Comparisons of Performance in Ireland PISA 2000 to PISA 2009: A Preliminary Report to the Department of Education and Skills

Jude Cosgrove
Gerry Shiel
Peter Archer
Rachel Perkins

Educational Research Centre,
St Patrick's College, Dublin 9

December 7, 2010

Contents

1. Overview: The Performance of Students in Ireland on PISA 2009

The first cycle of the OECD Programme for International Student Assessment (PISA) took place in 2000. Since then, PISA has been implemented in 2003, 2006 and 2009. Representative samples of 15-year olds in Ireland have participated in all four cycles to date.

In each PISA cycle, one domain is designated as a 'major' assessment domain and two or more as 'minor' assessment domains (Table 1). In 2009, reading literacy was designed as a major domain. In reporting on performance over time, the OECD compares each domain to when it was last a major domain.[1] Hence, the comparisons for 2009 are: Reading (2000 vs. 2009), Mathematics (2003 vs. 2009) and Science (2006 vs. 2009).

Table 1. PISA Assessment Domains (2000-2009)

Year	Major Domain	Minor Domains
2000	Reading Literacy	Mathematical Literacy, Scientific Literacy
2003	Mathematical Literacy	Scientific Literacy, Reading Literacy, Cross-curricular Problem Solving*
2006	Scientific Literacy	Mathematical Literacy, Reading Literacy
2009	Reading Literacy	Mathematical Literacy, Scientific Literacy

*Additional minor domain, assessed only in 2003.

Ireland dropped 9 scale points on PISA reading between 2000 and 2006 (one-tenth of a standard deviation, which was not statistically significant), and this needs to be factored in to any evaluation of the changes between 2000 and 2009. Nonetheless, the decline in the reading achievement in Ireland between 2000 and 2009 is the largest in all participating countries (31 points[2]) and is statistically significant; the next largest country decline is 22 points. An examination of the proportions of students at each reading proficiency level suggests that the decline in performance is evenly distributed across levels, and so cannot be attributed to higher or lower achievers doing exceptionally poorly (Table 2; see also Figure 1). In 2000, just 11.0% of 15-year olds in Ireland scored at or below Level 1 on PISA reading literacy. By 2009, this had increased to 17.2%, indicating that there were almost 50% more 'poor' readers in 2009 than in 2000 (Figure 1). Similarly, in 2000, 14.2% of students were categorised as achieving at Level 5 (the highest level on the PISA 2000 combined reading literacy scale). By 2009, this had dropped to 7.0% achieving at or above Level 5 (a new category, Level 6 had been added, but the cut-off point for the bottom of Level 5 did not change).

[1] In practice, however, as discussed later in this report, the PISA Consortium linked reading performance in 2009 to performance in 2006, with backwards links to 2003 and 2000.

[2] These changes should be interpreted with respect to an OECD average of approximately 500 and a standard deviation of about 100 for each domain.

Table 2. Comparisons of Performance Benchmarks for Reading, Mathematics and Science (Ireland)

Reading	2000	2009	Change
Mean score (all)	527	496	-31
Mean score (males)	513	476	-37
Mean score (females)	542	515	-27
Gender difference	-29	-39	-10
% at or below Level 1 (all)	11.0	17.2	6.2
% at or below Level 1 (males)	13.5	23.2	9.7
% at or below Level 1 (females)	8.3	11.3	3.0
% at (or above) Level 5 (all)	14.2	7.0	-7.2
% at (or above) Level 5 (males)	11.2	4.5	-6.7
% at (or above) Level 5 (females)	17.4	9.6	-7.8
Mathematics	2003	2009	Change
Mean score (all)	503	487	-16
Mean score (males)	510	491	-19
Mean score (females)	495	483	-12
Gender difference	15.0	8.0	-7.0
% at or below Level 1 (all)	16.8	20.9	4.1
% at or below Level 1 (males)	15.0	20.6	5.6
% at or below Level 1 (females)	18.7	21.0	2.3
% at or above Level 5 (all)	11.3	6.7	-4.6
% at or above Level 5 (males)	13.7	8.1	-5.6
% at or above Level 5 (females)	9.0	5.1	-3.9
Science	2006	2009	Change
Mean score (all)	508	508	0
Mean score (males)	508	507	-1
Mean score (females)	508	509	1
Gender difference	0	-2	-2
% at or below Level 1 (all)	15.5	15.1	-0.4
% at or below Level 1 (males)	16.6	16.0	-0.6
% at or below Level 1 (females)	14.5	14.3	-0.2
% at or above Level 5 (all)	9.4	8.7	-0.7

This indicated that there were half as many ‘very good’ readers (Level 5 or higher) in 2009 as in 2000. It is of note that the number of boys achieving at or below Level 1 has increased by almost 10 percentage points to 23.2% while the equivalent change for girls is only 3 percentage points, with 11.3% at or below Level 1.

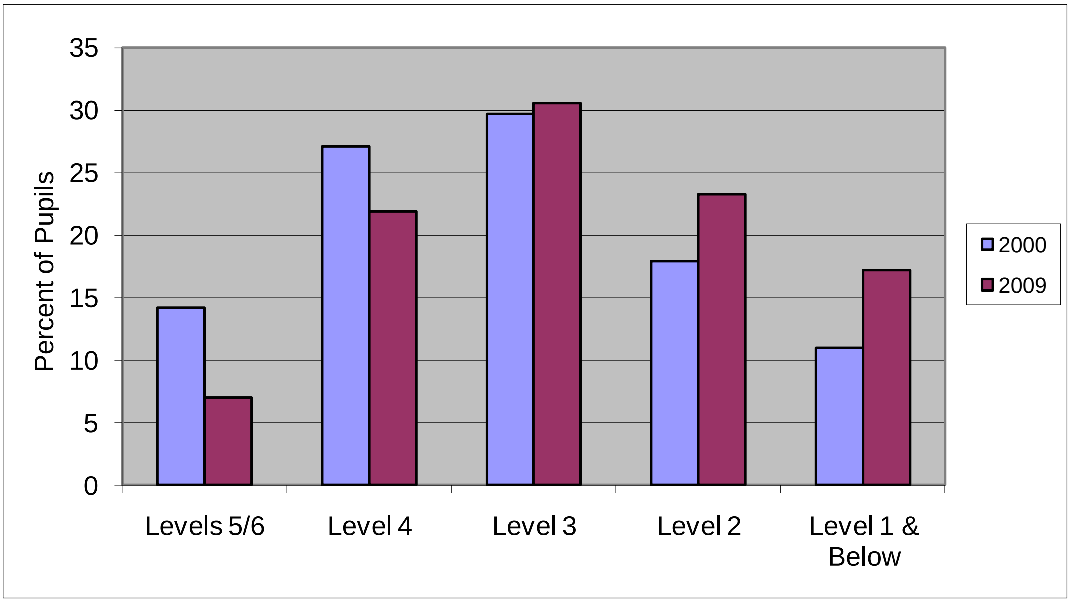


Figure 1: Distribution of Proficiency Levels on Combined Reading Literacy, by Year

As for mathematics, comparing PISA 2003 and PISA 2009, Ireland experienced a statistically significant decline in achievement of 16 scale score points – the second largest decline among participating countries (the largest decrease was 24 points). In the case of mathematics, therefore, the decline, while large (about one-sixth a 2003 international standard deviation) is half the size of the decline in reading (almost one-third of a 2000 international standard deviation).

As in reading literacy, there is evidence that the decline in mathematics performance is evenly distributed across the full spectrum of achievement, though the decline is marginally greater among high achievers. In 2003 16.8% of students in Ireland achieved at or below Level 1 (which, according to the OECD, indicates that they are not well prepared for the mathematical requirements of real life and further education), while in 2009, 21% of pupils achieved at or below Level 1 (Figure 2). Hence, more pupils achieved very low scores in PISA 2009 mathematics than in PISA 2003 mathematics. Conversely, in 2003, 11.3% of students in Ireland achieved at or above Level 5, compared with 6.7% in 2009.

In science, there has been no change in achievement since 2006, with mean scores of 508 in both years. The decrease in the proportion of students achieving at or below Level 1 (15.5% in 2006, 15.1% in 2009) is too small to be statistically significant. Similarly, the decline in the proportion achieving at or above Level 5 (9.4% in 2006, 8.7% in 2009) is again too small to be significant.

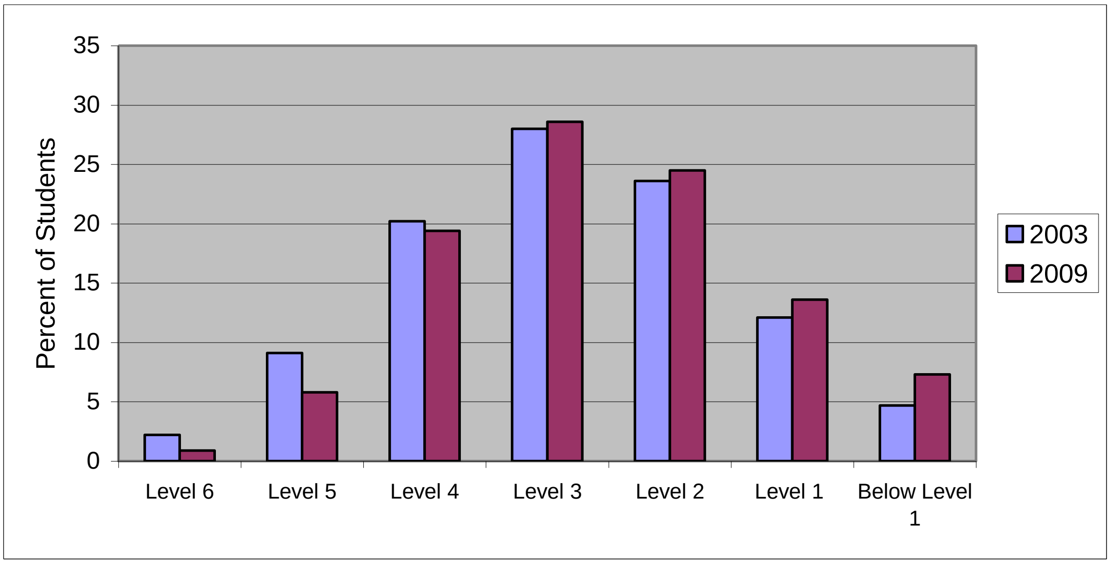


Figure 2: Distribution of Proficiency Levels on Combined Mathematical Literacy, by Year

In 2009, Ireland's mean score on reading literacy was not significantly different from the OECD average, whereas in 2000, 2003 and 2006 it was significantly higher. In 2009, Ireland's mean score on mathematics was significantly lower than the corresponding OECD country average, whereas in 2000, 2003 and 2006[3], it was not significantly different. In science, students in Ireland achieved a score that was significantly above the OECD average in 2009 – the same outcome as that achieved in 2000, 2003 and 2006.

The ranks for Ireland are shown in Table 3 for the period 2000-2009, based on all participating countries in PISA. It should be noted that the number of participating countries in PISA changes from cycle to cycle, and may also vary within cycles (e.g., a country may be included in the rankings for mathematics but not for reading if there is a technical problem with its achievement scores). In 2009, a number of high-scoring countries joined PISA, potentially lowering other countries' rankings. The number of OECD countries in PISA has also changed, with four new countries designated as OECD members in the international reports for PISA 2009 (Chile, Estonia[4], Israel and Slovenia). This means that rankings based on OECD member countries would also be expected to change (in the same way as OECD mean scores in 2009 now include contributions from these new or soon-to-be members). The rankings in Table 3 should be interpreted with reference to the accompanying rank ranges. Hence,

[3] For trend comparisons each domain in 2009 is compared with when it was a major domain, i.e. 2000-2009 for reading, 2003-2009 for mathematics, and 2006-2009 for science. The decision to monitor trends in this way was taken by the OECD.

[4] However, Estonia is still an accession candidate as of December, 2010 The others are now full member states, bringing the total to 33 member states, or 34 member and accession states. In this report, OECD 'member countries' include the 33 member states and Estonia.

for example, Ireland ranked 21st in reading among 65 participating countries in PISA 2009. We can be 95% certain that Ireland's rank is in the range of 15th to 27th.

Table 3 Ranks and Ranges of Ranks for Ireland in PISA Reading, Mathematics and Science (All Participating Countries)

Domain	2000 (31 countries)	2003 (40 countries)	2006 (57 countries)	2009 (65 countries)
Reading Literacy	5 (3-9)	7 (6-10)	6 (5-8)	21 (15-27)
Mathematics	15 (15-19)	20 (17-21)	22 (17-23)	32 (28-35)
Science	9 (9-12)	16 (13-18)	20 (15-22)	20 (16-23)

Table 4 shows Ireland's relative rankings for all domains and all cycles for the OECD, for EU25, and EU15 countries.

Table 4. Ranks for Ireland in PISA Reading, Mathematics and Science Relative to OECD, EU25 and EU15 Countries

Domain	Ireland's ranking relative to:											
	OECD countries				*EU25 countries*				*EU15 countries*			
	2000	*2003*	*2006*	*2009*	*2000*	*2003*	*2006*	*2009*	*2000*	*2003*	*2006*	*2009*
Science	9th	13th	14th	14th	4th	7th	10th	8th	4th	6th	7th	5th
Mathematics	15th	17th	16th	26th	8th	10th	11th	17th	8th	9th	8th	10th
Reading	5th	6th	5th	17th	2nd	2nd	2nd	8th	2nd	2nd	2nd	6th
Participating countries	27*	29**	30#	33##	18*	19**	23	23	14*	14**	15	15

* One further country, the Netherlands, participated, but school response rates were too low to permit computation of reliable data.

** One further country, the United Kingdom, participated, but response rates were too low to ensure reliable achievement estimates.

\# Reading literacy scores were not reported for the United States.

\#\# Now includes Chile, Israel and Slovenia (all joined in 2010) and Estonia. Without these countries, Ireland's relative rankings for 2009 of 30 OECD countries are, for science, mathematics and reading, respectively, are 12th, 23rd and 16th.

Again, the ranks should be interpreted in terms of a range of ranks, i.e. the point estimates are approximate. Also, although all domains and all cycles are shown, results should, strictly speaking, be compared only for major domain to major domain. Across the OECD, Ireland's mean score in reading now ranks it 17th compared to 5th in 2000. [5] Ireland's ranking for mathematics changed from 17th in 2003 to 26th (2003-2009) and, while for science, it was the 14th in both 2006 and 2009. However, Ireland's relative rankings on science and reading (but not mathematics) are somewhat higher amongst EU25 and EU15 countries.

[5] Note that the OECD makes reference to 34 OECD countries – this is because Estonia is currently of accession status.

The relatively large decline on PISA reading literacy in Ireland, the somewhat smaller decline in PISA mathematical literacy, and the stability in performance in scientific literacy suggest that a range of factors were implicated in influencing performance in each domain.

It is notable that male students accounted for a greater decline in achievement than females on PISA reading. Male students in Ireland had a mean score in 2009 that was some 39 points lower than in 2000, while female students had a mean score that was 27 points lower. Similarly, in mathematics, male students in 2009 had a mean score that was 19 points lower than in 2003, while females had a mean score that was 12 points lower. In the case of science, male students had a mean score that was one point lower in 2009 than in 2006, while female students had a mean score that was one point higher. Unlike reading literacy and mathematics, the differences in science are too small to reach statistical significance.

It might be noted that the mean reading literacy score for Northern Ireland decreased substantially between 2000 and 2006 by 24 points (from 519 to 495), while performance in mathematical literacy between 2003 and 2006 (from 514 to 494), and scientific literacy (524 in 2003 to 508 in 2006)[6] also fell (Bradshaw et al., 2007). To date, no satisfactory explanation for the change in performance in Northern Ireland has been provided, although the consistency in the declines across all three PISA domains suggests that the 2006 sample may have been different from earlier samples.

In addition to reporting performance on an overall achievement scale in each assessment cycle in which a domain is designated as ‘major’, PISA reports on performance on the components of the domain. Since reading is the major assessment domain in 2009, subdomain mean scores can be compared with the corresponding mean scores for 2000 (Table 5). It should be noted, however, that the OECD does not report these comparisons in Volume 5 of the international report on PISA 2009 (i.e., the volume that reports results on trends). In general, differences on subscales between 2000 and 2009 are similar to the overall difference of 31 points. It is noteworthy that the relative advantage on Reflect and Evaluate in Ireland held between 2000 and 2009, even though the actual mean score in 2009 dropped.

[6] However, the OECD advise against comparing mean scores for science between 2003 and 2006, since science was a minor domain in 2003 and a major domain in 2006, and it argues that comparisons should only be drawn between major and minor (rather than minor and major). It should also be noted that Northern Ireland sample is part of the UK sample, and is not adjudicated separately by the OECD).

Table 5. Mean Scores on Reading Literacy Subdomains in PISA 2000 and 2009 - Ireland

Subdomain		2000	2009	Diff (2009-2000)
Text Structure	Continuous	528 (3.2)	497 (3.3)	**-31**
	Non-continuous	530 (3.3)	496 (3.0)	**-34**
Aspect	*Retrieve	524 (3.3)	498 (3.3)	**-26**
	**Interpret	527 (3.3)	494 (3.0)	**-33**
	Reflect & evaluate	533 (3.1)	502 (3.1)	**-31**
Combined reading		527 (3.2)	496 (3.0)	**-31**

*Access and retrieve in 2009; **Integrate and interpret in 2009.
Significant differences are shown in bold.

Table 6. Mean Score Gender Differences on Reading Literacy Subdomains in PISA 2000 and PISA 2009 - Ireland

Subdomain		Gender Difference (Boys - Girls, 2000)	Gender Difference (Boys - Girls, 2009
Text Structure	Continuous	-34	**-41**
	Non-continuous	-17	**-39**
Aspect	*Retrieve	-23	**-44**
	**Interpret	-27	**-37**
	Reflect & evaluate	-37	**-38**
Combined reading	(All Items)	-29	**-39**

*Access and retrieve in 2009; **Integrate and interpret in 2009
Significant differences are shown in bold.

We can also look at differences in reading subscales across genders between 2000 and 2009. Gender differences have increased (as they have on average across OECD countries). It is noticeable that the gender difference has widened considerably on the non-continuous texts scale since 2000, with the difference increasing from -17 to -39 (Table 6). The gender difference has also increased noticeably on Retrieve items (from -23 to -44), and on combined or overall reading (from -29 to -39). However, this latter difference reflects a larger gender difference between males and females across OECD countries, which increased from -33 to -40 between 2000 and 2009. It is unclear why the overall gender difference in Ireland increased – i.e., whether it is a function of the sample of students selected, the nature of the test items on the 2009 test (which could favour girls more than boys), less persistence among boys in 2009, or some combination of these reasons.

This report, which should be viewed as a work in progress, investigates possible reasons for changes in achievement in Ireland, focusing principally on changes in reading achievement between 2000 and 2009. However, it should be acknowledged that the change in mathematics achievement also merits more in-depth exploration. Section 2 examines issues related to sampling and seeks to ascertain if changes in demography or changes in sampling procedures might be responsible for some of the observed changes in performance. Section 3 presents information on procedural and more general issues that might have affected the performance of students in Ireland on PISA 2009. Section 4 looks at issues related to equating and scaling and the contribution that these areas could have made to observed changes in performance. Section 5 presents conclusions and provides suggestions for additional analyses. Six appendices are included. Appendix A provides additional tables based on analyses of the PISA 2000 and PISA 2009 samples in Ireland and the corresponding school populations. Appendix B provides information on the scaling of PISA 2009 data from the PISA Consortium. This document is in draft form and has not yet been published. Appendix C provides an analysis of the test design for PISA reading literacy, and is intended to support the analyses reported in Section 4. Appendix D contains national and international item parameters for reading literacy in 2009. Appendix E describes aspects of the reading literacy frameworks used in PISA and Appendix F contains some additional analyses; these are referred to in the relevant sections of this report.

