from 76.6 to 67.2 per cent) whereas the pass rate for Science fell far less dramatically by 3.2 percentage points (from 78.7 to 75.5 per cent).
- In the Spain, the Level 3 pass rate for Mathematics fell by 0.8 percentage points (from 52.3 to 51.5 per cent) whereas the pass rate for Science rose considerably by 3.2 percentage points (from 53.8 to 57.0 per cent).

Even aside from these obvious outliers, across the 60 countries, the median level of the (absolute) difference between the changes in the two pass rates is 1.6 percentage points. Thus a fairly typical case would be France where the Level 3 pass rate in Mathematics fell by 2.1 percentage points (from 57.6 to 55.5 per cent) but the Science pass rate fell by only 0.3

[7] That is, 0.66 squared is 0.44 – only just above 0.40.
[8] Minus an additional 0.1 due to rounding.

percentage points (from 58.7 to 58.4 per cent). Thus, even in the average case, the change in one pass rate was not a good indicator of the change in the other.

Figure 2: The relationship in changes in pass rates at Level 3 in Mathematics and Science between 2009 and 2012

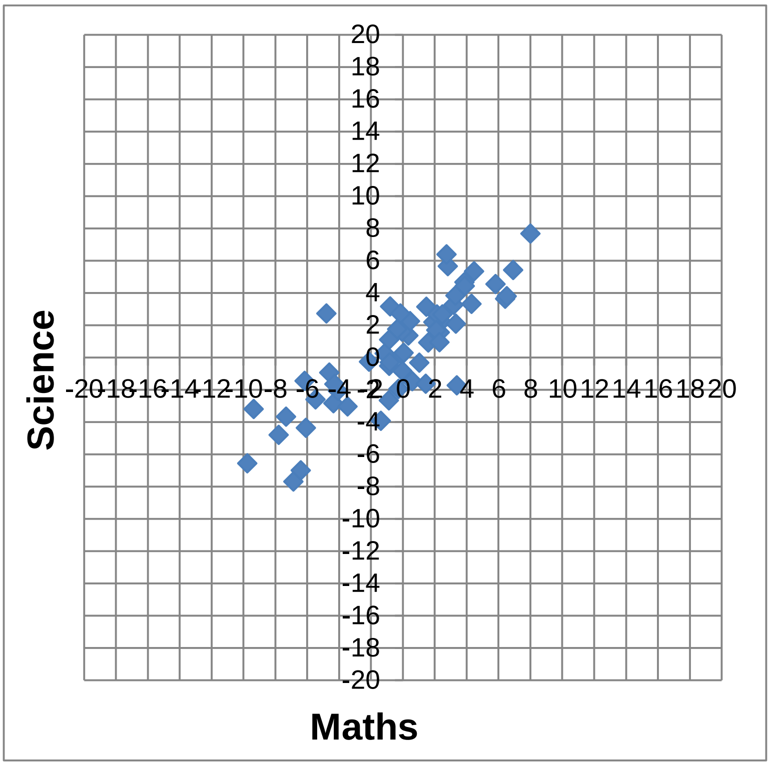


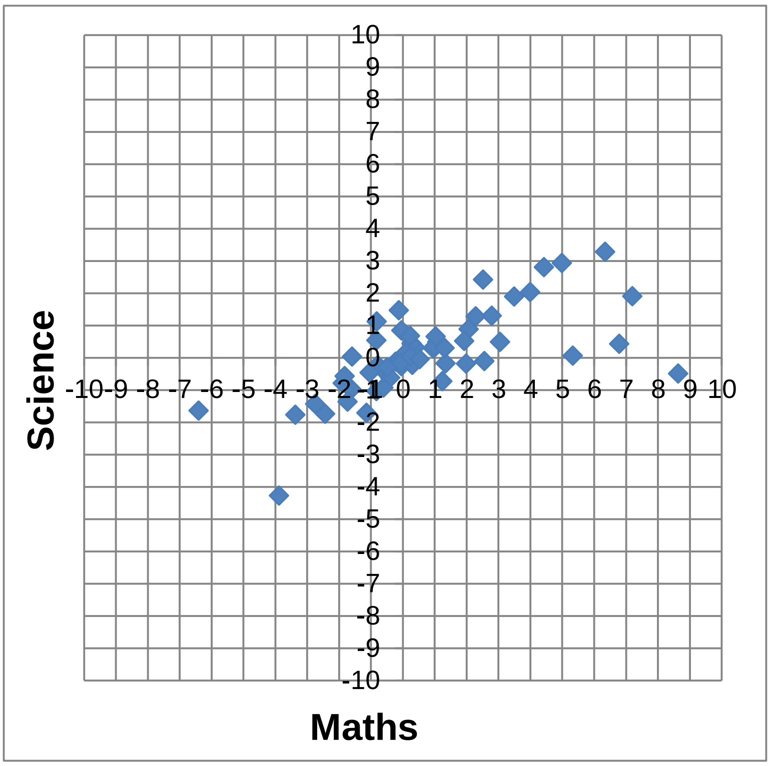
Figure 3 shows that the situation at Level 5 is no better. For example (and deliberately focussing on different countries to those noted above):

- In Korea, the Level 5 pass rate for Mathematics rose by 5.3 percentage points (from 25.6 to 30.9 per cent) whereas the pass rate for Science remained steady at around 11.6 per cent.
- In the United Kingdom, the Level 5 pass rate for Mathematics rose by 2.0 percentage points (from 9.8 to 11.8 per cent) whereas the pass rate for Science fell slightly by 0.2 percentage points (from 11.4 to 11.2 per cent)[9].

As can be seen by the tight clustering of points at the centre of Figure 3, and also by the change of the scale, for many of the countries examined, the changes in pass rates at Level 5 are quite small. Thus, across the 60 countries, the median level of the (absolute) difference between the changes in the two pass rates is only 0.7 percentage points. Nonetheless, from a practical perspective, differences even only slightly larger than this may be crucial when it comes to setting standards in GCSEs. Indeed, at present, Ofqual's approach to standard setting recommends that, for many subjects, grade boundaries are set so that the GCSE pass rate is within one percentage point of a prediction based on Key Stage 2 (KS2) results. In this context, even the small differences seen for many countries in Figure 3 indicate an unhelpfully large level of uncertainty.

[9] Note that both of these results can be easily verified from Tables 1.2.1b and 1.5.1b in Annex B1 of OECD (2014a).