It should be noted that the ERC has been in contact with the OECD and the PISA Consortium regarding the size of the declines in average achievement in reading and mathematics in 2009. According to both the OECD and the Consortium, other countries in which large declines in achievement were found are satisfied with the explanations provided to them by the Consortium.

2. PISA 2000 and PISA 2009 Sampling Outcomes

The focus of this section is on the quality of the samples for PISA 2000 and PISA 2009, and on identifying any differences in the samples that might have an impact on the scores of students in Ireland in either year, but particularly in 2009.

In considering the quality of the samples for Ireland, it should be noted that both the 2000 and 2009 samples were approved by the PISA Consortium prior to implementing the studies. An important difference between 2000 and 2009, however, is that in 2000 the sample was drawn by the ERC (and checked and confirmed for its quality by Westat of the PISA Consortium) while the 2009 sample (as well as those for 2003 and 2006) was drawn by Westat. The response rates achieved by Ireland in all PISA cycles to date met the requirements set out by the OECD (a weighted response rate of 85% participation at school level and 80% at student level).

It should also be noted that Ireland took part in two international studies in Spring 2009 – the fourth cycle of PISA and the International Civics and Citizenship Education Study (ICCS). Since ICCS assesses students in Second Year (post-primary) and PISA assesses students who are 15 years of age (usually distributed over the Second, Third, Transition and Fifth years, with the majority in Third year), the two studies drew from the same pool of post-primary schools, using a split-sample design. The ICCS sample was drawn first, and then the PISA sample. No schools selected for ICCS were selected for PISA (or vice versa), although there was overlap across the studies with respect to selection of replacement schools. In practice, no school was asked to implement both studies.

Stratification Variables Used in PISA 2000 and PISA 2009

Both the PISA 2000 and PISA 2009 samples were stratified (grouped) on the basis of school size (number of 15-year olds in the school), school sector (secondary, community/comprehensive, vocational), and percentage of female students enrolled (a categorical variable). School size was an explicit stratifying variable, while school sector and percentage of female students were implicit. In 2009, an additional implicit stratifying variable, percent of students in the school with a Junior Certificate fee waiver (split into quartiles), was also used.

The 2000 and 2009 samples were compared on the distribution of stratification variables (enrolment size – small, medium, large; sector – community/comprehensive, secondary, and vocational; and gender composition – split into quartiles by percent female). This results in a total of 36 cells (3*3*4). Sampling fluctuations were minimal when comparisons were made between 2000 and 2009 (Table A1, Appendix A). However, it was found that:

- There was an increase in the percentage of males (from 37.9% to 43.4%), and a corresponding decrease in the percentage of females (from 51.0% to 44.0%), in large secondary schools between the 2000 and 2009 samples.
- There was a decrease in the percentage of males (from 16.3% to 12.5%), and an increase in the percentage of females (from 11.0% to 13.6%), in large community/comprehensive schools between the two samples.
- There was a small increase in the percentage of males (from 2.9% to 4.2%), and decrease of females (from 3.1% to 1.7%), in small vocational schools between the 2000 and 2009 samples.

The two samples were also compared to the respective populations on the basis of the same 36 cells. Again, these comparisons indicate that, at least on these stratification variables, the samples in both years are good representations of the corresponding populations (Tables A2 and A3).

The above two sets of comparisons are not particularly informative vis á vis the socioeconomic characteristics of the samples and whether these have changed for reasons relating to random fluctuations in sampling. Therefore, average DEIS[7] scores (originally computed in 2005) were compared for the two samples. The difference of 0.02 points is very small – less than one-fiftieth of a standard deviation. Table A4 (Appendix A) shows the components of the DEIS index and also includes a comparison with the ICCS sample. All samples (PISA 2000, PISA 2009 and ICCS) appear to be, on average, slightly less disadvantaged than the general population of schools. For example, the final mean DEIS scores for the PISA 2000 and PISA 2009 schools are 0.13 and 0.15 respectively, while, for the ICCS sample, the DEIS score is 0.18. The corresponding mean across all post-primary schools in the population is 0.00 with a standard deviation of 1.00.

Tables A6 and A7 (Appendix A) show the distributions of PISA 2000, PISA 2009 and ICCS schools for each decile on the DEIS index, compared to the population of schools (based on 2008 data). Again, comparisons indicate that the 2000 and 2009 samples are comparable with respect to the distribution of schools on this index. However, it might be noted that in PISA 2000, 15.2% of schools were in the two lowest DEIS deciles (most disadvantaged), compared with 17% in PISA 2009, while, conversely, 22.4% of schools were in the two highest deciles (most advantaged) in PISA 2000, compared with 19.4% in 2009.

[7] DEIS, Delivering Equality of Opportunity in Schools, is a scheme designed to address low achievement and other difficulties in the most disadvantaged schools in Ireland. Schools were assigned to the School Support Programme under DEIS and allocated resources based on the level of disadvantage, as measured by a combined variable (the DEIS index) consisting of average performance on the Junior Certificate Examination over a three-year period (2002-2004), average retention rates to Junior Certificate level over three years, and the percentage of pupils in the school with medical card access, again averaged across three years. The small number of schools missing DEIS scores – mainly private schools – were given a DEIS score equal to the mean.

Comparisons were made on other demographic characteristics of the achieved samples in 2000 and 2009. The following were analysed: response rates, exclusion rates due to special educational needs, percent of PISA-eligible students that had left school, gender, home language, distribution across year levels, rates of parental unemployment, socioeconomic background (as indicated by parental occupation), and average age. Results indicate the following (see Table 7):

- Student response rates are quite stable and meet OECD requirements. In 2000, 86% of selected students participated, compared with 83% in 2009. The requirement is 80%.
- Not unexpectedly, given that PISA uses an age-based sample, student age is very stable.
- Exclusion rates[8] were similar in 2009 (2.7%) relative to 2000 (2.8%) and within the requirements of the PISA technical standards on both occasions. It is unknown how many students identified as having a special educational need participated in PISA 2000, but in PISA 2009, 3.5% did so. Of these 139 students, 4 had a physical disability, 77 had a general learning disability or behavioural/emotional difficulty, and 58 a specific learning disability (these are PISA-defined categories of special educational needs).
- The rate of leaving school in 2009 of PISA-eligible students (1.5%) was marginally lower than in 2000 (2.1%). This is based on an analysis of the students selected for PISA in each school whose principal indicated that they had left school between submission of student data to the DES and the PISA assessment (about 6 months).
- The percent of female and male students (across all schools) is quite stable and close to 50% in both PISA 2000 and 2009.
- The percentages of students in Junior Cycle at the time of the PISA study are similar in 2000 (64.9%) and 2009 (62.5%). In contrast, the percentage in Transition Year was higher in 2009 (23.5%) compared with 2000 (16.0%) and this is mirrored by a lower percentage of students in a Leaving Certificate programme in 2009 (13.9%) compared with 2000 (18.6%).

[8] These students were excluded from testing by their school because they had a special educational need, or limited experience of the language of the test, that, in the view of the school principal, meant they could not attempt the test.

Table 7. Student-Level Demographic Characteristics of the Sample, PISA 2000, 2003, 2006 and 2009

Characteristic	2000	2009
Response Rate (student level)	85.6	83.0
% students with a special educational need participating	Unknown	3.5
*Within-school Exclusion Rate (%)	2.8	2.7
*% Left School	2.1	1.5
% Female	50.4	49.9
Mean age (SD)	15.7 (0.28)	15.7 (0.28)
% Other Language	0.9	3.6
Year level (weighted %)		
Second Year	3.3	2.3
Third Year	61.6	60.2
Fourth Year	16.0	23.5
Fifth Year	18.6	13.9
Occupation mean (SD)	48.4 (15.6)	49.92 (16.3)
Parental unemployment rate (%)	5.2	8.3

*Unweighted figures.

- The percentage of students speaking a language other than English or Irish has increased fourfold from 0.9% in 2000 to 3.6% in 2009[9].
- Student socioeconomic status as measured by parents' occupations was lower in 2000 – by just one-tenth of a standard deviation (note that this SES measure does not reflect parental (un)employment status).
- Parental unemployment rates were higher in 2009 (8.3%) compared with 2000 (5.2%) (though this indicator is not used by the OECD in conditioning test scores[10]).

Fluctuations in Sampling and Demographic Composition – Application of Adjustments for Language Spoken, Special Educational Needs, and Rates of Early School Leaving

It is possible to make broad adjustments one at a time for changes in (i) language spoken, (ii) special educational needs, and (iii) early school leaving. However, these adjustments should be interpreted with caution since they involve a number of assumptions; for example that these sub-groups of students are equivalent to one another in terms of achievement in both

[9] The OECD reports for PISA 2009 define 'other' language as a language other than the language of the test; hence students who report speaking Irish at home taking mathematics/science components of PISA 2009 in English are regarded as taking the test in another language and vice versa. According to the OECD, in PISA 2009, 5.6% of students in Ireland took the PISA test in a language other than the language spoken at home.

[10] Conditioning refers to a procedure whereby test scores are adjusted on the basis of background characteristics. This is done so as to allow unbiased analyses of subgroups of interest in the population.

2000 and 2009, which may not be the case. Furthermore, it cannot be assumed that these sub-groups are mutually exclusive.

Table 7 showed a fourfold increase in the percentage of students with a first language other than English or Irish. What would the average score for reading in 2009 for Ireland be if this increase had not taken place? Table 8 shows (i) the percentages of students in 2000 and 2009 by language status, (ii) their average achievement, and (iii) the estimated score change had the number of students speaking a language other than English or Irish remained the same as in 2000.

If the number of students with another first language had remained stable between 2000 and 2009 at 0.9%, this would have resulted in an increase of around 1.6 score points on PISA 2009 reading (i.e. (500.4 - 443.9)*.027). Note, however, that the issue is not clearcut since the composition of newcomers has changed. In fact the table shows that in 2000, students with another first language had a mean reading score marginally higher than (but not significantly different to) English/Irish speakers while in 2009 it was considerably lower. So a fairer adjustment to 2009 would be to assume that the other language group has the same average scores as English/Irish speakers in which case the increase would be 2.1 score points (i.e. (500.4 - 443.9)*.036).

Table 8. Percent of Other Language Speakers, Mean Reading Scores of Other Language and English/Irish Speakers, and Expected Change in PISA 2009 Reading Scores Adjusting For Characteristics of the 2000 Sample

Year	% other language	Mean score other language speakers	Mean score English/Irish speakers	Score change if 2009 the same as 2000	Score change if 2009 the same as 2000 assuming equivalent achievement
2000	0.9	532.8	527.4	--	--
2009	3.6	443.9	500.4	+1.6	+2.1

Since PISA 2000, there have been substantial policy developments in the area of special educational needs, perhaps most notably as documented in the EPSEN (2004) Act. One consequence of this is that the population of students with special educational needs has changed since more such students are being identified for the provision of extra support. Any adjustments relating to special educational needs should be treated with extreme caution, particularly given that the percentage of 15-year-olds in special schools has remained constant in the period from 1999-2006 (in the region of 1.1 to 1.3% of the population of 15-year-olds according to annual statistical reports of the Department of Education and Skills).

In PISA 2000, we know the percentage of students that were excluded due to special educational needs but we have no data on the percentages of students with special educational needs that did participate in the assessment. For PISA 2009 we have both sets of information. In 2000, 2.8% of students were excluded due to a special educational need or limited experience of the test language while in 2009 this was similar at 2.7%. As noted previously, in PISA 2009, 139 students with SEN did participate in the assessment (that is, 3.5% of all participating students).

Based on analyses by Cosgrove (2005, unpublished doctoral dissertation), it is estimated that students with special educational needs in ordinary schools in 2003 who did not participate in the PISA assessment achieved a Junior Certificate English score that was 4.22 standard deviations lower than students who did attend the assessment, or 1.27 standard deviations lower than all students sampled for PISA (that is present, absent, special educational needs and left school) for whom 2003 Junior Certificate data were available. The former (4.22 SD) implies a less conservative adjustment while the latter (1.27 SD) is more conservative.

Therefore it is estimated that, given 3.5% in 2009 with a special educational need, and using the Junior Certificate standard deviation differences as a basis to make an adjustment, that if no students with a special educational need had participated in 2009, the score point difference would increase by between 4.2 and 14.2 score points (adjustments made for 1.27 and 4.22 standard deviation units, respectively). This estimate should be interpreted with caution however since the standard deviation adjustment is not based on PISA but rather Junior Certificate English, and it also assumes that students participating and not participating in PISA due to a special educational need would have equivalent levels of achievement which is unlikely to be the case. Also, we don’t know how many students with a special educational need participated in 2000. Therefore the more conservative estimate of 4.2 points is preferred but should not be considered reliable due to the many assumptions applied in its estimation.

As noted in Table 7, it is estimated on the basis of the student tracking form data (i.e. lists of sampled students) that the rate of potentially PISA-eligible students who left school prior to the assessment has dropped by 0.6% from 2.1% in 2000 to 1.5% in 2009. Work by Cosgrove (2005, unpublished doctoral dissertation) examining the average Junior Certificate English scores of PISA-sampled students who left school prior to the PISA assessment in 2003 found that their average score was 4.31 standard deviations lower than students that did attend the assessment.

Therefore it is estimated that, given a reduction of 0.6% in 2009 of students leaving school, and using the Junior Certificate standard deviation differences as a basis to make an adjustment, if the rate of early school leaving had remained the same in 2009 as in 2000 and the achievement differences of these students had remained constant, the score point difference would increase by about 2.5 score points on the PISA 2009 scale. Note that as with the adjustment for special educational needs, a number of assumptions apply here.

Fluctuations in an Alternative Measure of Achievement: The Junior Certificate

So far, Section 2 has demonstrated that the PISA 2000 and 2009 samples are comparable to one another and to the population of schools in terms of stratification variables (school sector, school size, percentage of female students), SES (DEIS scores), and a number of student demographic characteristics.

However, to explore further whether the achieved samples for 2000 and 2009 might be otherwise biased in terms of an alternative measure of achievement such as examination results relative to the population, comparisons were made between the PISA 2000 sample and schools that did not participate in PISA 2000, and between the PISA 2009 sample and schools that did not participate in PISA 2009, in terms of Junior Certificate English results for all students for 2000 and 2009.

The average results and distributions of results for the PISA samples and the non-PISA samples in both 2000 and 2009 are almost identical, indicating that the samples are not biased with respect to Junior Certificate English performance. Similarly, differences with respect to overall performance on the Junior Certificate[11] are minimal. Table 9 shows that, among schools that participated in PISA 2009, the average Junior Certificate English score (EOPS) was 9.3, whereas it was 9.1 among schools not taking part in PISA 2009. Similarly, the average overall Junior Certificate scores (JOPS) for students in schools taking part in PISA 2009 was 65.9, whereas the corresponding average for schools not taking part was 64.9.

[11] The scales used here are Junior Certificate English Performance Scale (EOPS) and the Junior Certificate Overall Performance Scale (JOPS). These scales were constructed by allocating points to students, based on grades achieved in English (EOPS) and across nine Junior Certificate subjects (JOPS).

Table 9. Descriptive Statistics for JOPS (JCE overall Performance Scale) and EOPS (JCE English Performance Scale) by whether Pupils Are in a School that Participated in PISA 2000 and PISA 2009

PISA 2000	Score	N	Mean	SD	Skewness	Std. Error	Kurtosis	Std. Error
No	JOPS	42114	65.2	13.0	-1.222	0.012	2.208	0.024
	EOPS	41889	9.2	1.9	-0.763	0.012	0.454	0.024
Yes	JOPS	12110	65.2	12.9	-1.144	0.022	1.816	0.045
	EOPS	12062	9.2	1.9	-0.77	0.022	0.489	0.045
PISA 2009	**Score**	**N**	**Mean**	**SD**	**Skewness**	**Std. Error**	**Kurtosis**	**Std. Error**
No	JOPS	39890	64.9	13.1	-1.195	0.012	2.065	0.025
	EOPS	39668	9.1	1.9	-0.753	0.012	0.411	0.025
Yes	JOPS	14334	65.9	12.4	-1.223	0.02	2.259	0.041
	EOPS	14283	9.3	1.8	-0.783	0.02	0.577	0.041

The fact that ICCS took place at the same time as PISA 2009 and the consequent need to split the sampling frame in half was also examined in terms of performance on the Junior Certificate. Could it be the case, for example, that the lower amounts of replacement schools available for PISA would have affected the achieved sample? In PISA 2009, 160 schools were sampled, and 144 participated, including two 'replacement' schools (i.e. schools similar to the original ones on the sampling stratification variables). The concurrent administration of the ICCS study, however, meant that the list of schools had to be split into two equivalent groups to avoid sampling any school twice (i.e. for both PISA and ICCS). This has the effect of (i) reducing the availability of the replacement schools and (ii) increasing the selection probabilities of schools. A comparison of the average Junior Certificate Performance Scores for (i) the achieved sample of 144 schools, (ii) the 18 non-participating schools and (iii) all 160 schools as originally sampled indicates that the mean scores are not notably different at 9.27, 9.19 and 9.25, respectively, i.e. differing by at most 0.04 of a standard deviation.

Taken together, the outcomes of the analyses reported so far in Section 2 suggest that the PISA samples in both 2000 and 2009 are valid, representative and unbiased and also comparable to one another and to their respective populations. While the PISA 2000 and 2009 samples are comparable, factors such as a small increase in SES could affect conditioning/scaling of achievement data in a negative way[12], while changes in student demographics (e.g., an increase in the proportion speaking a language other than English/Irish at home, or a decrease in early school leaving) could also negatively impact on achievement. Section 4 on scaling discusses these issues further.

[12] Though it should be noted that, according to the OECD report on PISA 2009, the increase in average SES for Ireland (0.06 standard deviations) is not significant.

Sampling Fluctuations: The Distribution of Achievement and Socioeconomic Status

It is possible that the particular schools sampled in the two surveys might, due to other fluctuations, be somewhat different in terms of social composition and how this relates to achievement. It should be noted that an additional stratification variable was used in 2009 for the first time – school average Junior Certificate fee waiver (split into quartiles) – in order to improve sampling precision, and it is possible that this might affect the achieved sample in terms of characteristics other than those considered up to now. This stratum was introduced on the basis of participating in ICCS and the ICCS and PISA samples had to be drawn from a split sampling frame using the same stratification variables.

Multilevel modelling was used to examine whether changes have occurred in the extent to which schools differ from one another across PISA 2000 and PISA 2009. Results indicate that between-school variance in average reading achievement increased from 18% to 23% and between-school variance in average socioeconomic composition (indicated by parental occupation or HSEI) increased from 13% to 18%. Further, total variation in achievement has increased by 11.6% while the total variation in SES has increased by about 8.7% (Table 10).[13]

Table 10. Variance Components for Reading Achievement and Parental Occupation (HISEI), 2000 and 2009

Component	Achievement 2009	HISEI 2009
Between	2233.009	48.578
Within	7451.358	218.895
Total	9684.367	267.473
% between schools	23.1	18.2
	Achievement 2000	HISEI 2000
Between	1532.436	31.704
Within	7144.092	214.345
Total	8676.528	246.049
% between schools	17.7	12.9

This suggests a school sample in 2009 that is somewhat more stratified by both SES and achievement compared with 2000. Note, however, that the between-school variance in the OECD PISA report for 2009 is computed on the basis of a composite measure of Economic, Social and Cultural Status

[13] It should be noted that the variance components associated with achievement are slightly different to those reported in the OECD PISA 2009 international reports due to a difference in the methods used to compute those components. In the case of the analyses reported here, a student weight which includes the school weight components was applied at the student-level portion of the model, while the OECD PISA 2009 reports have split this weight into school and student components and applied these to their respective portions. Nonetheless, both sets of figures indicate an increase in between-school variation in achievement from 2000 to 2009.

(ESCS) and the figure for Ireland is 23.1%, lower than the OECD average of 25.5%. Results for ESCS in 2000 did not include a comparable analysis of variance.

However, the strength of the relationship between SES and achievement is the same in 2000 and 2009 (if one compares the two figures in the first row of Table 11). A further observation with respect to this modelling exercise is that gender is more strongly associated with achievement in 2009 compared with 2000 (a pattern that is also found across many other OECD countries) (Table 11).

Table 11. Percentage of Variation in Achievement Explained by (i) School and Student HISEI and (ii) School and Student HISEI and Gender, 2000 and 2009

Model	2000	2009
School & Student SES	14.7	14.5
School & Student SES & Student Gender	17.0	19.2

To investigate differences in between-school variance further, an examination of the distributions of school average reading achievement in 2000 and 2009 was carried out. In 2000, school average scores for reading, mathematics and science, respectively, ranged from 410-615, 413-583, and 385-598. In 2009, the ranges for reading, mathematics and science, respectively, were 302-589, 334-590, and 331-590. Thus, there is more variation in school average scores in all three domains in 2009 relative to 2000 (Figures 3 and 4).[14] Focusing on the left portion of Figures 3 and 4, one can see that generally, schools with low achievement in reading tended to have low achievement in the other two subject domains. However, in PISA 2000, there is a tendency for some schools to have a profile of mathematics achievement that is lower than that for reading and science, while in 2009, a few schools have low science scores relative to the other two domains. In PISA 2009, the eight schools with very low mean scores on reading were selected for further procedural analysis. These schools had a mean reading score more than 100 points below the national student average. Test administration records for these eight schools were examined and failed to show any issues with test administration. It is estimated that if these eight schools had not participated in PISA, the Irish average reading score would increase by up to 7 scale points (see also Tables F1-F3). Table 12 shows the school average scores in reading, mathematics and science for the eight lowest-scoring schools in 2000 and 2009. It can be seen that the scores are considerably lower in all three domains in 2009 compared with 2000. Also, there is somewhat more variation across domains in 2009 compared with 2000.

[14] In the Statistics Canada report reviewing the outcomes of PISA 2009 for Ireland, seven rather than eight outliers are identified, and this is due to the fact that in the present report, the outliers were identified on the basis of the first plausible value for reading only, whereas the Statistics Canada analysis used all five plausible values.

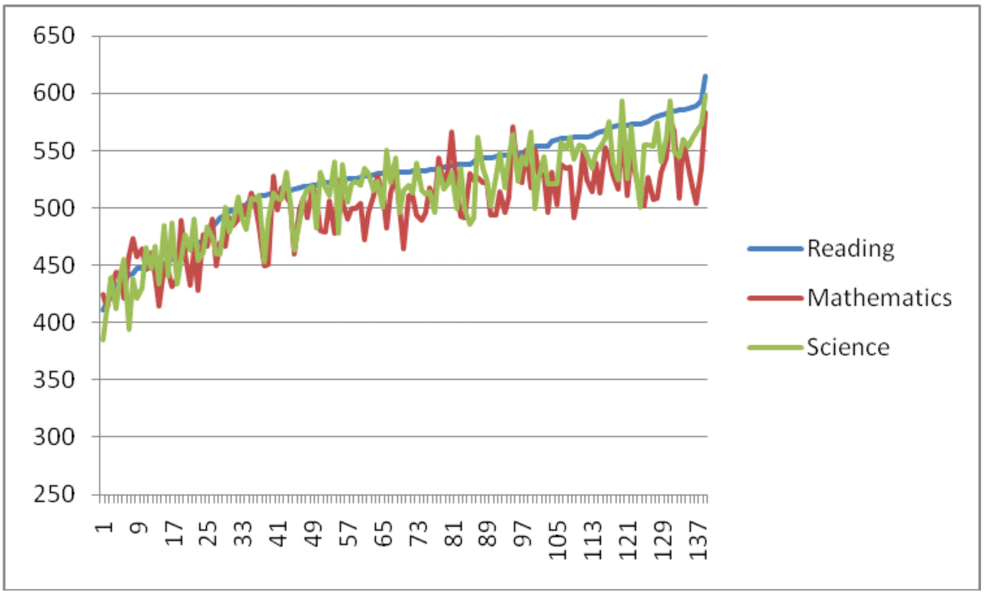


Figure 3. Distribution of School Mean Scores, PISA 2000

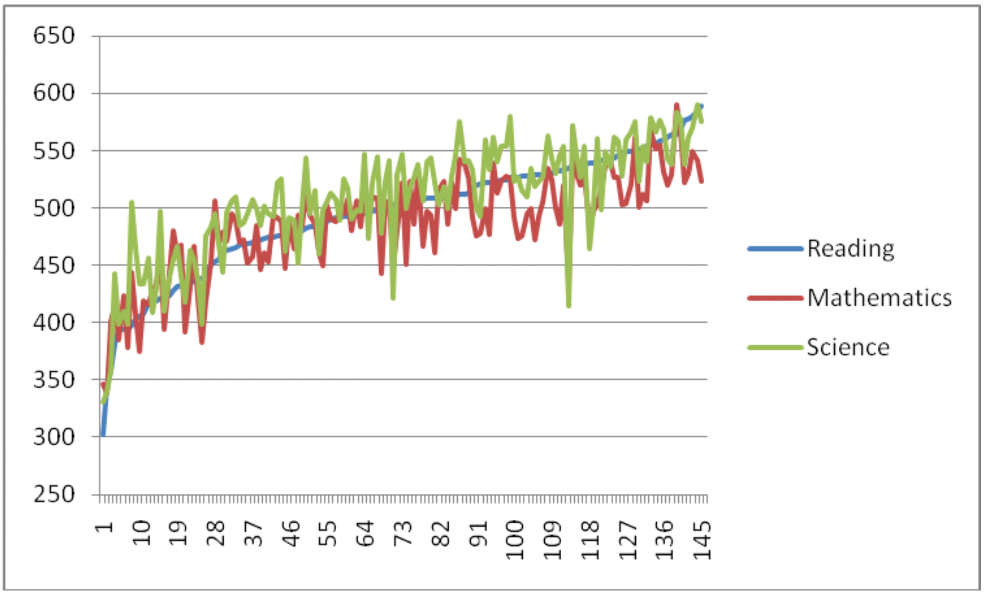


Figure 4. Distribution of School Mean Scores, PISA 2009

When analyses of variance components were conducted again on the 2009 dataset by excluding the eight very low performing schools, we find that

- between-school variation in achievement decreases from 23% to 18% (the latter being almost identical to the 2000 figure)
- total variation in achievement decreases by about 9%, again coming close to the 2000 value
- between-school variation in SES based on parent occupations decreases from about 18% to 17%, and overall variation in SES remains almost identical to the analyses that include the eight outliers.

Table 12. School average scores in reading, mathematics and science for the eight lowest-performing schools in reading in PISA 2000 and 2009

	2009			2000		
	Reading	Mathematics	Science	Reading	Mathematics	Science
	298.0	343.7	327.9	410.5	424.7	384.8
	344.7	337.9	339.5	417.3	412.6	412.7
	367.4	402.2	363.2	421.8	429.6	439.5
	378.4	410.7	440.5	432.2	444.4	412.1
	393.7	422.8	410.4	434.5	443.2	440.0
	393.9	381.3	394.5	437.5	421.5	455.1
	395.3	381.2	401.3	441.1	458.3	393.8
	396.7	442.3	503.1	442.8	473.7	438.3
(Average)	371.0	390.3	397.5	429.7	438.5	422.0

Estimates were computed using sampling weights and on the basis of the first of five plausible achievement scores.

Therefore, not surprisingly, the increase in between-school variance in achievement (estimated on the basis of all five plausible values) can be accounted for by the inclusion of the eight very low performing schools, but the increase in between-school and overall variation in SES is unrelated to the eight very low performing schools. It should also be noted that a comparison of the Junior Certificate results for very low performing and other schools indicates that the eight schools have an average English score that is 1.7 standard deviations below the other schools, and an average overall Junior Certificate score that is 2.0 standard deviations below the other schoolsd; therefore these eight schools are low achieving ones on a measure that is independent of PISA.

Appendix F contains some additional analyses of the characteristics of the 2009 very low performing schools. In summary, the results indicate that:

- Students in these eight schools had close to three times as many missing responses on their test booklets
- 30% of students in the eight schools skipped more than 25% of the questions presented to them
- The eight schools have a mean score in all three domains that is one standard deviation below the mean of the other schools
- The eight low performing schools have a mean SES score (based on parental occupation) that is 0.7 standard deviations below the other schools, and twice the rate of Junior Certificate fee waiver
- The eight schools have fewer girls
- These schools have 14% of ‘other’ language speakers compared with 3% of the other schools

- Participation rates are lower in the eight low-scoring schools compared to others, at 57% and 84%, respectively.
- The eight schools tend to be vocational and had more female test administrators.

Broadly speaking, it can be concluded that the eight low performing schools in the 2009 sample arise from a combination of low SES, high proportions of newcomer students, and lower proportions of girls, while low participation rates in PISA should also be factored in.

It should also be noted that two of these schools (38 students) were missing questionnaire data and this may have had an effect on conditioning. However, ACER (personal communication, September 28, 2010) has confirmed, through a comparison of school average plausible values and WLEs (two versions of the achievement scores that they compute) for all participating schools in Ireland in PISA 2009, that the conditioning process has had no effect on the means for schools in general, including the eight outliers.

Finally, drawing on CSO quarterly household survey data, we wanted to examine the possibility that observed SES changes in PISA may wholly or in part be simply mirroring changes in the population.

Table 13. Distribution of Persons by Major Occupational Grouping, Spring 2000 and Spring 2009, Males, Females, and All

Group	Males		Females		All	
	Apr-June 2000	Apr-June 2009	Apr-June 2000	Apr-June 2009	Apr-June 2000	Apr-June 2009
1. Managers and administrators	22.1	20.8	12.2	11.7	18.1	16.7
2. Professional	9.2	11.6	10.8	13.9	9.9	12.7
3. Associate professional and technical	6.3	7.7	11.1	12.5	8.3	9.9
4. Clerical and secretarial	5.0	5.6	22.8	21.1	12.2	12.7
5. Craft and related	21.3	19.2	2.1	0.9	13.4	10.9
6. Personal and protective service	6.9	8.4	14.1	17.3	9.9	12.5
7. Sales	5.3	6.2	12.1	12.0	8.1	8.8
8. Plant and machine operatives	13.9	11.3	6.8	2.4	11.0	7.2
9. Other	9.9	9.1	7.9	7.9	9.1	8.6
Total	100.0	100.0	100.0	100.0	100.0	100.0

Source: CSO.

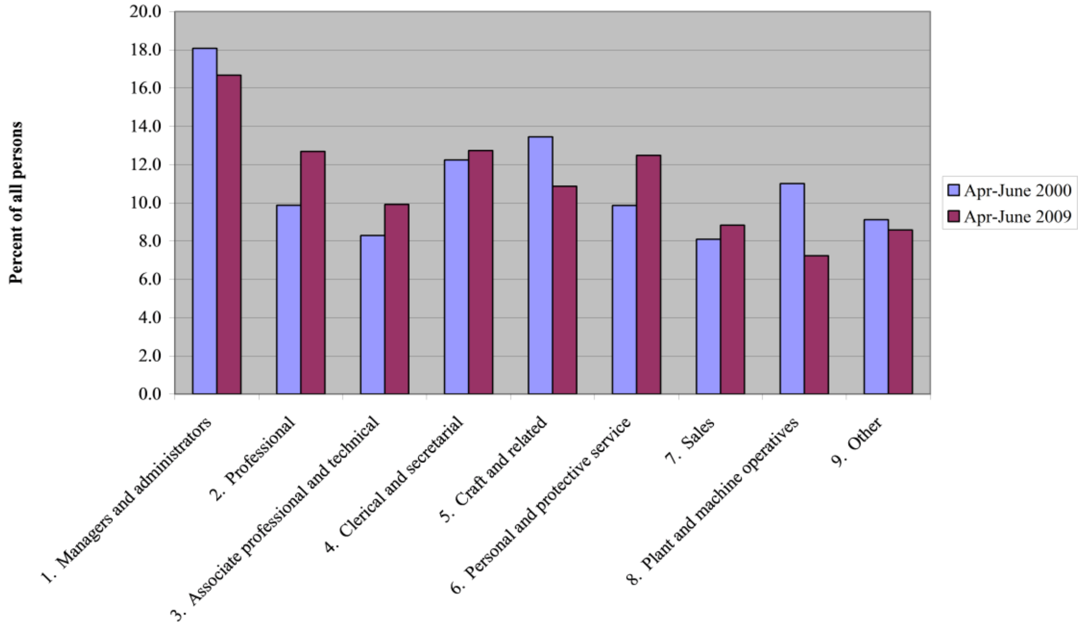


Figure 5. Distribution of All Persons by Major Occupational Grouping, Spring 2000 and Spring 2009
Source: CSO.

Table 13 (Figure 5) compares the distribution of persons across occupational groups for the two periods of spring 2000 and spring 2009. It would appear that the distribution across occupational groupings has not changed appreciably during this time period, although it should be noted that the data in Table 12 are very general in nature; PISA provides a more precise measure of SES.

Achievement Fluctuations in Specific Sub-Groups

It has already been noted that the gender gap in achievement in reading has increased, with some evidence of a larger decline in boys' reading achievement compared to that of girls. Here, we examine, in addition, achievement by immigrant status, language, and grade level.

The percentage of students in Ireland with immigrant status increased from 2.3% in 2000 to 8.3% in 2009. The mean score of immigrant students was 79 points lower in 2009 than in 2000. The percentage with a different home language to the language of the assessment was 2.0% in 2000 and 5.8% in 2009 (though this also includes students who speak Gaeilge at home and did the test in English and vice versa). There was a decline of 70 points in the mean score of students who spoke a different language to the language of assessment between 2000 and 2009 (Table 14).

Table 14. Changes in Background Variables Reported by OECD (2000-2009) and Performance on Reading Literacy

Students with. . .	Ireland 2000		Ireland 2009		Difference (2009-2000)	
	%	Mean	%	Mean	%	Mean
Immigrant Status	2.3	552	8.3	473	**+6.0**	**-79**
Different Home Language to Language of the Assesment	2.0	537	5.8	467	**+3.8**	**-70**

Significant differences are in bold.

Tables 15 and 16 present this information with reference to the OECD averages. Table 15 shows that in 2000, there were fewer immigrants in Ireland compared to the OECD average, while figures are similar for Ireland and the OECD average in 2009. Furthermore, there was a (non-significant) achievement difference of 24 points is in favour of immigrant students in 2000, compared with 46 points on average in favour of non-immigrants across OECD countries. In 2009, the trend for Ireland reversed, with non-immigrant students having a mean reading score that is 29 points higher than immigrants, although this is lower than the OECD average score difference of 44 points.

Table 15. Achievement differences in reading literacy in immigrant and non-immigrant students, Ireland and OECD average, 2000 and 2009

Comparison	Achievement difference 2009 (Non-immigrant – Immigrant)	Percent of immigrants 2009
Ireland	**29.0**	8.3
OECD	**43.7**	10.0
Comparison	Achievement difference 2000 (Non-immigrant – Immigrant)	Percent of immigrants 2000
Ireland	-24.0	2.3
OECD	**46.2**	8.3

Significant differences are shown in bold.
Comparisons are made on the basis of 27 OECD countries.

Table 16 shows patterns of results by language use that are similar to those shown in Table 15, i.e. in 2000, no difference in the mean scores of test and other language speakers in Ireland compared with a 55-point score difference in favour of test language speakers across the OECD on average, and an increase in the percentage of other language speakers in Ireland in 2009, now similar to the OECD average, with a significant (34-point) difference in favour of test language speakers in Ireland. Note that the 34-point difference between test and other language speakers is smaller than that across the OECD on average (50 points).

Table 16. Achievement differences in reading literacy in students speaking the language of the test and students speaking another language, Ireland and OECD average, 2000 and 2009

Comparison	Achievement difference 2009 (Speaks – Doesn't Speak Test Language	Percent of other language speakers 2009
Ireland	**34.0**	5.8
OECD	**50.0**	7.6
Comparison	Achievement difference 2000 (Speaks – Doesn't Speak Test Language	Percent of other language speakers 2000
Ireland	-9.0	2.0
OECD	**55.0**	7.5

Significant differences are in bold.
Comparisons are made on the basis of 27 OECD countries.

It is worth noting that in 2000, the immigrant and other language populations had substantially higher SES scores (based on parental occupation) than 'native' English/Irish speakers, while in 2009, the groups do not differ on average SES (Table 17). In other words, the SES advantage associated with membership of the migrant and/or other language group that was evident in 2000 had disappeared in 2009.

Table 17. Differences in average SES scores by immigrant and language group, 2000 and 2009

Characteristic	2000		2009	
	Mean	SD	Mean	SD
Native	48.2	15.7	49.8	16.1
Immigrant	54.5	18.2	52.2	18.7
Total	48.4	15.9	50.0	16.3
	Mean	SD	Mean	SD
Test language	48.3	15.8	49.9	16.1
Other language	58.1	22.5	50.6	18.7
Total	48.4	15.9	49.9	16.3

Table 18 shows mean achievement scores by grade level for all three domains, along with percentages of students by grade level, across all four PISA cycles. There has been a marked increase of students in Transition Year (from 16.0% in 2000 to 24.0% in 2009) and a corresponding decrease of students in Fifth Year (from 18.6% in 2000 to 14.4% in 2009). Although the mean reading scores of students in all grade levels declined significantly between 2000 and 2009, the drop was greatest for Fifth Year students (from 547.9 to 498.2) and smallest for those in Third Year (from 516.9 to 487.9). There was a significant drop in the mean mathematics scores for Third, Transition and Fifth Year students between 2003 and 2009, and this drop was smallest in Third year (from 492.3 to 480.1), and largest in Transition Year (from 537.3 to 509.5). There were no significant changes in the mean science scores of students at any grade level between 2006 and 2009. In science, Third Year students achieved a marginally higher score in 2009 than in 2006 (501.7 and 499.3, respectively).

Table 18. Distribution of students across grade levels, and mean achievement in reading, mathematics and science, all domains

Grade Level	2000%		2003%		2006%		2009%		
Second Year	3.3		2.8		2.7		2.4		
Third Year	62.0		60.9		58.5		59.1		
Transition Year	16.0		16.7		21.2		24.0		
Fifth Year	18.6		19.6		17.5		14.4		
				Reading scores					
	2000		2003		2006		2009		Diff 2000-2009
	Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean
Second Year	410.7	9.55	406.2	10.01	420.2	13.06	376.0	10.88	**34.7**
Third Year	516.9	3.60	502.8	3.23	506.9	3.85	487.9	3.43	**29.0**
Transition Year	568.4	4.52	562.0	4.48	547.8	4.70	525.3	4.42	**43.1**
Fifth Year	547.9	4.30	530.8	4.36	530.9	4.56	498.2	5.51	**49.7**
				Mathematics scores					Diff 2003-2009
Second Year	409.1	12.14	406.8	9.48	414.9	9.54	384.8	11.63	22.00
Third Year	495.4	3.11	492.3	2.97	492.3	2.95	480.1	3.07	**12.20**
Transition Year	537.3	5.72	542.9	4.56	530.1	4.30	509.5	3.88	**33.40**
Fifth Year	516.6	4.48	515.1	5.32	511.5	4.18	496.1	4.86	**19.00**
				Science scores					Diff 2006-2009
Second Year	425.8	10.49	400.5	9.95	408.5	11.0	403.7	10.24	4.80
Third Year	504.6	3.86	494.1	3.30	499.3	3.5	501.7	3.74	-2.40
Transition Year	550.9	5.61	548.6	4.71	537.1	4.3	532.9	4.93	4.20
Fifth Year	529.6	5.15	518.8	5.23	519.6	4.3	510.0	5.57	9.60

The relatively large decline in reading in Fifth year in 2009 may be due to a shift of more able students from Fifth year to Transition year (with relatively weaker students proceeding directly to Fifth year). The larger decline among Transition year in mathematics may reflect the fact that students in Transition year may not be exposed to systematic mathematics instruction in the same way as they would in Third or Fifth years. It is also possible that Transition year may include some students who would previously have left school or awaiting the official school-leaving age of 16 years, thus reducing performance in Mathematics.