Figure 3: The relationship in changes in pass rates at Level 5 in Mathematics and Science between 2009 and 2012



A possible major criticism of the analyses in Figures 2 and 3 is that they do not take account of the effects of measurement and sampling error. For example, it is possible that differences in the changes in pass rates are simply due to the unreliability of the measurement process and that if these same measurements had been undertaken using longer tests assessing each subscale, and with larger samples, then the changes in pass rates would be far more closely aligned. Fortunately, the data provided by PISA is explicitly designed to allow us to take account of the impact of measurement and sampling error. Specifically, rather than assigning a single score to each candidate, PISA provides 5 *plausible values* for each subject which indicate the possible range of each candidate's ability given the measurement error that is implicit in the PISA assessments. Furthermore, resampling weights are provided to enable analysts to ascertain the likely impact of sampling error. The OECD has also provided SAS macros (see OECD, 2009) that make it easy to perform analysis to ascertain the standard errors of any estimates whilst taking account of both measurement and sampling error. Using these macros it was possible to calculate standard errors for the difference between the changes in the pass rate for each subject using the following procedure:

1. Note that the difference in the change in the pass rate between two subjects is mathematically equivalent to the change in the difference in the pass rates.
2. Evaluate the standard error for the difference as the pass rate in year j (which we will label SED_j) using the SAS macro `proc_means_pv.sas` applied to a variable that is the difference between an indicator of whether a student had achieved a given Level or above in one subject and whether they had achieved the same given Level or above in another.
3. Estimate the standard error in the change in the difference between years j and k as $\sqrt{SED_j^2 + SED_k^2}$.

Using these standard errors we can then calculate approximate 95 per cent confidence intervals for the differences in the changes in pass rates within each country. These confidence intervals are shown in Figures 4 and 5[10]. For example, Figure 4 shows that, as discussed earlier, for Chile the difference in the change in pass rates at Level 3 is 3.1 percentage points. Furthermore, even after taking account of measurement and sampling error, the 95 per cent confidence interval for the difference in the change indicates that for Chile as a whole, the percentage of pupils at Level 3 and above in Mathematics rose by at least 0.7 percentage points more than for Science. More strikingly, at the other end of the chart, the analysis indicates that the percentage of pupils in Finland at Level 3 or above in Mathematics fell by at least 4 percentage points more than in Science.

In fact, across Figure 4 as a whole, almost a third of countries (19 out of 60) display confidence intervals that do not overlap with zero. In other words, for a notable number of countries the change in their pass rate in one subject is significantly different from their change in the pass rate in another. This provides strong evidence that changes in attainment in one subject do not provide an accurate indicator of likely changes in another – even if, as in this case, the two subjects are very highly correlated at the individual level.

Figure 4: Differences in changes in Level 3 pass rates between Mathematics and Science between 2009 and 2012 together with confidence intervals accounting for measurement and sampling error

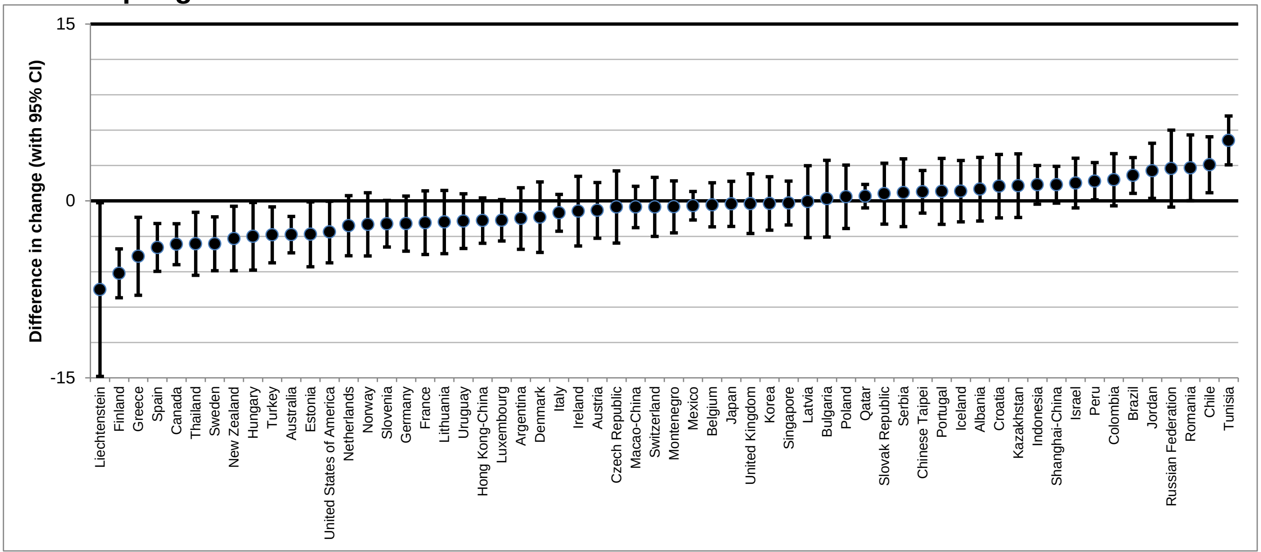
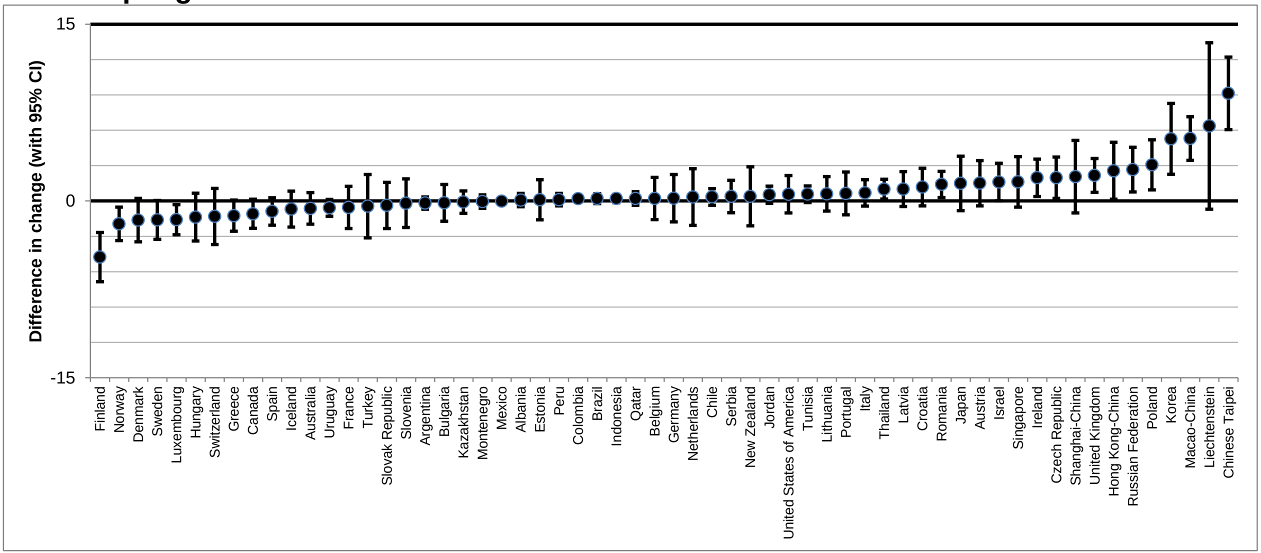


Figure 5, provides similar information at Level 5. In this chart a quarter of countries (15 out of 60) display a significant difference in the changes in pass rate between the two subjects. For example, in Korea, the analysis indicates that the proportion of pupils at Level 5 or above in Mathematics rose by at least 2.2 percentage points more than in Science.