Table 19 shows the mean SES scores by grade level. These fluctuate somewhat but only to a minor degree. Having said this, the SES measure does not fully capture aspects of socioeconomic background, e.g. rates of current unemployment. The table shows that the change of the distribution of students by grade level is significant in all cases except Second year, while the small changes in average SES (occupation) are significant only in the case of Third year where the mean SES has increased. However, it might be noted that the trend for Transition year is an increase in SES while in Fifth year there has been a small increase since 2000.

Table 19. Distribution of students across grade levels and mean SES scores, 2000 and 2009

	2000		2009		2000		2009		
	%	SE	%	SE	Mean SES	SE	Mean SES	SE	Diff
Second Year	3.3	0.33	2.4	0.34	42.4	1.68	41.2	2.06	-1.20
Third Year	62.0	1.01	**59.1**	**1.03**	48.3	0.50	**50.1**	**0.55**	+1.80
Transition Year	16.0	1.26	**24.0**	**1.44**	51.4	0.93	52.5	0.80	+1.10
Fifth Year	18.6	1.31	**14.4**	**1.10**	47.4	0.75	46.4	0.81	-1.00

Note: Significant differences are in bold in the 2009 columns.

Section 3. Procedural and General Changes Associated with PISA 2009

As noted in the previous section, an additional stratifying variable (SES, based on rates of Junior Certificate fee waiver, split into quartiles) was used in sampling. However, the extent to which this change is associated with changes in performance in reading and mathematical literacy is unclear and it is acknowledged that more work is required to disentangle the effects of including the SES stratifier from other potential demographic changes in the achieved sample. Furthermore, there may be complex interactions between the procedural changes noted in this section and other changes that may prove difficult or impossible to disentangle.

In order to incentivise student participation (which has been a problem in terms of attaining response rate standards), for the first time in 2009, participating students in each school were entered into a 'draw' and three students per school received a €15 euro voucher. It is unclear if this served to attract a somewhat higher number of disengaged lower achievers (who might otherwise be absent on the day of the assessment) but the analyses of the sampling outcomes shown in Section 2 suggest that this is unlikely to be the case.

A second change in PISA 2009 was that in 76% of schools, teachers in the school[15] administered the assessment, while in the remaining 24%, external staff administered the assessment. Previously, external staff administered PISA in all schools. This change, which is in line with practice in some other PISA countries, was made in order to increase schools' engagement, which had been a significant problem in previous cycles of PISA. All individuals administering the assessment received training in how to do so by ERC staff. Independent quality monitoring in six of these schools indicated no problems with adherence to test administration procedures. The mean score of students in schools where a teacher administered the test was 5 points or about one-twentieth of a standard deviation lower than in schools with an external administrator but, on its own, this difference is not statistically significant. Assuming all other procedures and conditions in PISA 2009 were equivalent to PISA 2000, and assuming external test administrators had administered the assessment in all PISA 2009 schools rather than in just 24%, there would have been a 3.8 point increase in the overall average score for 2009.

However, this result should be interpreted with respect to the phenomenon of survey fatigue which may be more widespread in Ireland currently than it was in 2000. Some PISA 2009 schools participated in the TALIS main study and the ICCS field trial the previous year and this may have had a further effect on schools' engagement with PISA 2009. It is possible that school staff may have

[15] Note that it was stipulated that the teachers who administered the test could not be one of the students' subject teachers.

implicitly or otherwise communicated this fatigue or disengagement to the students, but of course it is not possible to quantify this or otherwise verify it. It is notable in this regard that the percentage of students skipping 25% or more of questions in their booklets has increased (see the rest of this section for further discussion of this issue).

Appendix F (Tables F8 and F9) provides more detailed comparisons of schools with internal and external test administrators. The difference worth noting is that schools with an external test administrator are more advantaged in terms of SES, as evidenced in both student occupational scores and school average Junior Certificate fee waiver; hence the small observed performance difference is likely to be accounted for by differences in SES.

A third change in PISA 2009 was that for the first time, an electronic reading assessment was administered to some students in each school (selected at random) after they had completed the two-hour paper assessment and 40-minute questionnaire. This might have resulted in a reduced level of motivation for taking the paper assessment among these students. However, a comparison of the mean (paper-based assessment) scores of students who did not participate in the electronic assessment with those who did participate in it indicates that the means are more or less identical for reading, mathematics and science and this is true for both males and females, with differences as small as 2 score points or less (i.e. less than one-fiftieth of a standard deviation).

For the first time in PISA 2009, information was collected from test administrators on whether any students in the assessment session appeared to sit without attempting any questions. This information was not collected in 2000. However, an analysis of student percent missing scores indicates that there has been an increase in skipping of questions. For example, in 2000, 2.6% of students skipped more than 25% of the questions, and in 2009 this was double, at 5.2%. Table 20 shows the average percentage of missing items (all domains) by gender and language spoken, together with the correlation between the percentage of missing items with SES (parental occupation). There has been an increase in the average number of skipped responses, notably for males and students speaking a language other than the test language, while the correlation with SES is similar in both 2000 and 2009.

Table 20. Percentage of Missing Responses (all items, across all domains) in 2000 and 2009, by Gender and Test Language and Correlation with SES

Characteristic	2000	2009
Gender		
Males	6.3	8.0
Females	4.4	5.4
Language		
Test language	5.3	6.2
Other language	6.5	11.9
SES (r)	-.195	-.200

More detailed analyses of missing responses were conducted on the PISA 2000 and 2009 data. A regression analysis was conducted with percent of missing responses as the outcome. It should be noted that although some school-level variables were included, single- rather than multi-level analysis was conducted. We included the following explanatory variables:

- Student gender (0=male, 1=female)
- Student SES (parental occupation, mean = 0, SD=1)
- Language spoken by student (0=English/Irish, 1=other)
- Test administrator gender (0=male, 1=female)
- Test administrator position (three dummies, principal/deputy principal; careers guidance or SEN teacher, external administrator; with subject teacher as the reference group)
- School SES (student average parental occupation, mean = 0, SD = 1)
- School percentage female (two dummies, highest and lowest quartiles, with interquartile range as the reference group)
- School sector (two dummies, community/comprehensive and vocational, with secondary as the reference group).

Tested together (with corrected standard errors, and testing for overall change in the R-squared statistic in the case of dummy variable sets), the final model shown in Table 21 indicates that student gender, student SES, language spoken, and school SES account for about 7.4% of the variation in missing responses and that the remaining variables listed above are not significant.

Table 21. Final regression model of student percent missing responses: PISA 2009

Variable	PE	SE	CI95L	CI95U	R2
Student gender (male-female)	-2.558	0.077	-2.709	-2.407	.074
Student SES	-1.255	0.043	-1.339	-1.171	
Language (English/Irish-Other)	4.397	0.207	3.991	4.803	
School SES	-1.209	0.044	-1.295	-1.123	

Table 22 shows a regression of missing responses using the same variables that were in the 2009 model for PISA 2000. The results are similar with respect to student and school SES, but it can be seen that the parameter estimate associated with gender is bigger in 2009 suggesting a somewhat increased level of skipped responses among boys compared with 2000. Also, the parameter estimate associated with language much smaller in 2000 compared to 2009.

Table 22. Final regression model of student percent missing responses: PISA 2000

Variable	PE	SE	CI95L	CI95U	R2
Student gender (male-female)	-1.775	0.060	-1.893	-1.658	0.079
Student SES	-0.929	0.033	-0.993	-0.865	
Language (English/Irish-Other)	1.611	0.308	1.007	2.215	
School SES	-1.258	0.033	-1.323	-1.194	

To assess whether the observed increase in missing responses is a phenomenon largely confined to Ireland, the international student percentages of skipped responses by booklet for 2000 and 2009 were compared. Table 23 compares the weighted student-level percent of missing items (not reached items are not treated as missing for this analysis; which is different to the Statistics Canada report) for countries participating in both 2000 and 2009, as well as the proportion of students skipping more than 25% and 50% of test items in 2000 and 2009. Across the 32 countries that can be compared (note that this does not include countries administering PISA 2000 in 2001), there has been a decrease of almost 2% in missing responses, a decrease of 1.5% in students skipping more than 25% of questions, and a small increase of 0.2% of students skipping more than 50% of questions. The data in the table indicate that:

1. Percent of missing responses increased in 6 of the 32 countries between 2000 and 2009 and Ireland has the second highest increase in percent missing.
2. Percent of students skipping more than 25% of questions increased in 12 countries, and Ireland has the third highest increase on this indicator.
3. Percent of students skipping more than 50% of questions increased in 24 countries, and Ireland has the eighth highest increase on this indicator.

Table 23. Patterns of student-level missing responses (all domains) to PISA test items 2000 and 2009 for countries participating in both cycles

Country	2000 % missing	2000 >25% missing	2000 >50% missing	2009 % missing	2009 >25% missing	2009 >50% missing	Diff % missing	Diff >25% missing	Diff >50% missing
Australia	6.08	4.11	0.20	5.33	4.43	0.47	0.76	-0.31	-0.27
Austria	7.50	5.21	0.19	8.99	9.96	0.93	-1.49	-4.75	-0.75
Belgium	7.72	6.69	0.28	6.18	5.79	0.60	1.54	0.90	-0.32
Brazil	15.46	18.41	0.93	3.96	4.17	0.39	11.50	14.25	0.53
Canada	4.83	2.24	0.04	4.35	2.71	0.27	0.48	-0.47	-0.23
Czech Republic	9.55	7.01	0.22	10.07	10.09	0.55	-0.52	-3.08	-0.33
Denmark	10.64	10.82	0.64	7.53	6.50	0.50	3.11	4.32	0.15
Finland	5.61	2.91	0.11	4.37	2.38	0.27	1.24	0.52	-0.16
France	9.11	6.39	0.22	9.87	10.19	1.21	-0.76	-3.79	-0.99
Germany	10.76	11.49	0.48	7.33	7.00	0.71	3.43	4.49	-0.23
Greece	12.62	14.06	0.80	10.56	11.65	0.92	2.06	2.41	-0.12
Hungary	10.46	8.95	0.32	7.99	6.69	0.13	2.47	2.26	0.19
Iceland	7.54	5.36	0.37	6.13	5.22	0.80	1.40	0.14	-0.43
Ireland	5.40	2.89	0.20	6.71	5.99	0.56	-1.31	-3.10	-0.36
Italy	12.29	12.77	0.47	9.57	10.32	0.82	2.72	2.45	-0.35
Japan	9.66	9.14	0.21	8.38	9.16	0.95	1.28	-0.02	-0.73
Korea	5.09	2.26	0.07	3.90	2.40	0.14	1.19	-0.14	-0.07
Liechtenstein	10.44	10.69	0.63	6.30	3.90	0.00	4.14	6.79	0.63
Luxembourg	13.12	16.04	1.50	9.51	10.63	1.25	3.61	5.41	0.25
Mexico	9.44	6.19	0.05	2.57	1.52	0.09	6.87	4.67	-0.04
Netherlands	2.23	0.36	0.00	2.07	0.73	0.09	0.16	-0.36	-0.09
New Zealand	5.56	3.93	0.14	5.37	4.52	0.36	0.19	-0.60	-0.22
Norway	9.11	8.21	0.40	7.55	6.67	0.54	1.55	1.54	-0.14
Poland	12.24	14.49	1.09	7.11	5.73	0.20	5.13	8.77	0.89
Portugal	10.70	10.25	0.20	7.55	6.18	0.45	3.15	4.07	-0.25
Russian Federation	11.30	11.99	0.60	9.90	10.60	0.78	1.40	1.39	-0.18
Spain	8.45	6.18	0.21	8.82	7.95	0.93	-0.37	-1.77	-0.72
Sweden	8.19	6.05	0.14	8.14	8.45	0.79	0.05	-2.40	-0.64
Switzerland	9.28	8.96	0.41	6.63	5.85	0.52	2.65	3.11	-0.11
United Kingdom	6.20	4.08	0.09	6.58	5.15	0.57	-0.38	-1.07	-0.47
United States	4.94	3.14	0.16	2.61	1.13	0.04	2.33	2.01	0.12
Average	8.76	7.78	0.37	6.84	6.25	0.54	1.92	1.54	-0.18

These data support the hypothesis of lower student engagement with the test and/or lower proficiency in 2009 relative to 2000.

Table 24 compares the percentages of skipped questions for 2000 and 2009 by domain for Ireland only. In the case of reading and mathematics, the instances of skipped responses have increased markedly since 2000. For example, close to 1% of students in 2009 skipped more than half of the reading items presented to them, compared with just 0.24% in 2000; in the case of mathematics, 2.5% skipped over half of the questions presented to them compared to just under 1% in 2000. In science, there is again a higher rate of skipping over 50% of questions in 2009 (0.2% compared with 0.9%). However, on the other two indicators of missingness, science is lower in 2009 than 2000, which is the reverse trend compared to reading and mathematics. This is an unexpected finding that may indicate differential rates of engagement/proficiency in science, mathematics and reading in 2009, or it may relate to the particular item set used to assess each domain in 2000 and 2009. To investigate this further, an analysis by domain and item type (multiple choice versus written response) was conducted for all countries participating in PISA 2000 and PISA 2009. Table 25 shows the results of this analysis.

Table 24. Comparison of missing responses by assessment domain, Ireland, 2000 and 2009

Indicator	2009		2000	
	Mean	SD	Mean	SD
% missing reading	5.2	8.9	4.2	7.3
>25% missing reading	4.3	20.4	2.5	15.6
>50% missing reading	0.90	9.5	0.24	4.8
% missing mathematics	10.4	14.2	11.1	12.6
>25% missing mathematics	13.7	34.4	14.5	35.2
>50% missing mathematics	2.5	15.6	0.92	9.6
% missing science	3.9	10.0	6.8	10.9
>25% missing science	3.5	18.3	7.0	25.5
>50% missing science	0.89	9.4	0.20	7.7

Looking first at the international averages for science, it can be seen that there has been a marked drop in the percentage of skipped written response items from 2000 to 2009 and the Irish data for science follow this international trend. In the case of reading, internationally, there has been an increase in the skipping of written responses, but whereas Ireland had below-average rates of missing on both types of reading items in 2000, in 2009 these were around the international average. In the case of mathematics, both in Ireland and internationally, rates of skipped items were quite high for both item types, and the Irish averages were below the international ones in 2000. However in 2009, the rate of skipped responses was much lower for multiple choice items and higher for written response items, and the Irish averages are similar to the corresponding international averages. These results suggest that changes in patterns of skipped responses may in part be a function of the test design.

Table 25. Comparison of missing responses by assessment domain and item type, for countries participating in both cycles

Country	2000						2009					
	Reading		Mathematics		Science		Reading		Mathematics		Science	
	MC	WR	MC	WR	MC	WR	MC	WR	MC	WR	MC	WR
Australia	4.1	5.6	9.4	11.2	1.8	23.9	1.5	8.8	1.9	13.8	3.2	4.5
Austria	5.6	6.7	10.4	12.8	1.9	26.9	3.3	15.1	3.5	17.4	6.7	9.5
Belgium	6.0	7.1	10.6	11.5	3.3	13.4	2.2	10.3	2.6	14.6	4.4	6.3
Brazil	12.5	14.2	20.4	26.7	6.1	18.0	2.5	11.1	4.9	22.4	5.8	7.3
Canada	3.4	4.4	8.7	9.1	1.4	28.9	1.3	7.5	1.5	11.3	2.4	3.4
Czech Republic	6.5	8.7	16.1	18.1	3.1	32.8	2.8	17.0	3.8	25.3	7.1	9.6
Denmark	8.1	9.1	14.1	18.0	4.5	19.9	2.1	12.0	2.5	18.8	5.6	7.9
Finland	3.6	4.5	11.5	13.4	1.6	10.2	1.1	7.3	1.7	11.5	2.8	3.2
France	6.4	8.7	14.4	13.0	3.1	25.6	3.3	16.2	3.2	21.9	6.4	9.8
Germany	8.3	9.7	14.2	18.7	3.9	25.7	2.3	13.0	3.0	16.1	5.2	7.0
Greece	9.3	11.1	21.4	22.7	4.3	31.0	3.0	15.2	4.0	27.3	7.8	11.3
Hungary	7.9	9.8	15.8	17.8	4.2	16.7	1.4	13.8	2.2	20.9	4.6	7.9
Iceland	5.4	6.6	13.1	14.2	2.6	19.1	2.0	9.8	2.7	13.7	4.4	5.9
Ireland	3.6	4.7	12.0	10.8	1.8	19.2	2.1	10.9	2.5	17.0	4.2	5.5
Italy	8.3	10.6	22.1	24.0	5.2	30.8	2.6	15.2	3.6	23.4	6.8	9.6
Japan	8.0	10.1	13.4	13.8	1.7	26.3	1.3	15.5	2.0	18.8	5.0	7.9
Korea	3.9	4.1	16.1	10.5	1.0	30.7	0.7	7.2	1.0	8.8	2.4	3.6
Liechtenstein	8.6	10.6	12.8	9.0	4.1	15.5	1.4	11.5	1.7	13.1	4.6	5.9
Luxembourg	9.8	12.2	21.0	23.4	5.3	22.4	3.3	15.7	3.9	19.5	7.0	9.8
Mexico	7.1	8.2	19.2	20.0	3.0	30.7	2.1	6.2	3.1	17.9	3.3	4.1
Netherlands	1.8	2.0	4.0	3.6	0.6	11.3	0.5	3.5	0.7	6.6	1.2	1.4
New Zealand	3.9	5.3	9.2	10.2	1.8	25.4	1.7	8.8	2.0	13.4	3.4	4.5
Norway	6.5	7.7	16.2	19.2	3.4	5.3	2.1	11.3	3.7	19.9	5.6	7.2
Poland	9.1	11.9	19.3	21.0	4.5	13.2	1.1	11.7	1.4	18.7	4.8	7.6
Portugal	7.9	9.7	21.2	19.1	3.3	18.4	1.6	12.4	2.0	20.0	4.7	7.0
Russian Federation	8.7	10.5	17.3	18.0	4.9	20.4	3.0	15.7	4.7	24.4	6.7	9.0
Spain	5.9	7.0	17.6	17.4	3.3	20.4	2.6	13.4	3.7	23.0	5.8	8.2
Sweden	5.7	7.0	13.4	16.1	3.0	24.0	2.8	13.1	4.2	15.5	6.3	8.2
Switzerland	7.4	9.2	9.7	10.8	3.4	22.5	2.1	11.3	2.4	13.1	4.6	6.5
United Kingdom	4.6	5.7	10.0	11.6	1.7	14.4	2.3	10.6	3.1	16.0	3.9	5.5
United States	3.6	4.5	8.5	9.8	1.5	11.0	0.6	4.3	0.9	7.8	1.4	1.8
International Average	6.5	8.0	14.3	15.3	3.1	21.1	2.0	11.5	2.7	17.2	4.8	6.7

MC=multiple choice; WR=written response.