[10] Note that the suggested method does not take account of linking error – that is, the uncertainty in the equating method between years. However, any such errors would have the effect of either moving all countries up in the chart (that is, adjusting the change in Mathematics upwards for all countries relative to the change in Science) or moving all countries down in the chart. Since the countries are roughly evenly split between those improving more in Mathematics and those improving more in Science this would have little effect on our overall conclusions. For this reason, for our purposes, the link error can be safely ignored.

Figure 5: Differences in changes in Level 5 pass rates between Mathematics and Science between 2009 and 2012 together with confidence intervals accounting for measurement and sampling error



Similar analyses to those illustrated in Figures 4 and 5 were undertaken in each of the three possible pairs of subjects. The results of these analyses are summarised in Table 4 in terms of the numbers of countries displaying a significant difference in the change in pass rates between two subscales. As can be seen, at both Levels, the results that we have illustrated for Mathematics and Science are the best possible case of the three pairs that could have been considered. Indeed, at Level 3, the table shows that the *majority* of countries show differential improvement between Mathematics and Reading between 2009 and 2012. This further illustrates that, even when attainment in two subjects is highly correlated, using changes in attainment over time in one subject to infer the likely changes over time in another is extremely dangerous.

Table 4: Number of countries where change in pass rate in one subject between 2009 and 2012 is significantly different from the change in another subject

First subject in pair	**Second subject in pair**	**Number of countries with significant difference in change in pass rates**		**Percentage of countries with significant difference in change in pass rates (out of 60)**	
		Level 3	**Level 5**	**Level 3**	**Level 5**
Mathematics	Reading	34	17	57%	28%
Mathematics	Science	19	15	32%	25%
Reading	Science	26	20	43%	33%

Discussion

Maintaining examination standards over time is doubtless a very difficult task. As such, the prospect of a well-designed reference test that will accurately monitor changes in national performance over time is an attractive one. Indeed, the temptation to hope that this may provide valuable evidence beyond the two subjects for which it has been designed (Mathematics and English) is strong. However, this report shows very clear reasons why this should not be done. It is likely that the design of the NRT (in terms of sampling design and methods used for linking) will be very similar to that employed for international surveys such as PISA. If we accept that such procedures are robust, the results in this report clearly show that improvements in national performance in one particular subject do not necessarily imply improvements in others. Furthermore, we have shown that, for a substantial number of countries, the differences in the changes in pass rates go well beyond the level we would expect due to sampling and measurement error and are statistically significantly different.

It should be noted that the detailed analysis provided so far has only considered the differences in the changes in pass rates over a relatively short period of time - three years. It would be expected that over a longer period of time the differences would become far more pronounced. This is briefly examined in Appendix 1, which recreates the analysis but looks at changes in Mathematics and Reading[11] performance between PISA 2003 and PISA 2012. This analysis shows a remarkably weak relationship between changes in performance in the two subjects. For example, the correlation in the change in the pass rate at Level 5 is only 0.41 between the two subjects. At both Level 3 and Level 5, more than two-thirds of the 39 countries with available data show a significantly different change in pass rates between the two subjects. A striking case is provided by France which saw its pass rate at Level 3 in Reading rise from 59.7 to 62.2 per cent whilst in Mathematics the pass rate fell from 63.1 to 55.5 per cent. This gives a clear example of how, even within a large country, national performance in different subjects may drift apart over time.

Taken together, the results imply that using results from an NRT in English and Mathematics to infer how pass rates in other subjects should change would be unwise. After all it should be remembered that the correlations between Mathematics and Reading performance as measured in PISA are generally somewhat higher than the correlations between Mathematics and English grades and achievement in other GCSE subjects[12]. For this reason, even if the NRT proves to be successful for monitoring standards within English and Mathematics themselves, deciding how to maintain standards in other subjects will remain an on-going question. This means that identifying practical methods of maintaining standards, including how to make best use of expert judgement will remain an important research topic.

Further details of countries' pass rates at Level 3 and Level 5 in PISA Reading and Mathematics are shown in Appendix 2. This provides the interested reader with further precise examples of where changes of the pass rate in one subject do not necessarily go alongside a similar change in the other. It also provides some idea of the likely volatility over time in the percentage of students attaining given performance standards nationally when this is estimated via a sample survey such as PISA or (potentially) the NRT.

[11] Science performance in PISA 2012 is not comparable to Science performance in any PISA surveys prior to 2006.

[12] The greater correlations for PISA might possibly be due to greater alignment in the style of assessment questions, or possibly due to skills in multiple subjects all being tested concurrently within the same test booklets.

References

Benton, T. and Sutch, T. (2013). *Exploring the value of GCSE prediction matrices based upon attainment at Key Stage 2*. Cambridge Assessment Research Report. Cambridge, UK: Cambridge Assessment.

Hong Kong Examinations and Assessment Authority (2011). *Grading procedures and standards-referenced reporting in the HKDSE Examination*. Hong Kong: HKEAA.

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

OECD (2014a). *PISA 2012 Results: What Students Know and Can Do – Student Performance in Mathematics, Reading and Science (Volume I)*. Paris: OECD. http://dx.doi.org/10.1787/9789264201118-en.

OECD (2014b). *PISA 2012 Technical Report*. Paris: OECD.

Stringer, N. S. (2012). Setting and maintaining GCSE and GCE grading standards: the case for contextualised cohort-referencing. *Research Papers in Education*, 27(5), 535-554.