Table 26 examines the percent of skipped responses by domain and item type for Ireland only, for 2000 and 2009, by gender, language, and SES. The purpose of this analysis is to examine whether there have been differential changes in the rates of skipped responses by certain subgroups of the population, and also whether the assessment domain matters or not. In the case of reading, it can be seen that the percentage of skipped written responses has more than doubled for both males and females, and now exceeds 13% for males. The rates of skipped responses by students with a first language other than the language of the test have also increased and particularly so for written response items – from 6% in 2000 to 23% in 2009. In the case of mathematics, the gender differences are not that marked, but as with reading, in 2009 boys skipped more items than girls. Also, the rate of skipped responses for written mathematics questions was higher in 2009 for other language speakers compared with those speaking the language of the assessment (21% versus 17%), though is much less marked than for reading.

Table 26. Percent of missing responses, all domains, by gender, language spoken and SES, 2000 and 2009 – Irish students

Characteristic	RMC2000	RMC2009	RWR2000	RWR2009
Reading				
Gender				
Males	4.4	2.5	5.8	13.5
Females	2.7	1.7	3.4	8.2
Language				
Test language	3.5	1.8	4.6	10.1
Other language	4.6	6.5	5.7	22.6
SES (r)	-0.156	-0.091	-0.162	-0.176
Characteristic	MMC2000	MMC2009	MWR2000	MWR2009
Mathematics				
Gender				
Males	12.1	3.0	11.8	18.6
Females	11.8	2.0	9.7	15.4
Language				
Test language	11.9	2.2	10.7	16.7
Other language	7.9	4.9	10.9	20.7
SES (r)	-0.115	-0.11	-0.181	-0.212
Characteristic	SMC2000	SMC2009	SWR2000	SWR2009
Science				
Gender				
Males	2.0	5.1	15.2	6.3
Females	1.6	3.2	11.5	4.7
Language				
Test language	1.8	3.8	13.2	5.0
Other language	3.0	7.4	19.1	9.4
SES (r)	-0.082	-0.134	-0.148	-0.129

In columns 2 to 5, the first letter denotes the domain (R, M or S), the second two denote the item type (MC, WR), and the last four digits indicate the year of the assessment (2000, 2009).

Turning now to science, there is little variation by gender or language group and again, it is unexpected that the rate of skipped responses on written science items in 2009 has decreased across all groups examined, compared to 2000. Correlations between SES and the rate of skipped responses for these six item types indicate weak negative relationships that does not vary appreciably across item type, domain or cycle.

Skipping responses could be indicative of disengagement with the test, an inability to answer some questions, and/or changes in strategies for taking the test (e.g. a more cautious student would skip a question if they were not sure of the answer, while a less cautious student would be more inclined to guess an answer). This leads to the question as to whether rates of skipped responses are related to achievement scores. If students are skipping responses due to caution rather than disengagement or an inability to respond, there may be little or no correlation with achievement. As an illustration, Figures 6 and 7 show the relationship between achievement in reading and percent of skipped responses to reading questions for Irish students. The correlation between percent of skipped questions and achievement is negative: -.59 in 2000 and -.55 in 2000. However there is a good deal of clustering at the lower end of the missing distribution. Together with increases in skipped responses overall, Figures 6 and 7 support the hypothesis that there has been an increase in student disengagement with the test and/or their ability to answer questions, which appears to vary by domain.

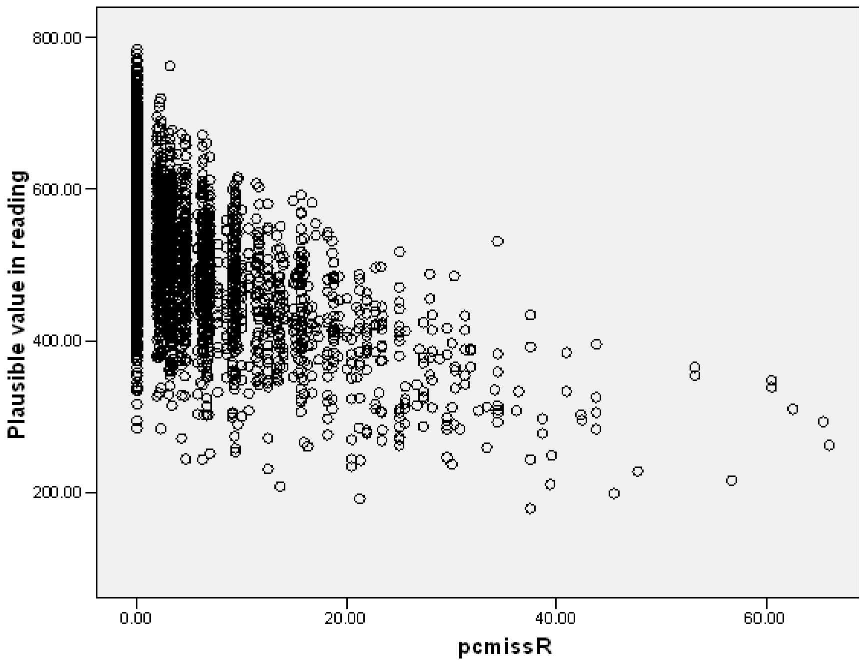


Figure 6. Scatterplot of percent missing reading questions and reading achievement, 2000

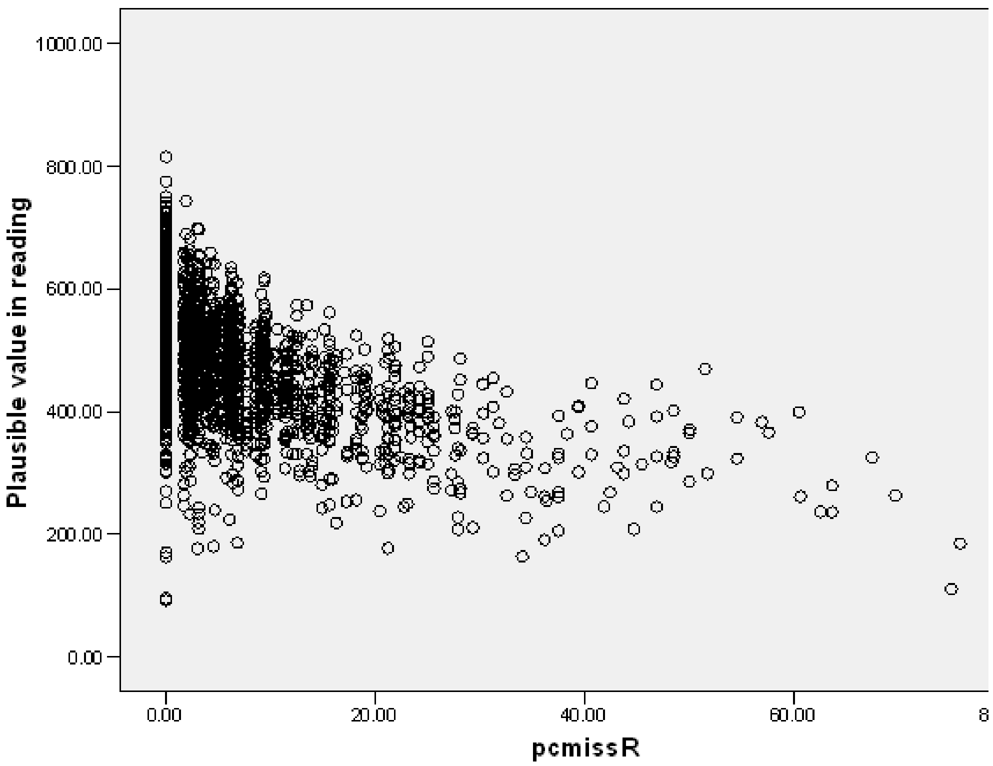


Figure 7. Scatterplot of percent missing reading questions and reading achievement, 2009

Finally, the revisions to the PISA assessment framework and administration of new reading items may also have an effect not only on how students respond to these new items but may also interact with students' responses to the link items and this is also an issue that merits further investigation. The Statistics Canada analysis (LaRoche & Cartwright, 2010) indicates that the reading questions new to 2009 are systematically more difficult for students in Ireland relative to the link items. This issue will be investigated further by comparing item percent correct for Ireland and other participating countries when the data become available.

Section 4. Linking, Scaling and Related Issues

A key feature of PISA is the establishment of links from cycle to cycle so that performance across different cycles can be tracked and the progress of education systems monitored. However, it has to be acknowledged that PISA Consortium is still coming to terms with the complexity of this task and with the objective of establishing more stable links, with some evidence that the approach to linking adopted by PISA may operate more favourably for some countries than for others (see Gebhardt & Adams, 2007).

What one seeks in surveys such as PISA are for trends to be both precise (accurate) and stable, yet sensitive to change if differences are real – such as performance differences arising from changes in demography or positive outcomes for initiatives designed to improve performance. What one doesn't want is a large number of mean scores that are significantly different from earlier assessments in the absence of adequate explanation. In PISA 2009, 13 countries show significant increases in reading achievement, and six, including Ireland, show declines.[16] We are unsure why there are so many significant differences – 19 out of a pool of 36 countries for which comparative data are available for 2000 and 2009. As was noted by Mazzeo and von Davier (2008) in their review of PISA trends for the OECD, the number of countries with significant differences in PISA is large compared with other large-scale assessments[17] (e.g., NAEP). Mazzeo and von Davier interpret this to indicate that PISA is less stable than some other large-scale assessments.[18] However, their comparison is between a *national* assessment and an international one. In the 2007 TIMSS study (an IEA study of mathematics and science), grade 8 (second year) students were surveyed in 56 countries/regions (not including Ireland) (Mullis et al., 2008). The results for 2007 were compared with those for 2003 and 1999. Comparisons were possible for 42 countries/regions between 2003 and 2007, and for 28 countries/regions between 1999, 2003 and 2007. Focusing on comparisons for mathematics, first between 2003 and 2007, significant changes were found in 23 countries/regions. Changes exceeded 40 scale points in Malaysia, Tunisia and Ghana. A comparison of 1999 and 2007 indicates significant changes in achievement in 11 countries/regions. These exceeded 30 score points in Lithuania, Jordan and Québec. Hence, the rate of significant changes in achievement appears broadly similar in TIMSS to what is observed in PISA (with about half of countries showing significant changes during those

[16] In mathematics, there were 12 significant decreases in achievement since 2003 (including Ireland, with the second largest difference), and 9 significant increases. In science, there were 8 decreases and 9 increases.

[17] This is a commentary on the relatively small standard errors associated with PISA trend differences, arising from link error. In fact, Ireland's decline of 11 points on PISA reading literacy between 2000 and 2003 was reported by the OECD (2004) as being statistically significant at the .05 level. However, the OECD (2007) reported the 11 point difference as being significant only at the less conventional .10 level, after the ACER consortium had made a correction to the link error factor.

[18] A corollary of this is that a large decline, such as that experienced in Ireland for reading literacy in PISA 2009, could be reversed in the next PISA cycle.

two periods) and suggests that the estimation of trends is an evolving methodology in international studies, and not confined to PISA.

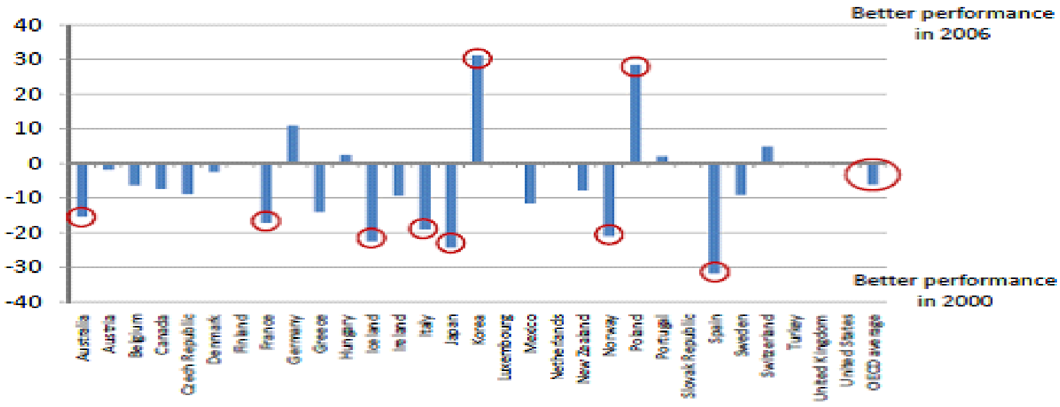
Up to now, we know that at least two countries (Japan and Mexico) suffered declines between 2000 and 2003 (following through to 2006) on reading due, at least in part, to the link items selected. Ireland also suffered a decline, albeit smaller than in these countries and significant only at the unconventional and less conservative .90 significance level) between 2000 and 2006, and, again, the decline may have been due to link error. If it the case that the declines in Japan and Mexico arise from the choice of link items, then the changes are an artefact of the particular link items selected. And there may be countries in which increases in achievement are also related to the particular link items selected. However, the Statistics Canada report (LaRoche & Cartwright, 2010) suggests that the selection of link items has not disadvantaged Ireland in 2009 at least, although that report provides some indication that the new reading items administered in 2009 show a systematic bias that may result in an underestimate of Irish students' performance. Note that in scaling the reading items to form the 2009 scale, the PISA Consortium assumes that the new and old test items have equivalent measurement properties.

Past Changes in Performance: Reading Literacy

Figure 8 shows that performance on PISA reading literacy changed significantly in about a third of OECD countries between 2000 and 2006. Large significant declines were observed in Spain (-32 points), Japan (-24), Iceland (-22), France (-17) and Australia (-15).[19] In Ireland, the decline was 9 points (a difference that was not statistically significant at the conventional .05 level). The OECD average declined just 6 points in the same time period (in the case of boys, the difference is -10, and in the case of girls, just -3). While eight countries showed significant declines in achievement (countries below the axis that are marked in red), just two (Korea and Poland) showed significant increases (countries above the axis, marked in red).

[19] The decline in performance in Northern Ireland noted in Section 1 has not been highlighted by the OECD, as analyses are based on the United Kingdom, rather than its constituent countries.

Figure 8: Changes in Reading Literacy from 2000 to 2006 (Adams, 2009)



It should be noted that the performance of students in Ireland in 2006 actually represented a non-significant increase in achievement since 2003, when the mean score was 11 points lower than in 2000. Hence, much of the decline in achievement between 2000 and 2006 in Ireland appears to have occurred between 2000 and 2003. This suggests that efforts to understand the decline in reading performance in Ireland between 2000 and 2009 need to take into account the decline of 11 points between 2000 and 2003, and the decline of 9 points between 2000 and 2006, and whether or not the two declines are linked to one another.

Ireland's decline of 31 points between PISA 2000 and 2009 is unprecedented in terms of size, with the exception of the decline in Spain between 2000 and 2006 (-32) (Table 27).

Table 27. Countries with Large Changes in Performance in PISA Reading (2000-2006)

Country	Reading 2003-2000			Reading 2006-2003			Reading 2006-2000		
	Diff	M	F	Diff	M	F	Diff	M	F
France	-9	***-14***	-5	+3	+0	+7	**-17**	**-21**	**-14**
Iceland	**-15**	**-25**	-7	-7	-3	**-13**	**-22**	**-28**	**-19**
Ireland	***-11***	-12	-11	+2	-1	+4	-9	-13	-8
Italy	***-12***	-14	***-13***	-7	-8	-6	**-19**	**-22**	**-18**
Japan	**-24**	**-21**	**-28**	0	-4	+4	**-24**	**-25**	**-24**
Korea	+9	+7	***+14***	**+22**	**+13**	**+27**	**+31**	**+20**	**+41**
Mexico	**-22**	**-23**	**-22**	11	4	**+17**	***-11***	**-18**	-5
Norway	-6	-10	-4	**-15**	***-13***	**-17**	**-21**	**-24**	**-21**
Poland	**+17**	**+15**	**+19**	***+11***	+11	***+11***	**+29**	**+26**	**+30**
Spain	**-12**	**-21**	-6	**-20**	**-17**	**-21**	**-32**	**-38**	**-27**

Bold: difference is statistically significant at .05 level; bold + italics, statistically significant at .10 level. M = Male; F = Female.

Overall Percent Correct on all Reading Items, 2000 and 2009

Table 28 shows, for *all* reading items administered in Ireland in 2000 and 2009, the percent correct overall and for males and females. There has been a decline in percent correct in the region of 5.3 percentage points, and this holds for both males and females. This is consistent with the finding reported earlier, i.e. that in 2000, 2.6% of students skipped more than 25% of the questions, and in 2009 this was double, at 5.2%.

Table 28. Percentage Correct for All Reading Items, Overall and by Gender, 2000 and 2009: Ireland

Group	2000	2009
All	65.7	60.2
Males	63.0	57.4
Females	68.2	63.1

The Role of Link Items

A number of investigators, both within the PISA Consortium and working as independent researchers, have attempted to explain why large differences in achievement occur between PISA cycles for selected countries. Gebhardt and Adams (2007) attributed the differences in performance in Japan between 2000 and 2003 (-24 points) to two main factors:

- Link item choice (i.e., the performance of students on the particular items that the Consortium selected to establish links between 2000 and 2003), and the fact that students in Japan performed less well than might be expected on those particular items (i.e., differential item functioning pointed to several significant item by country interactions).
- A coding problem – a large number of values for socioeconomic status (parental occupation) were missing in 2003 compared with 2000, and their absence impacted on the conditioning aspect of scaling.

The decline in performance in Mexico was also attributed in part by Gebhardt and Adams (2007) to "an unfortunate choice of link items" (from Mexico's point of view).

The PISA Consortium has made some efforts to deal with linking error in PISA (see Appendix B). In practical terms, this means that the standard error of the difference includes a correction for linking error. Interestingly, in the international report on PISA 2003 (OECD, 2004), the decline of 11 points in Ireland was reported to be statistically significant at the conventional .05 level. However, when the same comparison was re-issued in the international report on PISA 2006 (OECD, 2007), the difference was reported to be significant only at the unconventional (less conservative) .10 level.

Performance on the Reading Link Items in Ireland

Given the importance of link items in establishing trends, can it be said that Irish students suffered unduly from the particular set of link items used in PISA?

Table 29 shows item percent correct scores of students in Ireland over four PISA cycles on the link items in reading. It shows a steady increase in the rate of missingness (skipped items) and not-reached items between PISA 2000 and PISA 2009, and also shows a decline in item percent correct on the link items, from 64.4% correct in 2000 to 60.0% correct in 2009. Of course, the Table does not explain whether the decline in performance on the link items is due to a change in demographics, sampling fluctuation, or some effects specific to the link items themselves in terms of, for example, their positions within the test booklets.

Table 29. Average Missing, Not Reached and Item Percent Correct for Link Items, All Cycles – Reading

Statistic (%)	2000	2003	2006	2009
Missing	4.9	5.7	6.2	7.9
Not reached	1.9	0.6	0.8	2.4
Correct (all)	64.4	64.4	63.1	60.0
Correct (females)	66.7	66.5	65.7	62.5
Correct (males)	62.7	62.3	60.5	57.5

Table 30 shows item percent correct scores for multiple choice and written response link items in 2000 and 2009. The table shows a drop in performance of 4.2% on the multiple-choice link items, and a drop of 4.8% on the constructed response (written) items. Thus, the decline in the performance of students in Ireland on the link items was marginally greater for constructed response than for multiple-choice items. For all items, there has been a decline of around 4.4%, which is the slightly larger for males (5.2%) than for females (4.2%).