Appendix 1: Changes in PISA Mathematics and Reading performances between 2003 and 2012

Figure 6: The relationship in changes in pass rates at Level 3 in Mathematics and Reading between 2003 and 2012 (correlation=0.71)

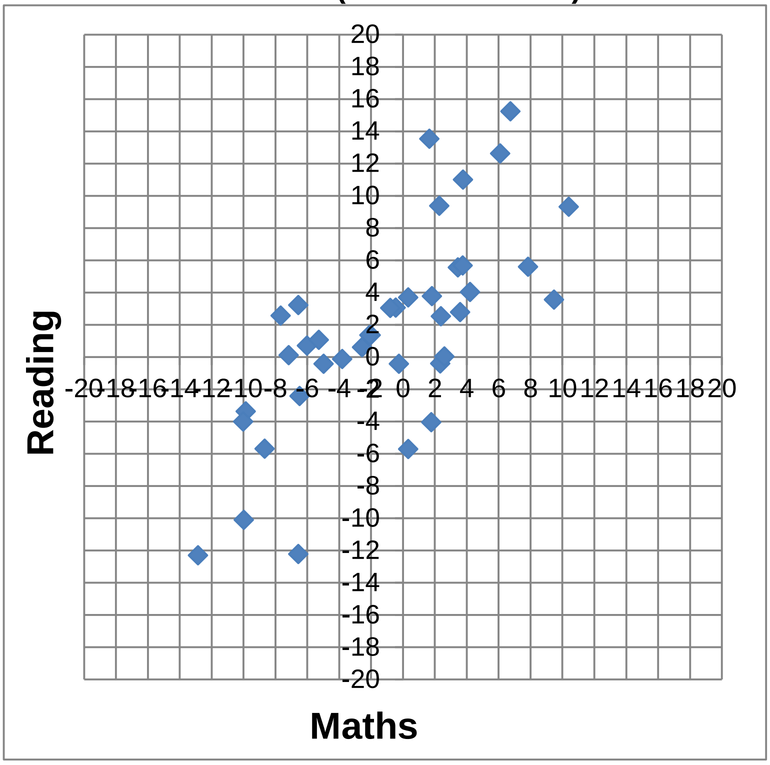


Figure 7: The relationship in changes in pass rates at Level 5 in Mathematics and Reading between 2003 and 2012 (correlation=0.44)

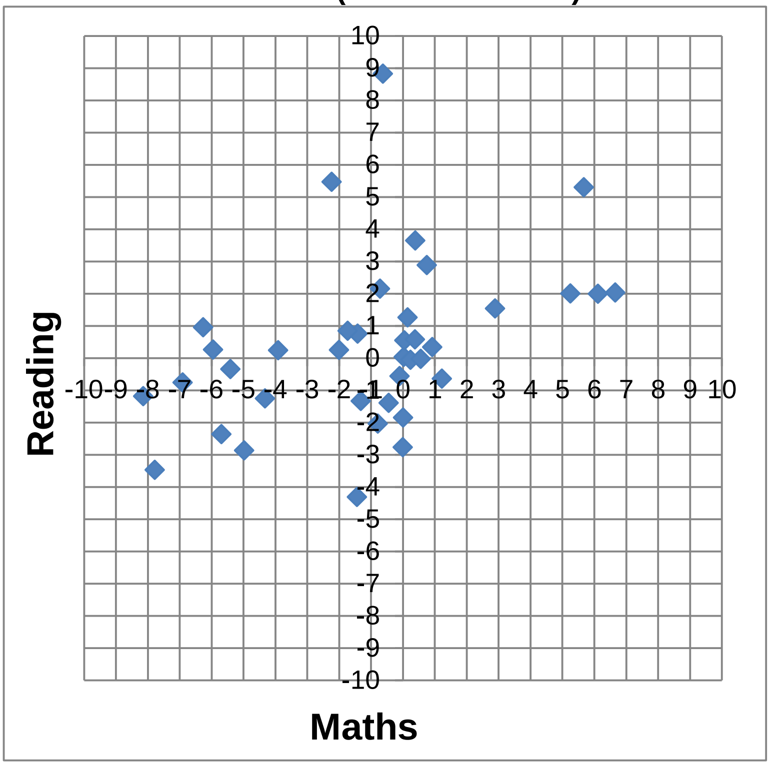


Figure 8: Differences in changes in Level 3 pass rates between Mathematics and Reading between 2003 and 2012 together with confidence intervals accounting for measurement and sampling error

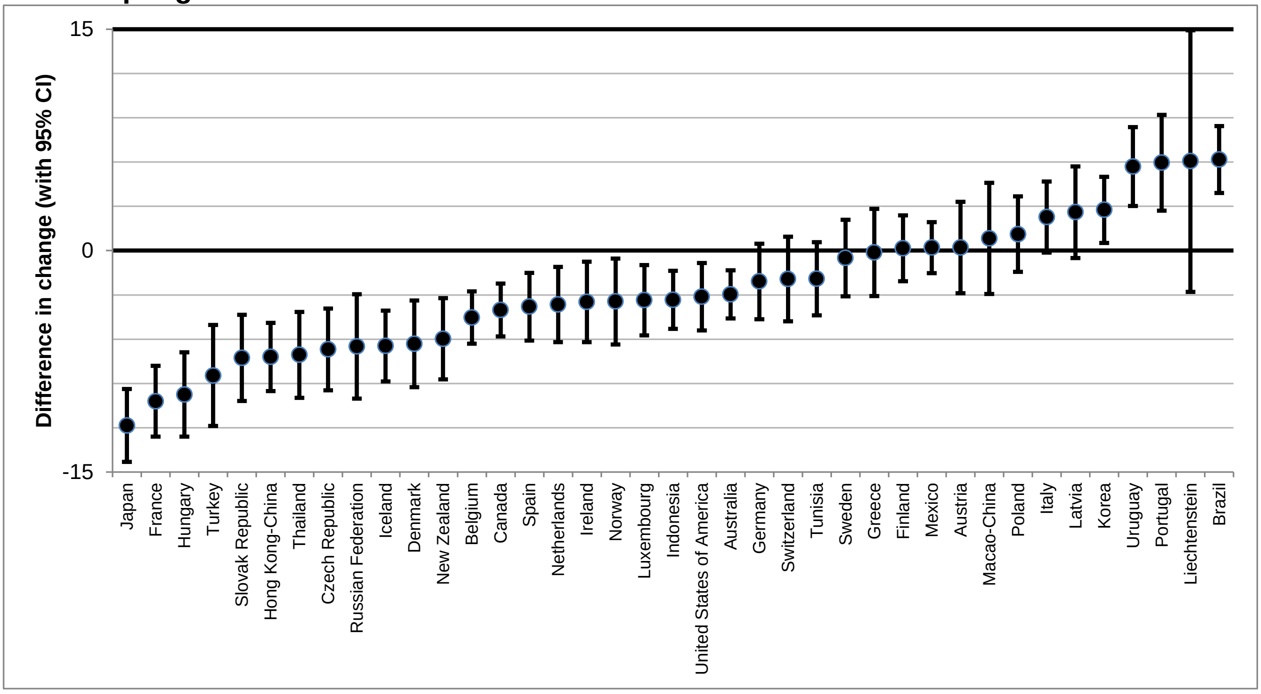


Figure 9: Differences in changes in Level 5 pass rates between Mathematics and Reading between 2003 and 2012 together with confidence intervals accounting for measurement and sampling error