Table 30. Item Percentage Correct for Reading Link Items by Item Type, Overall and by Gender, 2000 and 2009

Item type/Group	2000	2009
Multiple choice		
All	69.6	65.4
Males	67.4	63.2
Females	71.9	67.5
Written response		
All	61.5	56.7
Males	59.8	53.9
Females	63.4	59.4

There is a fairly consistent pattern of differential item functioning (DIF) in Ireland across the four cycles (Table 31). [20]

Table 31. Differential Item Functioning (item-by-country) in PISA 2000, 2003, 2006 and 2009

Item	DIF 09	DIF 06	DIF 03	DIF 00
R219Q01T		ETE	ETE	ETE
R219Q02		ETE	ETE	
R102Q04A				
R102Q05		ETE	ETE	
R102Q07	ETE			
R220Q04	HTE	HTE	HTE	
R220Q06	HTE	HTE	HTE	HTE
R227Q01	HTE	HTE	HTE	HTE
R227Q02			HTE	
R227Q03		HTE		
R111Q02B				ETE
R055Q03	ETE			
R104Q05		HTE	HTE	

HTE=harder than expected
ETE=easier than expected

Hence, while there is differential functioning of the link items in 2009 (three harder than expected, two easier than expected), there isn't strong evidence to support the view that differential item functioning strongly impacted on the change in performance among students in Ireland in PISA 2009. However, it could be the case that the selected link items had a negative impact on reading achievement in Ireland in 2003 (with five items designed 'harder than expected' and three 'easier than expected'). To the best our knowledge, other variables (e.g., demography, between-school variance) were more stable in 2003 than in 2009 (both compared to 2000), increasing the likelihood that Statistics Canada's conclusion that new reading items rather than the choice of link items beyond 2003, and the method used to establish the link back to 2000 (see below), may have contributed to the observed decline in Ireland's performance in reading in 2009.

[20] This DIF violates the assumption of the one-parameter Rasch model which assumes population invariance of item parameters. Note, however, that differential item functioning is accepted only if significant at the .95 level. Given the requirements of the Rasch model, it can be argued that a more conservative .90 level might be appropriate.

Use of Link Items in Scaling

In PISA 2009, the following steps were used to establish links for all three assessment domains (see Appendix B):

Step 1. Item parameter (item difficulty) estimates for reading and mathematics were obtained from the PISA 2009 calibration sample;

Step 2. The above item parameters estimates were transformed through the addition of constant, so that the mean of the item parameter estimates for the link items was the same in 2009 as it was in 2006;

Step 3. The 2009 student abilities were estimated with item parameters anchored at their 2009 values;

Step 4. The above estimated student abilities were transformed with the shift estimated in Step 2.

Table 32 indicates that there has been a uniform decline in the reading scores of students at the lower and upper ends of the ability distribution. Again, the findings by Statistics Canada (LaRoche & Cartwright, 2010) indicate that, given that the new items were more difficult for Irish students, steps 3 and 4 above, which assume an equivalent test across new and link items, are problematic in the case of Ireland.

Table 32. Reading Scores of Students in Ireland at the 10th and 90th percentiles – 2000-2009

	10th Percentile	90th Percentile
PISA 2000	401	641
PISA 2003	401	622
PISA 2006	395	633
PISA 2009	373	611

Other Aspects of the PISA Design that May Contribute to Unstable Trends

There are a number of other reasons why the design of the PISA assessment may impact favourably or unfavourably on a country's performance. These factors may or may not have affected the performance of students in Ireland. According to Mazzeo & von Davier (2008), these include:

- Use of mixed domain booklets (doing a set of mathematics items first may have a different effect on reading performance than doing a set of science items, or a different set of reading items first). The use of mixed-domain booklets is viewed as being problematic for trend estimates because if in a previous administration of the survey the item parameters are based on a cluster preceded by a different domain this can affect the estimation of the parameters and current scaling models are incapable of modelling these complex interaction effects. For this reason, Mazzeo and von Davier prefer a design where booklets focus on just one domain.

- Position of a cluster within an assessment booklet (some reading clusters in PISA 2000 always preceded math/science clusters). Related to this, it should also be noted that in 2003, reading items were not selected in intact clusters and some items within clusters were deleted before making these into two intact 'trend' clusters from 2003 onwards. In the case of mathematics and science, intact clusters were selected. This is likely to further compound the stability of trend estimates in the case of reading.
- Use of a one-parameter (Rasch) model rather than a two-parameter model. The review by von Davier & Mazzeo (2008) suggested that the latter model was more successful at accounting for item-by-country interactions. The Rasch model assumes population invariance which as we have seen does not hold in many cases. See for example Table 31 which indicates that a substantial subset of link items for reading show differential item functioning in the case of Ireland which is indicative of a violation of the assumption of the Rasch model. This is not necessarily a problem if differential item functioning occurs at random (although this would affect the precision of estimates) but would be problematic of there was a systematic bias in differential item functioning that could serve to advantage or disadvantage particular countries.
- Passage effects. With the small number of passages used to establish trends, there is the risk of a passage effect in one or more countries (i.e., students perform particularly well or particularly poorly on one of the passages, and there is an insufficient number of such passages to compensate elsewhere). It should be noted that, for linking purposes, PISA 2009 included eight reading passages or units which is substantially lower than the number of passages for both mathematics and science link items (25 and 17, respectively). This indicates that trends in reading are more likely to be prone to passage effects.

Perhaps the biggest difficulty with PISA reading trends is the relatively small number of trend items. Just 28 trend items based on eight passages or stimulus texts were included in the 2003 and 2006 PISA reading assessments (Table 33), and the same number was used for estimating trend in 2009. In practice, this number may be even smaller, since some link items are dropped for all countries (2 in the case of PISA 2009 - R219Q01T and R219Q01E), and others for individual countries, or for students within a country taking a particular test booklet. Whereas there were actually 41 items in PISA 2009 that were drawn from PISA 2000, in fact, the same 28 as used in 2003 and 2006 were considered for establishing trends.

Table 33. Numbers of Items Used to Link Performance in PISA (2000-2009)

Comparison	Reading	Maths	Science
2000-2003	28	20	25
2003-2006	28	48	22
2006-2009	28 (26)	35(32)	53 (49)

Numbers in brackets indicate numbers of items following international item deletions.

The situation with link items is different for mathematics and science. First, there are more link items available in those domains. Second, performance on the link items tend to be more independent of one another (and less dependent on a specific text), potentially reducing the types of effects that may arise for reading literacy.

Trends in Mathematics and Science

As noted above, more link items were used in PISA mathematics than in PISA reading. This reduced the possibility of differences in performance across cycles being attributed to the particular set of link items selected. Table 34 shows that the difference in performance on the mathematics link items was 2.3% lower in 2009 than in 2006. This is equivalent to about one-tenth of a standard deviation (10 scale score points).

Table 34. Mathematics Percent Correct in Ireland 2003 and 2009

	2003 % Correct			2009 % Correct		
	Overall	Boys	Girls	Overall	Boys	Girls
All items (N=84)	49.5 (0.15)	51.4 (0.20)	47.5 (0.22)	N/A	N/A	N/A
Non-link items (N=49)	50.8 (0.20)	52.6 (0.27)	49.0(0.28)	N/A	N/A	N/A
Link items (N=35)	47.6 (0.24)	49.5 (0.32)	45.5 (0.35)	45.4	46.6	44.20
Link items (N=32)	49.0 (0.26)	51.0 (0.37)	47.0 (0.34)	46.7 (0.25)	48. (0.35)	45.4 (0.34)

It seems unlikely that the particular set of link items chosen for the PISA 2009 mathematics assessment contributed much to the differences observed between 2003 and 2009, as the same set of link items was used in 2006 (linking back to 2003), when no significant change in performance was observed. The greater decline in performance among boys in 2009 (3.0% on the 32 link items) may be related to a stronger representation of boys in the eight very low performing schools in the selected sample as well as more boys in small vocational schools relative to earlier cycles.

What of performance in science? Table 35 shows a drop of just one half of a percentage point between 2006 (when science was last a major domain) and 2009.

Table 35. Science Item Percent Correct Scores in Ireland 2006 and 2009

	2006 % Correct			2009 % Correct		
	Overall	Boys	Girls	Overall	Boys	Girls
All items (N=108)	52.6 (0.128)	52.5 (0.18)	52.7 (0.17)	N/A	N/A	N/A
Non-link items (N=55)	50.2 (0.177)	50.5 (0.25)	50.0 (0.24)	N/A	N/A	N/A
Link items (N=53)	55.1 (0.184)	54.6 (0.26)	55.6 (0.25)	54.5	54.3	54.7
Link items (N=49)	55.7 (0.19)	55.2 (0.26)	56.2 (0.26)	55.3 (0.20)	55.1 (0.29)	55.4 (0.28)

Why might performance in reading and mathematics decline, while performance in science remained constant? There are several possible reasons:

- Greater engagement by boys with the science items (fewer missed items) than with reading.
- An actual increase in achievement, masked by the relatively poorer performance of students in eight very low performing schools.
- The cohort that completed PISA science in 2006 was to first to study the revised Junior Certificate science syllabus. However, it may have been too early to look for effects of the revised syllabus in 2006, and such effects could well have begun to appear at this time.
- The 2009 cohort may also have benefitted from teaching in science at primary level associated with the revised Primary School Curriculum (1999) in that subject.
- A larger number of link items than in previous cycles, underlining the Consortium's improved understanding of the need to increase the number of such items.

Other Scaling Issues

Conditioning. It should be noted that although the samples for PISA 2000 and 2009 are similar along the dimensions examined in Section 2, changes in demographics such as the number of students who do not speak English or Irish at home (0.9% in 2000 and 3.6% in 2009) will undoubtedly have an effect on achievement. What we don't know is how this and other demographic changes might affect scores when conditioning is used. It is also possible that changes in two or more variables interact during conditioning to distort scores.

It has been shown in secondary analyses conducted by ACER that the estimation of trends can be affected at the stage at which students' scores are conditioned on background characteristics in terms of unwanted changes in demographics, particularly rates of missing data. In Japan, increases in rates of missing responses to questions on SES affected trend comparisons for that country between 2000 and 2003. In Ireland's case, the ERC has checked rates of missing data in 2000 and 2009 on a small set of core variables (gender, grade, language spoken, age, family structure, and measures of SES), but the fluctuations are miniscule and unlikely to explain the observed changes in achievement.

International vs National Item Parameters. [21] The use of international vs. national item parameters (for link items) may impact on scaling. PISA estimates trends with a common set of link items and uses the same set of international items parameters to estimate change across all participating countries before applying the (common) adjustment to each individual country. An alternative is to estimate a unique set of parameters for each country, with the scales made comparable by setting a common mean for the set of items used in all countries. A consequence of this is that trend estimates for each country would then be less comparable across countries, but, within countries, they would be based on more appropriate item estimates (since a separate estimate of linking error is computed for each country).

Gebhardt and Adams (2007) report an analysis of PISA 2000 and 2003 reading and science results in 27 OECD countries (including Ireland) and the Russian Federation, in which alternative approaches to equating, including use of national item parameters, were tried out. They found that, in the case of Ireland, the particular scaling method used (international item parameters, national item parameters with and without adjustment of mean scores for changes in socioeconomic status, gender, socioeconomic status, and language spoken at home) did not result in any change in trend (difference) scores in either domain.

As noted in Appendix D (also see Figure 9), the average international and national item parameters for Ireland in PISA 2009 are not statistically significantly different (the international parameters have an average of 0.00 while that for Ireland is -0.03; the averages for the link items internationally and for Ireland are -0.175 and -0.104, respectively), suggesting that, in the absence of other adjustments to the scaling methodology, use of national rather than international item parameters may not lead to a change in Irish mean scores for 2009.

On the other hand, in Gebhardt and Adams (2007), a significantly higher score was reported for Canada on reading literacy using national rather than international item parameters, suggesting that, for some countries, use of national parameters can provide a more accurate description of achievement, at least at national level. Urbach (2009) also reported improved results for Australia in mathematics between 2003 and 2006 by using national item parameters. Whereas the published mean score for mathematics was greater in 2003 than in 2006, use of national item parameters led to no significant difference. In the case of reading in Australia, the internationally-published data showed a

[21] Item parameters in a Rasch Model (such as that used in PISA) are indices of item difficulty. Appendix D of this report gives the international and national item parameters for reading for Ireland in 2009.

decline in reading among high achievers between 2000 and 2003. Urbach found that the decline in fact related to the 15% of lowest-achieving students between those years. Similarly, whereas published data showed a decline in Australian reading scores between 2003 and 2006 among high achieves, Urbach found a constant decline across all proficiency levels.

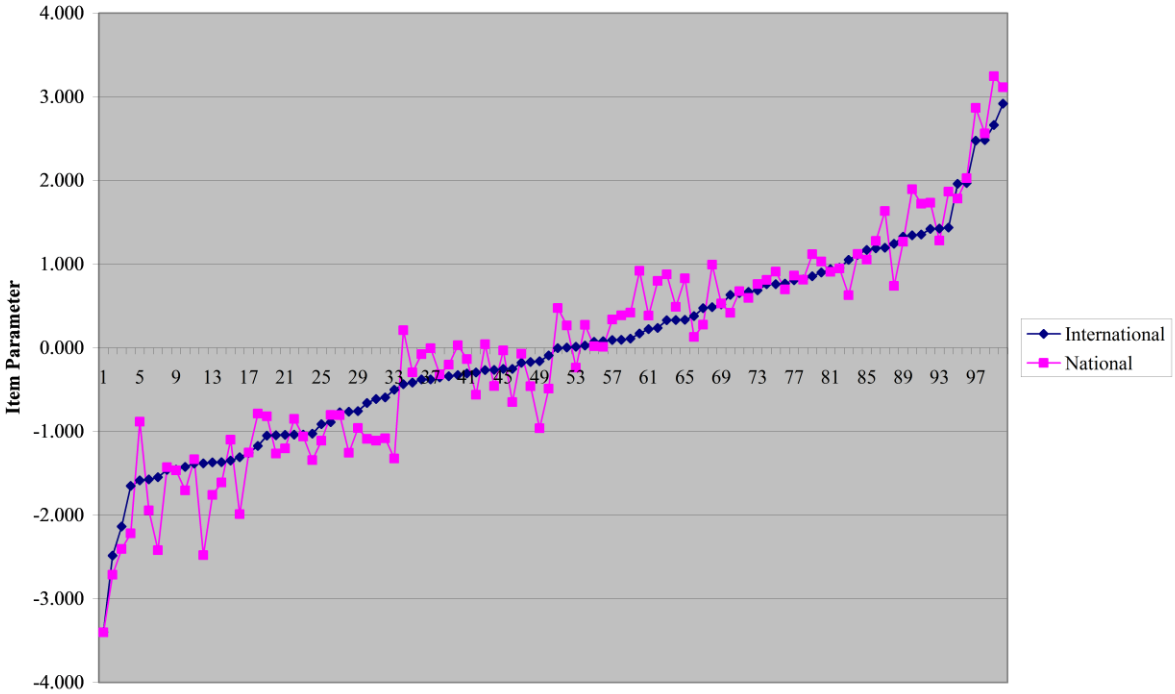


Figure 9. National (Irish) and International Item Parameters, All Reading Items, PISA 2009

Finally, it should also be noted that, in 2009, the PISA consortium decided to report performance only in terms of the PISA 2000 reading literacy scale. Had a new reading scale been devised for 2009 (with separate reporting for trends), it is possible that the 9-point decline incurred by Ireland in the period 2000 to 2006 would not have been reflected in the 2009 mean score (that is, performance on link items in 2000-2003 would have had less of an impact on overall performance).

It was noted earlier in this report that Ireland participated in ICCS, a survey run concurrent to PISA, although the same schools were not selected to participate in both surveys. While the measurement instruments and subject domains are not the same, and nor are the sample designs and target populations, it is still of interest to compare general country rankings for the 33 countries that participated in both assessments. This comparison is shown in Figure 10 (see also Table F5). Results vary widely, ranging from a +81-point difference in performance (ICCS higher; Denmark), to a difference of -15 (PISA higher; Netherlands). On average, these 33 countries had a higher mean score on ICCS compared with PISA. Ireland ranks 8 out of 33 countries on ICCS, and 13th on PISA reading. Note that the country-level correlation between these two scores is .83.

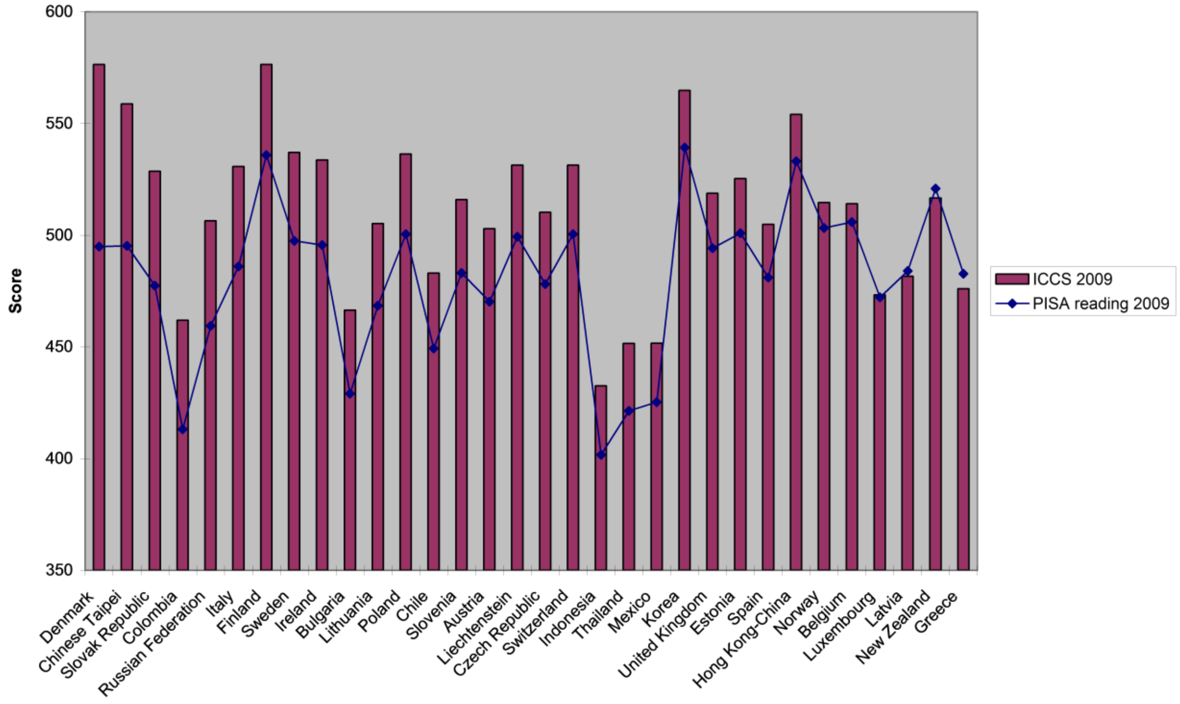


Figure 10. Country averages for PISA 2009 and ICCS 2009 (N=33)

Linking PISA Trends to Performance on National Assessments and Curriculum Change at

The 2004, the National Assessment of English reading was administered to a representative sample of pupils in Fifth class in that year. It was reported that the average performance of students in 2004 was not significantly different from that of students in a representative sample of Fifth class pupils in 1998 (Eivers et al., 2005). Students who were in Transition year and Fifth year in PISA 2009 (almost 40% of the achieved PISA 2009 sample) would have been eligible to participate in the 2004 National Assessment (Table 36). Hence, the year cohort of which 40% of the 2009 PISA sample belonged in 2004 did not experience a decline in achievement since the previous assessment in 1998. If the lower scores achieved by students in Transition Year and Fifth Year in PISA 2009 reflect a real decline in reading achievement, it would seem that the decline occurred between 2004 and 2009.

Table 36. Eligibility of the PISA 2009 Cohort for the 2004 National Assessments of Reading and Mathematics at Primary Level

Class Level in PISA 2009 (Representation in PISA)	Class Level in 2004	Corresponding National Assessment
Second Year (2.3%)	Third Class	None
Third Year (60.2%)	Fourth Class	National Assessment of Mathematics 2004
Transition Year (23.5%)	Fifth Class	National Assessment of English Reading 2004
Fifth Year (13.9%)	Fifth Class	National Assessment of English Reading 2004

A new Primary School English Curriculum (PSEC) was introduced in 1999, with implementation beginning in the 2001-02 school year. However, we would caution against using performance on the National Assessment of English Reading in 2004 or indeed PISA 2009 to make strong inferences about the impact of the revised primary school English curriculum on achievement. Pupils in Transition Year/Fifth year in PISA would have been entering Third class when the implementation of the 1999 PSEC began, and hence would have acquired basic reading skills under the pre-1999 curriculum and related textbooks. It is, of course, possible that some practices introduced by senior class teachers and textbook publishers in response to the implementation of the revised English curriculum could have had an effect that would only show up in achievement scores at a later stage (e.g., the relative emphasis given to novels vs. expository/information texts (NCCA, 2005). Similarly, a lack of implementation in some classrooms in respect of practices promoted by the curriculum (e.g., purposeful use of ICTs to teach reading/writing – NCCA, 2005; an emphasis on higher-order questioning and use of reading material as a stimulus for discussion – DES Inspectorate, 2005; use of formative assessment; Eivers et al., 2010) could also have a delayed negative impact. It could also be the case that skills and strategies introduced at primary level may not receive adequate follow-up at post-primary level if teachers of English do not have an adequate understanding of how to develop basic reading comprehension skills. Nevertheless, it seems unlikely that curricular changes (whether implemented or not) could account for a decline in reading literacy of the magnitude observed in Ireland between 2006 and 2009.

National Assessments of Mathematics Achievement were administered to representative samples of pupils in Fourth class in 1999 and 2004, with no change in average performance between the two years (Shiel et al., 2006). The Fourth class cohort in 2004 would have included Third year students eligible to participate in PISA 2009 (Table 36). As with English, a revised primary school mathematics curriculum (PSMC) was introduced in 1999, with implementation beginning in junior classes in the 2002-03 school year, and in senior classes in the following school year. Again, we can say that the performance of pupils in Fourth class was stable between 1999 and 2004, and that, if a real decline in standards occurred, it must have taken place after 2004. We can only make very limited inferences about the effects of the 1999 mathematics curriculum, since it was in place for a very short time when the 2004 national assessment was administered (perhaps as little as one year for the senior classes). However, it has to be acknowledged that there are some longstanding issues around the teaching of mathematics at primary level that pre-dated implementation of the 1999 PSMC, and that still need be addressed. These include using problems to teach basic skills and concepts and develop mathematical thinking, teaching problem-solving strategies, and enhancing generalisation and algebraic thinking (Eivers et al., 2010). It is also unclear to what extent post-primary mathematics builds on the skills that are currently taught at primary level (see Smyth, McCoy & Darmody, 2004), though one might expect greater compatibility between PSMC and Project Maths, in terms of teaching methodology.

With respect to science, it is relevant to note that a revised curriculum for primary schools in that subject was also introduced in 1999, with implementation beginning in 2003-04. Varley, Murphy and Veale (2008), in their review of implementation of this curriculum, noted some challenges to implementation, including an over-emphasis on life-sciences, and relatively few opportunities for pupils to apply science knowledge. Nevertheless, it may be that the introduction of science as a subject, and the engagement of pupils in scientific experiments from an early age may have combined with the implementation of the revised Junior Certificate Science curriculum at post-primary level to mitigate the effects of changes in demography and sampling that might otherwise have lowered performance in science in PISA 2009.

It is also worth noting that the correlations between PISA achievement outcomes and performance on the Junior Certificate have decreased over cycles (Table 37) – for example the decrease in shared variation between reading (English), mathematics and science from 2003 to 2009 is 7%, 17%, and 9%, respectively. Also, assuming that the application of marking schemes to the Junior Certificate English, mathematics and science examinations has remained relatively stable in the period 2000-2009, the performance of students on these examinations has remained extremely stable, which is not consistent with the changes in achievement on PISA reading and mathematics.

Table 37. Relationships between PISA and the Junior Certificate, PISA 2003, 2006 and 2009

Cycle		Reading (English)	Mathematics	Science
	r	0.673	0.754	0.674
	SE	0.012	0.008	0.014
2003	r2	0.453	0.569	0.454
	r	0.636	0.687	0.702
	SE	0.012	0.011	0.010
2006	r2	0.404	0.472	0.493
	r	0.621	0.629	0.602
	SE	0.013	0.012	0.016
2009	r2	0.386	0.396	0.362
change in r2 2003-2009		0.067	0.173	0.092
change in r2 2006-2009		0.019	0.076	0.130

Changes in Achievement in Schools that Participated in Multiple PISA Assessments: The Stability of Trends

As noted in the report prepared by Statistics Canada, 39 schools participated in both PISA 2000 and PISA 2009. Of these, five had a drop in average reading scores exceeding 100 scale points. Two of these five schools are also among the eight low-performing schoolsdiscussed previously.

Table F6 compares the five schools with the remainder of the sample. Results indicate that:

- Students in ‘large achievement drop’ schools had close to twice as many missing responses
- 19% of students in ‘large drop’ schools skipped more than 25% of the questions presented to them
- The ‘large drop’ schools have a mean score in all three domains that is half a standard deviation or more below that of the other schools
- ‘Large drop’ schools have a mean SES score (based on parental occupation) that is 0.4 standard deviations below the other schools, but only a slightly higher rate of Junior Certificate fee waiver
- ‘Large drop’ schools have fewer girls
- ‘Large drop’ schools have similar proportions of ‘other’ language speakers compared with other schools
- Participation rates are similar across ‘large drop’ and other schools
- ‘Large drop’ schools tended to have female test administrators employed as careers guidance or learning support teachers and in the community/comprehensive sector to a greater degree than other schools.

In summary, it could be the case that 'large drop' schools have experienced changes in the SES and language composition of their students in recent times.

Due to the concerns about the link between PISA 2000 and PISA 2009 it was decided to follow up these observations with some comparisons of schools that participated in both PISA 2006 and 2009. There are 36 such schools.

Table F7 shows the proportions of females, that speak another language, average grade level (where grade 9 = third year), average SES and school average performance on PISA reading, science, mathematics, and on Junior Certificate English, mathematics and science for these schools in 2006 and 2009. Overall, the table indicates relative stability in terms of averages.

Also, as shown in Table 38, the correlations between PISA achievement outcomes and Junior Certificate Examination outcomes are, broadly speaking, similar. However, correlations *across* the two cycles (e.g., Junior Certificate English 2006-2009) indicate that the performance on the Junior Certificate is more stable across the two years relative to PISA (Table 39).

Figures F1 to F10 show the frequencies for the differences in the ten characteristics examined in Table F7. Results indicate that:

- Generally, the proportion of females is quite stable but ranges from an increase of 23% to a decrease of 35%
- The percentage of other language speakers is also relatively stable but ranges from a decrease of 10% to an increase of 11%
- Again, average grade level is quite stable but changes range from a decrease of 0.38 of a grade to an increase of 0.45 of a grade
- SES is less stable and appears to have increased by about one-eighth of a standard deviation
- Changes in achievement on average are generally small (down 7 points on PISA reading, up 5 points on PISA mathematics, and less than one point change on PISA science; JCE results show average increases ranging from about a quarter to half a standard deviation). However, Figures F5 to F10 show a range of differences on these six achievement measures – particularly for the PISA results, which fluctuate more than the JCE results. It should nonetheless be noted that changes in achievement (*across cycles)* on both PISA and the JCE are positively and significantly correlated with one another: .526 for reading (English), .498 for mathematics and .423 for science.
- *Within* a cycle, average achievement (taking PISA reading for 2000 as an illustrative outcome) is significantly related to school SES (.759), proportion with another language

(-.614), and not significantly related to proportion female (.274) or average grade level (.251). However, *changes* in achievement are not related to changes in these four characteristics, again taking the change in reading achievement as an illustrative outcome; the correlations are as follows: change in proportion female = .132, change in proportion with another language = .089, change in SES = .202, and change in average grade = .309. Similarly, changes in achievement on the Junior Certificate are unrelated to changes on these four characteristics with the exception of performance on Junior Certificate English and average grade level (.399).

Table 38. Within-cycle correlations of achievement outcomes for schools common to PISA 2006 and 2009 (N=36)

Correlation within cycle	2006	2009
PISA reading-JCE English	0.862	0.907
PISA mathematics-JCE mathematics	0.832	0.797
PISA science-JCE science	0.791	0.657

Table 39. Cross-cycle correlations of achievement outcomes for schools common to PISA 2006 and 2009 (N=36)

Correlation across cycle	r
PISA reading	0.672
PISA mathematics	0.716
PISA science	0.588
JCE English	0.869
JCE mathematics	0.879
JCE science	0.834

Section 5. Conclusions

This report is an initial attempt by the staff of the ERC to explain or at least provide some relevant context for the results of PISA 2009.

In Section 1, the scale of the changes in achievement was outlined. It was noted that the size of the decline in reading (31 points, or almost one third of an international standard deviation since 2000, and 20 points since 2003) was considerable, while the decline in mathematics (16 points, or one-sixth of an international standard deviation between 2003 and 2009) was smaller. It was further noted that performance in reading and mathematics declined across the spectrum of achievement, with fewer students scoring at the highest proficiency levels, and more scoring at the lowest levels. In the case of science, no difference in achievement was observed between 2006 and 2009. The changes in achievement had the effect of lowering Ireland's ranking in reading literacy from 5th across all participating countries in 2000 to 21st in 2009. Ireland's ranking in mathematics fell from 20th in 2000 to 32nd in 2009. Ireland's ranking in science (20th) did not change between 2006 and 2009. Changes in rankings were also compared for OECD countries. In 2000, 27 OECD countries for whom results can be reported participated, while in 2009 33 OECD countries did so. OECD rankings for Ireland for reading in 2000 and 2009, respectively, were 5th and 17th; for mathematics for 2003 and 2009 respectively they were 17th and 25th, and in science, they were 14th and 13th, respectively, for 2006 and 2009. Of particular concern in 2009 is that Ireland's mean score in mathematics is significantly below the corresponding OECD country average. It was also noted that although the gender difference in reading in Ireland is similar to the OECD average of about 40 points, there has been a somewhat larger increase in the gender gap in Ireland relative to OECD average (an increase of 10 points, compared with 7). The gender gap increase is particularly marked for two of the reading subscales – retrieving information and non-continuous texts.

The second section compared the samples for 2000 and 2009 and the extent to which each was representative of its respective population. Overall, the samples are remarkably similar. However, differences do emerge. Some of these, such as the four-fold increase in the proportion of students who speak a language other than English at home relate to demographic changes. Others, such as the significant reduction in the proportion of students in Fifth year (with a corresponding significant increase in Transition year), and a small decline in the rate of early school leaving are linked to organisational changes in the educational system. Still others, such as the presence of eight very low performing schools in the 2009 sample (i.e., schools with mean reading achievement scores that are more than one student standard deviation or 100 score points below the mean score) are likely to have arisen due to random sampling fluctuation and not due to the introduction, in 2009, of a sampling stratifier for school SES (Junior Certificate fee waiver). It was also noted that (i) achievement in reading has declined more in boys than girls, (ii) the achievement gap by immigrant and language status has increased in Ireland (yet remains below the corresponding OECD averages), and (iii) fluctuations in performance depend on the domain and grade level considered, e.g. the decline in reading is largest in Fifth Year, while in mathematics, it is greatest in Fourth (Transition) year. Average SES by year level has remained quite stable across cycles, although there is evidence for a small (non-significant) increase in average SES among Transition Year students and a small (non-significant) decrease in average SES among Fifth Years.

Schools that participated in PISA 2000 and 2009 are very similar in terms of their performance on the Junior Certificate English Examination in those years, and are also comparable with respect to performance in English to non-sampled schools in both years. Furthermore, the overall performance of students on the Junior Certificate Examination in the participating schools (both samples) remains stable and not different from overall performance in schools in general over the period 2001 to 2008. At a more general level, if we assume that the marking schemes for English, mathematics and science Junior Certificate examinations have been consistently applied during the period 2000-2009, Junior Certificate results in these three subjects have remained extremely stable, which stands in contrast to the PISA results for reading and mathematics. Concerning the originally sampled PISA 2009 schools, their average performance in English is equivalent to the achieved sample (which included two replacement schools and in which 16 schools were unable or unwilling to participate). Participating schools are also similar with respect to a range of measures used to establish DEIS scores, including the overall DEIS index, though it is apparent that, in 2009, fewer participating schools were in the two highest (most advantaged) DEIS deciles, and more in the two lowest (most disadvantaged) deciles, relative to participating schools in 2000.

There has been an increase in between-school variation in achievement and in socioeconomic status between 2000 and 2009, accompanied by an increase in overall variation in both achievement and SES in both years. The extent to which this may have been due to the introduction of a measure of socioeconomic status as an implicit stratification variable in 2009 and whether it could have had an impact on achievement is a matter that requires further investigation. (See the Statistics Canada report which confirms that this has not had any measurable impact and hence this observed change is likely to be due to changes in the PISA populations since 2000.)

As already noted, eight schools in PISA 2009 had an average reading score that is more than one student standard deviation below the overall average. No such schools emerged in the PISA 2000 sample. These schools also had markedly lower performance on the Junior Certificate (i.e. two standard deviations below the other schools in the sample). It is estimated that, if these eight schools had not participated in 2009, the Irish average score would have been up to seven scale score points higher. There do not appear to have been any particular difficulty with test administration in these schools. It was also observed that these eight schools tended also to have low scores in mathematics and science, although school averages varied somewhat more across the three domains when compared to the corresponding means of the eight lowest-scoring schools in reading in 2000. When the eight very low scoring schools were removed from the sample, between-school variance in reading achievement reduces to 2000 levels. However, the removal of the eight schools only has a marginal effect on between-school variation in socioeconomic status. The most likely reason for the appearance of these schools in the PISA 2009 sample is random fluctuations in sampling rather than any systematic bias or problem (again see the Statistics Canada report on this issue).

Section 3 looked at administration of the PISA test in Ireland in 2009 and subsequent scoring of the data. As far as is known, all aspects of survey administration and data processing were correctly adhered to and, at a more general level, the PISA Consortium has indicated that Ireland has met all technical standards in 2009. There were, however, some differences in administration between 2000 and 2009 (previous cycles used external test administrators while in PISA 2009 in about three-quarters of schools the assessment was administered by a member of staff). The greater involvement of teachers as test administrators was associated with a small and statistically insignificantly lower level of achievement, but regression analysis of missing responses indicates that this difference is accounted for by the fact that schools with external test administrators had, on average, higher SES than those with school staff administering PISA. Interestingly, missing responses were also significantly higher in males and students with another first language. Comparing this regression model with the equivalent one for 2000 indicates an increase in missing responses for males and other language speakers in 2009 (after adjusting for student and school SES).

Compared with the 31 other countries common to PISA 2000 and 2009, Ireland shows the second highest increase of skipped responses which supports the hypothesis of a comparatively large decline in engagement and/or proficiency. However, comparisons of non-response in 2000 and 2009 for Ireland only indicate that while overall there has been an increase in skipped questions, this is most pronounced in reading and mathematics, while in science, the instance of skipping items has actually decreased. This matter was examined further by comparing patterns of skipped responses by domain and item type for countries that can be compared across 2000 and 2009. Similar to the international averages, Ireland has seen a decline in missing responses for written response science items. In the case of mathematics and reading, the increases in skipped responses are more marked for written response items than for multiple choice ones, and in the case of mathematics, there has been a marked decrease in the rate of skipped multiple choice questions. Comparisons of specific subgroups in Ireland for skipped responses by domain and item type indicate a particularly marked increase in skipped reading written response items by students speaking another language, while the rate of skipped written reading response items has more than doubled for both males and females. As for mathematics, the overall pattern of increases in skipped written items and decreases in skipped multiple choice items tends to hold across gender and language groups but again, other language speakers had the highest rate of skipped written mathematics items in 2009. Finally, the overall pattern of skipped responses in the case of science tends also to hold across gender and language groups. Correlations between SES and the six item type/domain combinations indicate that these are generally weak and negative and have not changed between 2000 and 2009, i.e. there has been no change in the relative advantage of higher SES students on this measure. It would be advisable, nonetheless, for further national reporting to examine more closely why and how the changes in missingness by item type and domain are not consistent. For example, it could be the case in science that the written response items selected in 2009 represented a comparatively easy item set relative to those used in 2000.

A further possibility that was investigated was whether changes in skipping patterns in 2000 and 2009 reflect changes in engagement/proficiency, or whether they may be indicative of changes in test-taking strategies. However, there are strong negative correlations between percent of missing responses and achievement in both 2000 and 2009 which provides further corroborating evidence for a decline in engagement and/or proficiency.

The introduction of the electronic reading assessment (ERA; an add-on given to some students after they have completed the paper-and-pen test) appears to have had no impact on student motivation in relation to achievement on the paper-and-pen test in reading since the mean paper-based scores of students doing and not doing the electronic assessment are identical.

However, changes made to the assessment framework for reading and the administration of new reading items may have impacted on student performance. The analysis by Statistics Canada supports the assertion that the revision of the framework and consequent introduction of new reading items has resulted in a systematic underestimation of student performance in Ireland, since student performance on individual items was better than predicted on the basis of their scaled reading scores in 65% of cases and this was particularly pronounced in the case of the new, as opposed to the link, items.

Section 4 looked at the PISA test design, scaling, and related matters. It was noted that significant changes in achievement have been identified in about half of countries for reading (2000-2009) and mathematics (2003-2009) and that this is similar to other international studies of achievement trends such as TIMSS; hence, this area can be seen as a complex and evolving one. The analyses in this section confirmed that students in Ireland in 2009 performed somewhat less well on link items in reading than in previous cycles, and that new reading items were more difficult than the link items, raising questions about the suitability of the new items. Parallel with lower percent-correct scores, higher levels of missingness on the link items were noted, particularly on constructed-response items. There is evidence that students skipped more of the test items generally in 2009 compared with 2000 (something that could arise from lower achievement, greater disengagement from the test, or some combination of the two). In 2009, twice as many students (5.2%) skipped more than 25% of questions presented to them than in 2000 (2.6%).

The question of whether the link items used in reading disadvantaged students in Ireland in a significant way was explored. It was argued that performance on link items may have contributed to a 9-point decline in reading performance between 2000 and 2006, though the Statistics Canada analysis indicates that some of the observed decline in achievement may be attributable to the new reading items introduced in 2009 and the scaling of these items, which assumes that they assess competencies equivalent to the link items,

The conditioning of student scores (a key component of the PISA scaling process) is unlikely to have been affected by differential rates of missingness in PISA 2000 and 2009 on key student background variables such as gender and SES. According to a statement received from the PISA Consortium on August 13th, 2010 "Preliminary analyses were conducted on PISA 2009 plausible values where unconditional and conditional trends (controlling for gender, grade, Mother/Father job, test language, immigration status and age) were compared between PISA 2006 and 2009 cycles. In case of IRL there were no significant differences found between conditional and unconditional trends for three domains". If this is correct, it can be concluded that conditioning during scaling did not affect trends for Ireland, though there would be value in having this checked independently.

It was also suggested that these technical issues should be considered in the broader context of curricular changes and/or implementation and some examples of these were provided in the concluding part of Section 4. In this section, it was also noted that the strength of the relationship between PISA reading, mathematics and science and the respective Junior Certificate examination results has weakened across cycles.

Finally, Section 4 included a comparison of the characteristics of schools (n=36) that participated in both PISA 2006 and 2009. Cross-cycle correlations on the school average PISA scores for the three domains are in the region of .60 to .70, and the equivalent correlations for the Junior Certificate examination results are stronger, at around .80 to .90, indicating greater stability on the Junior Certificate relative to PISA. Furthermore, average PISA scores for these 36 schools are very similar for 2006 and 2009, although it was noted that average scores changed from a decrease of 205 points to an increase of 88 points. It was also noted that changes across cycles on both PISA scores and Junior

Certificate results are unrelated to changes in the percentage of females, other language speakers, average grade level, and SES (the exception being a weak positive relationship between percent of other language speakers and performance on Junior Certificate English).