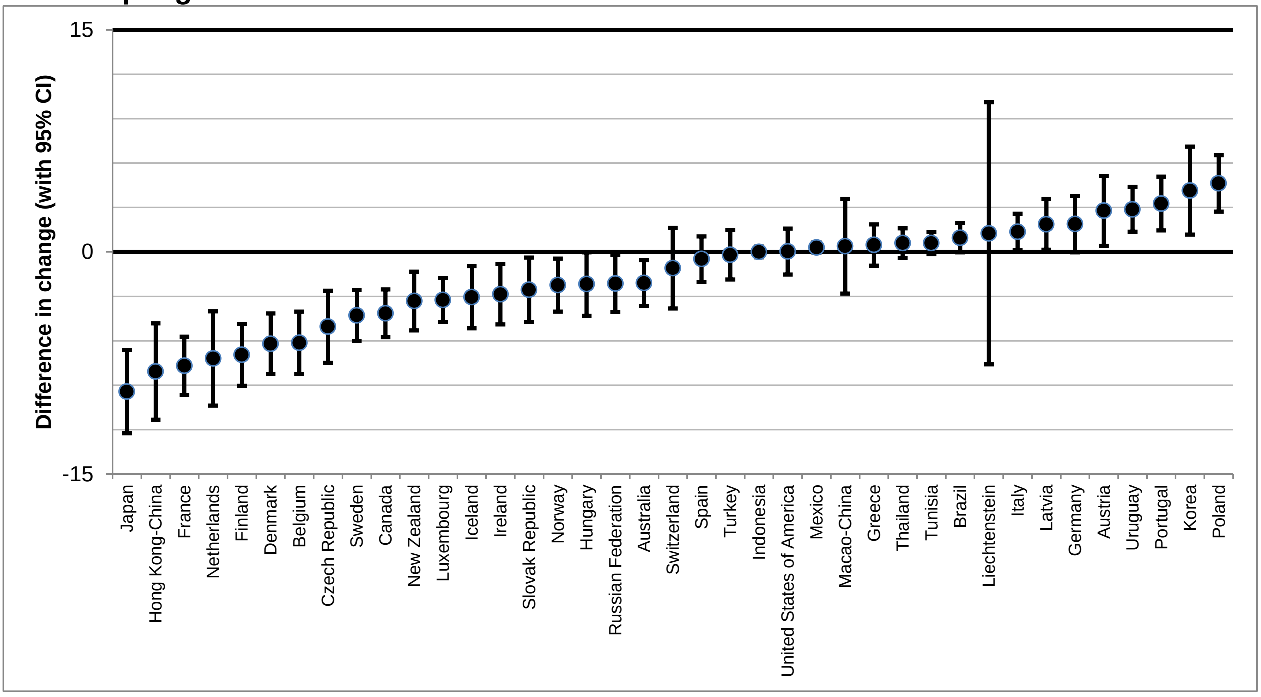


Table 8: Number of countries where change in pass rate in one subject between 2003 and 2012 is significantly different from the change in another subject

First subject in pair	Second subject in pair	Number of countries with significant difference in change in pass rates		Percentage of countries with significant difference in change in pass rates (out of 39)	
		Level 3	Level 5	Level 3	Level 5
Mathematics	Reading	26	26	67%	67%

Appendix 2: Detailed pass rates at Level 3 and Level 5 in Mathematics and Reading for countries in PISA 2003, 2009 and 2012

Table 9: Pass rates at Level 3 for each country in PISA 2003, 2009 and 2012 together with standard errors

	Mathematics pass rates (standard error)			Reading pass rates (standard error)			Difference in change in pass rates (standard error)	
Country	2003	2009	2012	2003	2009	2012	2003-2012	2009-2012
Albania	-	12.1 (1.2)	16.4 (1.0)	-	17.7 (1.4)	23.0 (0.9)	-	-1.0 (1.6)
Argentina	-	15.7 (1.5)	11.3 (1.1)	-	23.0 (1.7)	19.2 (1.3)	-	-0.5 (1.3)
Australia	67.1 (0.9)	63.9 (1.0)	58.4 (0.8)	69.9 (1.0)	65.3 (0.9)	64.2 (0.7)	-3.0 (0.8)	-4.3 (0.9)
Austria	59.6 (1.5)	55.6 (1.4)	59.4 (1.2)	56.7 (1.6)	48.3 (1.3)	56.3 (1.3)	0.2 (1.6)	-4.1 (1.5)
Belgium	67.6 (0.9)	63.5 (1.0)	62.6 (1.0)	63.9 (1.0)	62.0 (1.0)	63.5 (1.0)	-4.5 (0.9)	-2.4 (1.1)
Brazil	10.7 (1.4)	11.9 (0.9)	12.0 (0.8)	24.8 (1.5)	23.3 (1.1)	19.9 (0.8)	6.2 (1.2)	3.5 (0.9)
Bulgaria	-	29.5 (2.2)	31.9 (1.7)	-	35.6 (2.5)	38.5 (2.0)	-	-0.5 (1.9)
Canada	71.6 (0.8)	69.8 (0.8)	65.1 (0.9)	72.2 (0.8)	69.5 (0.7)	69.7 (0.8)	-4.0 (0.9)	-4.8 (0.9)
Chile	-	21.7 (1.4)	23.1 (1.3)	-	36.3 (1.5)	31.9 (1.4)	-	5.8 (1.3)
Chinese Taipei	-	71.7 (1.0)	74.0 (1.1)	-	59.7 (1.2)	70.4 (1.4)	-	-8.3 (1.1)
Colombia	-	9.2 (0.8)	8.4 (0.8)	-	22.3 (1.3)	18.0 (1.3)	-	3.4 (1.4)
Croatia	-	40.1 (1.5)	43.4 (1.6)	-	50.2 (1.5)	53.5 (1.7)	-	0.1 (1.7)
Czech Republic	63.3 (1.5)	53.4 (1.3)	57.3 (1.3)	56.0 (1.4)	49.5 (1.4)	56.7 (1.4)	-6.7 (1.4)	-3.3 (1.5)
Denmark	64.0 (1.3)	59.9 (1.4)	58.8 (1.2)	58.5 (1.5)	58.8 (1.1)	59.6 (1.2)	-6.3 (1.5)	-2.0 (1.5)
Estonia	-	64.6 (1.4)	67.5 (1.0)	-	61.0 (1.5)	68.2 (1.1)	-	-4.3 (1.5)
Finland	77.2 (0.8)	76.6 (1.0)	67.2 (0.9)	79.7 (0.8)	75.2 (0.9)	69.6 (1.0)	0.2 (1.1)	-3.7 (1.2)
France	63.1 (1.2)	57.6 (1.4)	55.5 (1.1)	59.7 (1.2)	59.2 (1.4)	62.2 (1.0)	-10.2 (1.2)	-5.2 (1.3)
Germany	59.4 (1.7)	62.6 (1.4)	62.9 (1.3)	57.8 (1.4)	59.3 (1.3)	63.4 (1.3)	-2.1 (1.3)	-3.8 (1.6)
Greece	34.8 (1.8)	43.3 (1.8)	37.2 (1.2)	49.8 (1.8)	53.1 (1.9)	52.3 (1.5)	-0.1 (1.5)	-5.3 (1.8)
Hong Kong-China	75.7 (1.9)	78.0 (1.1)	79.5 (1.3)	67.9 (1.9)	75.7 (1.0)	78.9 (1.2)	-7.2 (1.2)	-1.8 (1.1)
Hungary	53.2 (1.4)	54.5 (1.7)	46.7 (1.5)	52.7 (1.2)	58.7 (1.5)	56.0 (1.5)	-9.7 (1.5)	-5.1 (1.5)
Iceland	64.8 (0.9)	61.8 (0.9)	54.9 (0.9)	57.6 (0.9)	61.0 (0.8)	54.3 (0.9)	-6.5 (1.2)	-0.1 (1.1)
Indonesia	7.1 (1.0)	6.4 (1.1)	7.5 (1.4)	9.4 (1.1)	12.3 (1.6)	13.1 (1.6)	-3.3 (1.0)	0.2 (1.3)
Ireland	59.6 (1.3)	54.7 (1.5)	59.2 (1.2)	67.8 (1.4)	59.5 (1.3)	70.8 (1.3)	-3.5 (1.4)	-6.9 (1.8)
Israel	-	38.0 (1.2)	44.9 (1.9)	-	51.0 (1.3)	55.6 (1.8)	-	2.3 (1.4)
Italy	43.4 (1.3)	50.9 (0.9)	51.3 (1.0)	51.3 (1.3)	54.9 (0.8)	56.8 (0.9)	2.3 (1.2)	-1.6 (0.8)
Japan	70.3 (1.5)	70.1 (1.4)	72.0 (1.4)	60.1 (1.7)	68.4 (1.4)	73.6 (1.4)	-11.8 (1.3)	-3.2 (1.2)
Jordan	-	11.9 (1.3)	10.5 (0.9)	-	20.1 (1.3)	18.5 (1.2)	-	0.2 (1.3)
Kazakhstan	-	17.4 (1.2)	23.2 (1.7)	-	17.2 (1.2)	11.7 (1.0)	-	11.3 (1.6)
Korea	73.9 (1.3)	76.3 (1.7)	76.2 (1.4)	76.4 (1.4)	78.8 (1.4)	76.0 (1.4)	2.8 (1.1)	2.8 (1.2)
Latvia	50.8 (1.8)	50.2 (1.7)	53.4 (1.5)	56.3 (1.9)	53.6 (1.7)	56.3 (1.3)	2.6 (1.6)	0.5 (1.7)
Liechtenstein	70.4 (2.6)	75.5 (2.4)	70.7 (2.2)	71.0 (2.6)	60.3 (2.2)	65.2 (3.3)	6.1 (4.5)	-9.7 (4.4)
Lithuania	-	47.6 (1.3)	48.0 (1.3)	-	45.6 (1.2)	50.7 (1.3)	-	-4.7 (1.4)
Luxembourg	55.3 (0.7)	53.4 (0.6)	53.3 (0.6)	53.0 (0.9)	50.0 (0.6)	54.4 (0.7)	-3.4 (1.2)	-4.5 (1.0)
Macao-China	69.2 (1.6)	69.5 (0.7)	72.8 (0.6)	62.5 (1.6)	54.6 (0.7)	65.2 (0.6)	0.8 (1.9)	-7.4 (0.9)
Mexico	13.2 (1.2)	20.9 (0.8)	17.5 (0.6)	20.4 (1.4)	26.9 (0.8)	24.5 (0.7)	0.2 (0.9)	-1.0 (0.7)
Montenegro	-	17.0 (0.7)	19.1 (0.6)	-	22.4 (0.7)	27.5 (0.7)	-	-2.9 (1.0)
Netherlands	71.1 (1.8)	67.6 (2.3)	67.3 (1.6)	65.2 (1.7)	60.9 (2.7)	65.0 (1.6)	-3.7 (1.3)	-4.5 (1.5)
New Zealand	65.8 (1.0)	65.5 (1.1)	55.8 (1.0)	67.0 (1.1)	66.3 (1.0)	62.9 (1.0)	-6.0 (1.4)	-6.4 (1.3)
Norway	55.5 (1.3)	57.5 (1.3)	53.4 (1.2)	60.5 (1.3)	61.4 (1.2)	61.8 (1.4)	-3.4 (1.5)	-4.5 (1.4)
Peru	-	9.6 (1.2)	9.3 (1.1)	-	13.2 (1.3)	15.2 (1.4)	-	-2.3 (1.2)
Poland	53.1 (1.3)	55.5 (1.3)	63.5 (1.4)	58.7 (1.3)	60.5 (1.3)	68.1 (1.3)	1.1 (1.3)	0.4 (1.3)
Portugal	42.8 (1.6)	52.4 (1.4)	52.3 (1.8)	52.2 (1.7)	56.0 (1.6)	55.7 (1.9)	6.0 (1.7)	0.3 (1.6)
Qatar	-	13.2 (0.4)	15.3 (0.3)	-	18.2 (0.4)	21.0 (0.5)	-	-0.7 (0.7)
Romania	-	24.4 (1.5)	30.9 (1.8)	-	28.0 (1.8)	32.2 (1.8)	-	2.2 (1.7)
Russian Federation	43.3 (2.1)	42.9 (1.6)	49.5 (1.5)	35.6 (1.7)	41.1 (1.4)	48.2 (1.5)	-6.5 (1.8)	-0.6 (1.8)
Serbia	-	33.0 (1.4)	34.6 (1.6)	-	33.9 (1.2)	36.1 (1.6)	-	-0.6 (1.5)
Shanghai-China	-	86.4 (0.9)	88.7 (1.0)	-	82.7 (1.1)	86.1 (1.1)	-	-1.1 (0.9)
Singapore	-	77.1 (0.8)	79.5 (0.6)	-	69.0 (0.5)	73.4 (0.7)	-	-2.0 (1.0)
Slovak Republic	56.6 (1.6)	55.8 (1.4)	49.4 (1.6)	46.7 (1.5)	49.7 (1.3)	46.8 (1.7)	-7.3 (1.5)	-3.5 (2.0)

Country	Mathematics pass rates (standard error)			Reading pass rates (standard error)			Difference in change in pass rates (standard error)	
	2003	2009	2012	2003	2009	2012	2003-2012	2009-2012
Slovenia	-	57.2 (0.8)	56.3 (0.8)	-	53.2 (0.6)	51.7 (0.7)	-	0.6 (1.2)
Spain	52.3 (1.3)	52.3 (1.0)	51.5 (1.0)	52.8 (1.1)	53.6 (1.1)	55.8 (0.9)	-3.8 (1.2)	-3.0 (1.1)
Sweden	61.1 (1.1)	55.6 (1.3)	48.2 (1.1)	66.1 (1.1)	59.1 (1.3)	53.7 (1.2)	-0.5 (1.3)	-2.0 (1.3)
Switzerland	67.9 (1.2)	70.6 (1.2)	69.8 (1.3)	60.7 (1.6)	60.5 (1.0)	64.4 (1.2)	-1.9 (1.5)	-4.8 (1.5)
Thailand	20.6 (1.3)	20.2 (1.4)	22.9 (1.6)	21.7 (1.3)	20.3 (1.1)	31.0 (1.6)	-7.1 (1.5)	-8.0 (1.5)
Tunisia	7.3 (0.9)	7.7 (1.0)	11.1 (1.4)	13.6 (1.0)	18.3 (1.3)	19.3 (1.6)	-1.9 (1.3)	2.4 (1.4)
Turkey	25.7 (2.5)	32.6 (2.0)	32.5 (2.2)	32.3 (2.5)	43.3 (1.8)	47.5 (2.1)	-8.5 (1.7)	-4.4 (1.7)
United Kingdom	-	54.9 (1.3)	55.0 (1.5)	-	56.7 (1.1)	59.9 (1.6)	-	-3.2 (1.2)
Uruguay	27.7 (1.1)	27.3 (1.1)	21.2 (1.1)	36.3 (1.2)	30.1 (1.0)	24.0 (1.0)	5.7 (1.4)	0.0 (1.3)
USA	50.4 (1.3)	52.2 (1.6)	47.9 (1.7)	57.9 (1.4)	58.0 (1.5)	58.5 (1.7)	-3.1 (1.2)	-4.8 (1.2)

Table 10: Pass rates at Level 5 for each country in PISA 2003, 2009 and 2012 together with standard errors

Country	Mathematics pass rates (standard error)			Reading pass rates (standard error)			Difference in change in pass rates (standard error)	
	2003	2009	2012	2003	2009	2012	2003-2012	2009-2012
Albania	-	0.4 (0.2)	0.8 (0.2)	-	0.2 (0.1)	1.2 (0.2)	-	-0.6 (0.4)
Argentina	-	0.9 (0.3)	0.3 (0.1)	-	1.0 (0.2)	0.5 (0.1)	-	-0.1 (0.3)
Australia	19.8 (0.8)	16.4 (0.9)	14.8 (0.6)	14.6 (0.7)	12.8 (0.8)	11.7 (0.5)	-2.1 (0.8)	-0.6 (0.8)
Austria	14.3 (1.0)	12.9 (0.9)	14.3 (0.9)	8.3 (0.8)	4.9 (0.5)	5.5 (0.6)	2.8 (1.2)	0.7 (1.2)
Belgium	26.4 (0.8)	20.4 (0.7)	19.5 (0.8)	12.5 (0.5)	11.2 (0.6)	11.8 (0.6)	-6.1 (1.1)	-1.4 (1.0)
Brazil	1.2 (0.4)	0.8 (0.2)	0.7 (0.2)	1.9 (0.5)	1.3 (0.2)	0.5 (0.1)	1.0 (0.5)	0.7 (0.2)
Bulgaria	-	3.8 (1.0)	4.1 (0.6)	-	2.8 (0.5)	4.3 (0.6)	-	-1.3 (0.8)
Canada	20.3 (0.7)	18.3 (0.6)	16.4 (0.6)	12.6 (0.5)	12.8 (0.5)	12.9 (0.6)	-4.1 (0.8)	-2.0 (0.7)
Chile	-	1.3 (0.3)	1.6 (0.2)	-	1.3 (0.3)	0.6 (0.1)	-	0.9 (0.4)
Chinese Taipei	-	28.6 (1.5)	37.2 (1.2)	-	5.2 (0.8)	11.8 (0.8)	-	2.0 (1.7)
Colombia	-	0.1 (0.1)	0.3 (0.1)	-	0.6 (0.2)	0.3 (0.1)	-	0.4 (0.2)
Croatia	-	4.9 (0.7)	7.0 (1.1)	-	3.2 (0.4)	4.4 (0.7)	-	0.9 (0.9)
Czech Republic	18.3 (1.2)	11.6 (0.9)	12.9 (0.8)	6.4 (0.6)	5.1 (0.5)	6.1 (0.5)	-5.0 (1.2)	0.3 (1.2)
Denmark	15.9 (0.9)	11.6 (0.8)	10.0 (0.7)	5.2 (0.5)	4.7 (0.5)	5.4 (0.6)	-6.2 (1.0)	-2.3 (1.0)
Estonia	-	12.1 (0.8)	14.6 (0.8)	-	6.1 (0.6)	8.3 (0.7)	-	0.3 (0.9)
Finland	23.4 (0.8)	21.7 (0.9)	15.3 (0.7)	14.7 (0.7)	14.5 (0.8)	13.5 (0.6)	-6.9 (1.1)	-5.4 (1.1)
France	15.1 (0.9)	13.7 (1.0)	12.9 (0.8)	7.4 (0.6)	9.6 (1.0)	12.9 (0.8)	-7.7 (1.0)	-4.1 (1.0)
Germany	16.2 (0.9)	17.8 (0.9)	17.5 (0.9)	9.6 (0.6)	7.6 (0.6)	8.9 (0.7)	1.9 (1.0)	-1.7 (1.1)
Greece	4.0 (0.6)	5.7 (0.6)	3.9 (0.4)	5.7 (0.7)	5.6 (0.5)	5.1 (0.6)	0.5 (0.7)	-1.3 (0.8)
Hong Kong-China	30.7 (1.5)	30.7 (1.2)	33.7 (1.4)	5.7 (0.5)	12.4 (0.8)	16.8 (1.2)	-8.1 (1.7)	-1.3 (1.6)
Hungary	10.7 (0.9)	10.1 (1.1)	9.3 (1.1)	4.9 (0.6)	6.1 (0.7)	5.6 (0.8)	-2.2 (1.1)	-0.4 (1.0)
Iceland	15.5 (0.7)	13.6 (0.6)	11.2 (0.7)	7.1 (0.6)	8.5 (0.6)	5.8 (0.5)	-3.1 (1.1)	0.3 (0.9)
Indonesia	0.2 (0.1)	0.1 (0.0)	0.3 (0.2)	0.1 (0.1)	0.0 (0.0)	0.1 (0.1)	0.0 (0.2)	0.1 (0.2)
Ireland	11.4 (0.8)	6.7 (0.6)	10.7 (0.5)	9.3 (0.7)	7.0 (0.5)	11.4 (0.7)	-2.9 (1.0)	-0.4 (0.8)
Israel	-	5.9 (0.7)	9.4 (1.0)	-	7.4 (0.6)	9.6 (0.8)	-	1.3 (0.8)
Italy	7.0 (0.5)	9.0 (0.5)	9.9 (0.6)	5.2 (0.3)	5.8 (0.3)	6.7 (0.3)	1.4 (0.6)	0.1 (0.6)
Japan	24.3 (1.5)	20.9 (1.2)	23.7 (1.5)	9.7 (0.9)	13.4 (0.9)	18.5 (1.3)	-9.4 (1.4)	-2.3 (1.3)
Jordan	-	0.3 (0.2)	0.6 (0.4)	-	0.2 (0.1)	0.1 (0.1)	-	0.4 (0.4)
Kazakhstan	-	1.2 (0.4)	0.9 (0.3)	-	0.4 (0.1)	0.0 (0.0)	-	0.1 (0.5)
Korea	24.8 (1.4)	25.6 (1.6)	30.9 (1.8)	12.2 (1.1)	12.9 (1.1)	14.1 (1.2)	4.1 (1.5)	4.1 (1.5)
Latvia	8.0 (0.8)	5.7 (0.6)	8.0 (0.8)	6.0 (0.7)	2.9 (0.4)	4.2 (0.6)	1.9 (0.9)	1.1 (0.8)
Liechtenstein	25.6 (3.4)	18.1 (2.4)	24.8 (2.6)	13.0 (2.5)	4.6 (1.4)	10.9 (2.9)	1.3 (4.5)	0.5 (4.5)
Lithuania	-	7.0 (0.7)	8.1 (0.6)	-	2.9 (0.4)	3.3 (0.4)	-	0.7 (0.8)
Luxembourg	10.8 (0.6)	11.4 (0.6)	11.2 (0.4)	5.2 (0.4)	5.7 (0.5)	8.9 (0.4)	-3.2 (0.8)	-3.3 (0.8)
Macao-China	18.7 (1.4)	17.1 (0.5)	24.3 (0.6)	1.7 (0.5)	2.9 (0.2)	7.0 (0.4)	0.4 (1.6)	3.1 (0.8)
Mexico	0.4 (0.1)	0.7 (0.1)	0.6 (0.1)	0.5 (0.1)	0.4 (0.1)	0.4 (0.1)	0.3 (0.2)	-0.1 (0.1)
Montenegro	-	1.0 (0.2)	1.0 (0.2)	-	0.6 (0.2)	1.0 (0.2)	-	-0.3 (0.3)
Netherlands	25.5 (1.3)	19.9 (1.5)	19.3 (1.2)	8.8 (0.7)	9.8 (1.1)	9.8 (0.8)	-7.2 (1.6)	-0.5 (1.5)

Country	Mathematics pass rates (standard error)			Reading pass rates (standard error)			Difference in change in pass rates (standard error)	
	2003	2009	2012	2003	2009	2012	2003-2012	2009-2012
New Zealand	20.7 (0.7)	18.9 (0.9)	15.0 (0.9)	16.3 (0.8)	15.7 (0.8)	14.0 (0.8)	-3.3 (1.0)	-2.1 (1.1)
Norway	11.4 (0.6)	10.2 (0.7)	9.4 (0.7)	10.0 (0.7)	8.4 (0.9)	10.2 (0.7)	-2.2 (0.9)	-2.6 (1.0)
Peru	-	0.6 (0.2)	0.6 (0.2)	-	0.5 (0.2)	0.5 (0.2)	-	0.0 (0.2)
Poland	10.1 (0.6)	10.4 (0.9)	16.7 (1.3)	8.0 (0.6)	7.2 (0.6)	10.0 (0.9)	4.6 (1.0)	3.5 (1.0)
Portugal	5.4 (0.5)	9.6 (0.8)	10.6 (0.8)	3.8 (0.5)	4.8 (0.5)	5.8 (0.6)	3.3 (0.9)	0.0 (0.9)
Qatar	-	1.8 (0.2)	2.0 (0.2)	-	1.7 (0.2)	1.6 (0.1)	-	0.3 (0.3)
Romania	-	1.3 (0.3)	3.2 (0.6)	-	0.7 (0.2)	1.6 (0.4)	-	1.0 (0.6)
Russian Federation	7.0 (0.8)	5.2 (0.8)	7.8 (0.8)	1.7 (0.3)	3.2 (0.5)	4.6 (0.6)	-2.1 (1.0)	1.1 (0.9)
Serbia	-	3.5 (0.5)	4.6 (0.7)	-	0.8 (0.2)	2.2 (0.4)	-	-0.4 (0.7)
Shanghai-China	-	50.4 (1.2)	55.4 (1.4)	-	19.5 (1.1)	25.1 (1.2)	-	-0.6 (1.5)
Singapore	-	35.6 (0.8)	40.0 (0.7)	-	15.7 (0.5)	21.2 (0.6)	-	-1.1 (1.0)
Slovak Republic	12.7 (0.9)	12.7 (1.0)	11.0 (0.9)	3.5 (0.4)	4.5 (0.5)	4.4 (0.7)	-2.6 (1.1)	-1.6 (1.3)
Slovenia	-	14.2 (0.6)	13.7 (0.6)	-	4.6 (0.5)	5.0 (0.4)	-	-0.9 (0.9)
Spain	7.9 (0.7)	8.1 (0.5)	8.0 (0.4)	5.0 (0.5)	3.3 (0.3)	5.5 (0.3)	-0.5 (0.8)	-2.2 (0.6)
Sweden	15.8 (0.8)	11.4 (0.8)	8.0 (0.5)	11.4 (0.7)	9.0 (0.7)	7.9 (0.6)	-4.3 (0.9)	-2.3 (0.9)
Switzerland	21.2 (1.5)	24.1 (1.4)	21.4 (1.2)	7.9 (0.8)	8.1 (0.7)	9.1 (0.7)	-1.1 (1.4)	-3.8 (1.4)
Thailand	1.6 (0.4)	1.3 (0.4)	2.6 (0.5)	0.5 (0.1)	0.3 (0.2)	0.8 (0.2)	0.6 (0.5)	0.8 (0.5)
Tunisia	0.2 (0.1)	0.3 (0.2)	0.8 (0.4)	0.3 (0.1)	0.2 (0.1)	0.2 (0.2)	0.6 (0.4)	0.5 (0.4)
Turkey	5.5 (1.6)	5.6 (1.2)	5.9 (1.1)	3.8 (1.2)	1.9 (0.4)	4.3 (0.9)	-0.2 (0.9)	-2.2 (1.1)
United Kingdom	-	9.8 (0.7)	11.8 (0.8)	-	8.0 (0.5)	8.8 (0.7)	-	1.2 (0.9)
Uruguay	2.8 (0.4)	2.4 (0.4)	1.4 (0.3)	5.3 (0.7)	1.8 (0.3)	0.9 (0.3)	2.9 (0.8)	-0.2 (0.5)
USA	10.1 (0.7)	9.9 (1.0)	8.8 (0.8)	9.3 (0.7)	9.9 (0.9)	7.9 (0.7)	0.0 (0.8)	0.8 (0.8)