The view of the PISA Consortium at this time is that the declines in performance in reading (since 2000) and mathematics (since 2003) are real, and should be interpreted in the context of demographic and structural changes. Our analyses show that, relative to earlier PISA cycles, these include an increase in the proportion of students who speak a language other than English or Irish, a decrease in proportion of early school leavers, and decrease in the proportion of 15-year olds enrolled in a Leaving Certificate course. Also, we found that 3.5% of students with an identified special educational need participated in 2009, but the equivalent percentage is not available for 2000.While our analyses show that the samples of schools participating in earlier PISA cycles (especially PISA 2000) are equivalent to the 2009 sample in respect of achievement in Junior Certificate English, and overall Junior Certificate performance, and in terms of average scores on the DEIS index, there are some potentially important differences. These include a small increase in average SES, and the participation of a number of very low achieving schools in 2009. Paradoxically, the effect of removing these schools is to reduce between-school variance in reading achievement but not between-school variance in SES.

Finally, attention was drawn to potential problems with the scaling of data for PISA, when according to Statistics Canada, student performance on individual items in reading was better than predicted on the basis of their scaled reading scores in 65% of cases and this was particularly pronounced in the case of the new, as opposed to the link, items. Clearly, there is a need to examine the scaling of PISA 2009 data in more detail in the context of performance in other participating countries, to ascertain if the effects of scaling the data for Ireland were similar to the effects for other participating countries.

REFERENCES

Adams, R. (2009). Are they an outrageous fortune or a sea of troubles. Keynote address at PISA Research Conference, Kiel, Germany, September 14-16. Accessed at http://www.pisaresconf09.org/user_uploads/files/trends/Keynote_Adams.pdf

Bradshaw, J., Sturman, L., Vappula, H., Ager, R., & Wheater, R. (2007). Student achievement in Northern Ireland. *Results in science, mathematics and reading among 15-year olds from the OECD PISA 2006 study*. Slough: National Foundation for Educational Research.

Department of Education and Science. Inspectorate. (2005). *An evaluation of curriculum implementation in primary schools.* Dublin: Stationery Office.

Eivers, E., Shiel, G., Perkins, R., & Cosgrove, J. (2005). *The 2004 National Assessment of English Reading.* Dublin: Educational Research Centre. http://www.erc.ie/documents/naer04_report_full.pdf

Eivers, E.,. Close, S., Shiel, G., Millar, D., Clerkin, A., Gilleece, L., & Kiniry, J. (in press). *The 2009 National Assessments of Mathematics and English Reading.* Dublin: Stationery Office.

Gebhardt, E., & Adams, R. (2007). The influence of equating methodology on reported trends in PISA. *Journal of Applied Measurement, 8(3)*, 305-322. Access at: https://mypisa.acer.edu.au/images/mypisadoc/gebhardt.pdf

Joint Oireachtas Commitee on Education and Skills. (2010). *Staying in education: A new way forward. School and out-of-school factors protecting against early school leaving*. Dublin: Author.

LaRoche, S., & Cartwright, F. (2010). *Independent review of the 2009 PISA results for Ireland: Report prepared for the Educational Research Centre at the request of the Department of Education and Skills.* Dublin: Department of Education and Skills. (Available at www.erc.ie)

Mazzeo, J., & von Davier, M. (2008). *Review of the programme for international student assessment test design: Recommendations for fostering stability in assessment results.* Paper presented to the 26th meeting of the PISA Governing Board, The Hague, November 2-5. Accessed at https://edsurveys.rti.org/PISA/documents/MazzeoPISA_Test_DesignReview_6_1_09.pdf

Monsieur, C., & Berezner, A. (2007). The computation of equating errors in international surveys of education. Journal of Applied Measurement, 8(3), 323-335. Accessed at https://mypisa.acer.edu.au/images/mypisadoc/monseur_berezner.pdf

Mullis, I.V.S., Martin, M.O., & Foy, P. with Olson, J.F, Preuschoff, C., Erberber, E., Arora, A., & Galia, J. (2008). *TIMSS 2007 International Mathematics Report: Findings from IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grade.* Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

National Council for Curriculum and Assessment. (NCCA). (2005). Primary curriculum review. Phase 1. Dublin: Author. http://www.ncca.ie/uploadedfiles/Publications/PrimaryCurriculumReview.pdf

OECD. (2005). *PISA 2003 technical report*. Paris: Author. Access at: http://www.oecd.org/dataoecd/49/60/35188570.pdf

Shiel, G., Surgenor, P., Close, S., & Millar, D. (2006). The 2004 National Assessment of Mathematics Achievement. Dublin: Educational Research Centre. http://www.erc.ie/documents/nama04_main_report_complete.pdf

Skolverket. Swedish National Agency for Education. (2009). What influences educational achievement in Swedish schools. Stockholm: Author. http://www.skolverket.se/sb/d/190

Smyth, E., McCoy, S., & Darmody, M. (2004). Moving up: The experiences of first-year students in post-primary education. Dublin: Liffey Press.

Urbach, D. (2009). An investigation of Australian OECD PISA trend results. Paper presented at the PISA Research Confernece, Kiel, Germany, September 14-16, 2009. Access at: http://www.pisaresconf09.org/user_uploads/files/trends/room3/Urbach.pdf

Varley, J., Murphy, C., & Veale, O. (2008). *Science in primary schools. Phase 1. Final report.* Dublin: National Council for Curriculum and Assessment.

Appendix A: Comparison of the PISA 2000 and PISA 2009 Achieved Samples for Ireland

Table A1 gives a break down of the percentage of students in both the PISA 2000 and PISA 2009 samples by school type, school size and gender composition. There has been an increase in the percentage of males (from 37.92% to 43.42%), and a corresponding decrease in the percentage of females (from 51.0% to 44.0%), in large secondary schools between the 2000 and 2009 samples. This has been mirrored by a decrease in the percentage of males (from 16.3% to 12.5%), and an increase in the percentage of females (from 11.0% to 13.6%), in large community/comprehensive schools between the two samples. Also of note is the small increase in the percentage of males (from 2.9% to 4.2%), and decrease of females (from 3.1% to 1.7%), in small vocational schools between the 2000 and 2009 sample.

Table A1. Percentages of Students in PISA 2000 and PISA 2009 Samples by Stratification Variables

		School size															
		Small				Medium				Large				*Total*			
School type	*% girls*	2000		2009		2000		2009		2000		2009		*2000*		*2009*	
		Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
Secondary	0-25%	1.22	0.00	1.37	0.00	9.20	0.00	7.50	0.51	27.51	0.29	28.95	0.12	37.93	0.29	37.82	0.63
	26-50%	0.97	0.61	0.84	1.04	2.21	1.22	4.41	3.24	4.94	3.05	12.30	9.71	8.12	4.88	17.55	13.99
	51-75%	0.00	0.00	0.00	0.00	4.53	5.65	1.93	2.51	5.47	5.27	1.64	1.94	10.00	10.92	3.57	4.45
	76-100%	0.00	1.48	0.00	1.27	0.00	8.69	0.00	11.97	0.00	42.36	0.53	32.22	0.00	52.53	0.53	45.46
	Total	*2.19*	*2.09*	*2.21*	*2.31*	*15.94*	*15.56*	*13.84*	*18.23*	*37.92*	*50.97*	*43.42*	*43.99*	*56.05*	*68.62*	*59.47*	*64.53*
Vocational	0-25%	0.00	0.00	1.99	0.23	2.80	0.34	0.00	0.00	0.00	0.00	1.70	0.36	2.80	0.34	3.69	0.59
	26-50%	2.94	3.10	1.37	0.27	5.75	3.12	6.78	5.23	10.47	7.34	9.31	6.74	19.16	13.56	17.46	12.24
	51-75%	0.00	0.00	0.68	0.61	1.43	1.31	2.12	3.34	2.65	3.14	2.14	2.58	4.08	4.45	4.94	6.53
	76-100%	0.00	0.00	0.14	0.55	0.00	0.97	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.97	0.14	0.55
	Total	*2.94*	*3.10*	*4.18*	*1.66*	*9.98*	*5.74*	*8.9*	*8.57*	*13.12*	*10.48*	*13.15*	*9.68*	*26.04*	*19.32*	*26.23*	*19.91*
Comm/ Comp	0-25%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.45	0.00	0.00	0.00	1.45	0.00	0.00	0.00
	26-50%	0.00	0.00	0.00	0.00	1.96	1.10	1.53	1.28	11.99	7.82	9.34	8.50	13.95	8.92	10.87	9.78
	51-75%	0.00	0.00	0.22	0.40	0.00	0.00	0.59	0.74	2.83	3.19	3.12	3.83	2.83	3.19	3.93	4.97
	76-100%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.26	0.00	0.00	0.00	1.26
	Total	0.00	0.00	*0.22*	*0.40*	*1.96*	*1.10*	*2.12*	*2.02*	*16.27*	*11.01*	*12.46*	*13.59*	*18.23*	*12.11*	*14.8*	*16.01*

Table A2 presents a comparison of the PISA 2000 sample to the population of 15-year-olds in schools in 2000. Overall, the PISA 2000 sample appears to reflect the 2000 population of 15-year-olds quite well. However, there a couple of small discrepancies between the sample and population, for example the overall percentage of males in vocational schools seems to be slightly under-represented in the sample (26.0% compared to 27.9%), while the percentage of females in these schools is slightly over-represented (19.3% compared to 18.2%) - the reverse is the case in community/comprehensive schools. Also, there is a slightly higher percentage of males in the medium secondary schools category in the PISA sample compared to the population (15.9% compared to 14.7%).

Table A2. Percentages of Students in the School Population in 2000 and the PISA 2000 Sample by Stratification Variables

		School size															
		Small				Medium				Large				*Total*			
School type	*% girls*	Sample		Population		Sample		Population		Sample		Population		Sample		Population	
		Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
Secondary	0-25%	1.22	0.00	1.45	0.00	9.20	0.00	10.48	0.10	27.51	0.29	28.44	0.07	37.93	0.29	40.37	0.18
	26-50%	0.97	0.61	0.70	0.51	2.21	1.22	2.27	1.83	4.94	3.05	5.07	3.65	8.12	4.88	8.04	5.99
	51-75%	0.00	0.00	0.19	0.26	4.53	5.65	1.96	2.83	5.47	5.27	4.62	6.17	10.00	10.92	6.76	9.27
	76-100%	0.00	1.48	0.03	1.21	0.00	8.69	0.01	8.88	0.00	42.36	0.00	42.34	0.00	52.53	0.04	52.43
	Total	*2.19*	*2.09*	*2.36*	*1.98*	*15.94*	*15.56*	*14.73*	*13.65*	*37.92*	*50.97*	*38.12*	*52.23*	*56.05*	*68.62*	*55.21*	*67.86*
Vocational	0-25%	0.00	0.00	1.56	0.31	2.80	0.34	2.50	0.54	0.00	0.00	1.44	0.17	2.80	0.34	5.51	1.03
	26-50%	2.94	3.10	1.79	1.14	5.75	3.12	5.97	3.89	10.47	7.34	11.92	8.36	19.16	13.56	19.67	13.40
	51-75%	0.00	0.00	0.31	0.48	1.43	1.31	1.42	1.88	2.65	3.14	1.01	1.22	4.08	4.45	2.75	3.58
	76-100%	0.00	0.00	0.01	0.07	0.00	0.97	0.00	0.16	0.00	0.00	0.00	0.00	0.00	0.97	0.01	0.24
	Total	*2.94*	*3.10*	*3.67*	*2.01*	*9.98*	*5.74*	*9.89*	*6.48*	*13.12*	*10.48*	*14.38*	*9.75*	*26.04*	*19.32*	*27.94*	*18.24*
Comm/ Comp	0-25%	0.00	0.00	0.00	0.00	0.00	0.00	0.18	0.00	1.45	0.00	0.99	0.02	1.45	0.00	1.17	0.02
	26-50%	0.00	0.00	0.06	0.06	1.96	1.10	0.88	0.70	11.99	7.82	11.75	9.27	13.95	8.92	12.69	10.03
	51-75%	0.00	0.00	0.00	0.00	0.00	0.00	0.85	1.04	2.83	3.19	2.13	2.63	2.83	3.19	2.99	3.67
	76-100%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.17	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.17
	Total	*0.00*	*0.00*	*0.06*	*0.06*	*1.96*	*1.10*	*1.92*	*1.91*	*16.27*	*11.01*	*14.87*	*11.92*	*18.23*	*12.11*	*16.85*	*13.90*

Table A3 presents a comparison of the PISA 2009 sample to the population of 15-year-olds in schools in 2009. In general, the PISA 2009 sample is a good reflection of the population of 15-year-olds in schools. However, males are somewhat over-represented in large secondary schools (43.4% compared to 37.4% in the population), and under-represented in community/comprehensive schools overall (14.9% compared to 18.0% in the population).

Table A3. Percentages of Students in the School Population in 2009 and the PISA 2009 Sample by Stratification Variables

		School size															
		Small				Medium				Large				*Total*			
School type	*% girls*	Sample		Population		Sample		Population		Sample		Population		Sample		Population	
		Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
Secondary	0-25%	1.37	0.00	1.74	0.00	7.50	0.51	9.26	0.11	28.95	0.12	27.24	0.31	37.82	0.63	38.25	0.43
	26-50%	0.84	1.04	0.73	0.53	4.41	3.24	3.67	2.89	12.30	9.71	7.72	6.14	17.55	13.99	12.12	9.57
	51-75%	0.00	0.00	0.18	0.30	1.93	2.51	1.55	2.01	1.64	1.94	2.42	3.15	3.57	4.45	4.15	5.46
	76-100%	0.00	1.27	0.01	0.72	0.00	11.97	0.00	13.51	0.53	32.22	0.00	34.62	0.53	45.46	0.01	48.85
	Total	*2.21*	*2.31*	*2.66*	*1.56*	*13.84*	*18.23*	*14.49*	*18.53*	*43.42*	*43.99*	*37.38*	*44.22*	*59.47*	*64.53*	*54.53*	*64.31*
Vocational	0-25%	1.99	0.23	1.27	0.21	0.00	0.00	0.79	0.19	1.70	0.36	1.82	0.45	3.69	0.59	3.88	0.85
	26-50%	1.37	0.27	2.34	1.56	6.78	5.23	6.68	4.53	9.31	6.74	10.48	7.66	17.46	12.24	19.49	13.74
	51-75%	0.68	0.61	0.74	0.96	2.12	3.34	1.68	2.31	2.14	2.58	1.61	1.82	4.94	6.53	4.04	5.09
	76-100%	0.14	0.55	0.03	0.25	0.00	0.00	0.00	0.18	0.00	0.00	0.00	0.00	0.14	0.55	0.03	0.43
	Total	*4.18*	*1.66*	*4.38*	*2.98*	*8.90*	*8.57*	*9.15*	*7.21*	*13.15*	*9.68*	*13.91*	*9.93*	*26.23*	*19.91*	*27.44*	*20.12*
Comm/ Comp	0-25%	0.00	0.00	0.07	0.02	0.00	0.00	0.31	0.08	0.00	0.00	1.29	0.15	0.00	0.00	1.67	0.25
	26-50%	0.00	0.00	0.05	0.02	1.53	1.28	2.34	1.75	9.34	8.50	10.48	8.71	10.87	9.78	12.87	10.48
	51-75%	0.22	0.40	0.14	0.19	0.59	0.74	0.39	0.48	3.12	3.83	2.97	3.68	3.93	4.97	3.50	4.36
	76-100%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.26	0.00	0.49	0.00	1.26	0.00	0.49
	Total	*0.22*	*0.40*	*0.26*	*0.24*	*2.12*	*2.02*	*3.03*	*2.31*	*12.46*	*13.59*	*14.74*	*13.03*	*14.80*	*16.01*	*18.04*	*15.57*

Table A4 gives the average scores on the various DEIS indices for schools in PISA 2000, PISA 2009, ICCS and all schools (using 2005 data). The schools in the PISA 2000 and PISA 2009 samples are very similar in terms of the DEIS indices. However, schools in the PISA 2009 have, on average, a higher final index score indicating that they are slightly less disadvantaged than the PISA 2000 schools. All samples (PISA 2000, PISA 2009 and ICCS) appear to be, on average, less disadvantaged than the general population of schools.

Table A4. Average School-Level Scores on DEIS Indices, PISA 2000, 2009 and ICCS Samples (2005 Data)

	PISA 2000		PISA 2009		ICCS		All schools	
	Mean	(StDev)	Mean	(StDev)	Mean	(StDev)	Mean	(StDev)
JCE and JCSP revised medical card % average for 02, 03, 04	28.81	(13.89)	28.52	(15.10)	26.26	(14.78)	31.30	(16.37)
Average performance on JCE 02 and 03	64.69	(5.76)	65.05	(5.50)	65.12	(5.64)	64.34	(6.02)
JC average % retention for 95, 96 and 97 entry cohorts	94.96	(4.54)	94.90	(4.53)	94.82	(5.24)	94.14	(5.78)
Average LC retention for 95, 96 and 97 entry cohorts	78.43	(12.3)	78.65	(11.72)	78.36	(12.34)	77.15	(13.40)
Estimate of medical card possession among dropouts	5.04	(4.54)	5.10	(4.53)	5.18	(5.24)	5.86	(5.78)
Total poverty estimate based on revised medical card data	33.85	(17.23)	33.61	(18.56)	31.44	(18.65)	37.15	(20.47)
Final index based on revised medical card data	0.4547	(3.2)	0.5336	(3.23)	0.6148	(3.39)	0.0000	(3.66)
Final index (z score)	0.1304	(0.92)	0.153	(0.925)	0.1751	(0.9697)	0.0000	(1.05)

Table A5 presents the average scores on the various DEIS indices for schools in PISA 2000, PISA 2009, ICCS and all schools using data from 2008. The average scores for the PISA 2000 and 2009 samples are very similar; however, it seems that the PISA 2009 sample may be slightly less disadvantaged than the PISA 2000 sample. Again, both samples appear to be slightly less disadvantaged than the general population of schools.

Table A5. Average School-Level Scores on DEIS Indices, PISA 2000, 2009 and ICCS Samples (2008 Data)

	PISA 2000		PISA 2009		ICCS		All schools	
	Mean	(StDev)	Mean	(StDev)	Mean	(StDev)	Mean	(StDev)
JCE and JCSP revised medical card % average for 2008	29.93	(15.96)	27.77	(16.89)	25.15	(16.44)	30.45	(18.73)
Average performance on JCE 2008 (JC08_OPS_TOT)	65.95	(5.89)	66.69	(5.57)	66.59	(5.87)	65.87	(6.28)
Average performance on JCE 2008 (JC08_altOPS_TOT)	65.95	(5.96)	66.73	(5.54)	66.62	(5.91)	65.77	(6.51)
JC average % retention for 2001 entry cohort	96.17	(3.48)	96.44	(3.92)	95.99	(5.58)	95.39	(6.06)
Average LC retention for 2001 entry cohort	80.53	(11.49)	81.39	(11.21)	80.31	(2.34)	79.46	(13.36)

Appendix B: Draft Document on Scaling/Linking Provided by PISA Consortium – August 2010

Developing Common Scales for the Purposes of Trends

- The reporting scales that were developed for each of reading, mathematics and science in PISA 2000 were linear transformations of the natural logit metrics that result from the scaling as described above. The transformations were chosen so that the mean and standard deviation of the PISA 2000 scores was 500 and 100 respectively, for the 27 OECD countries that participated in PISA 2000 that had acceptable response rates (see Wu & Adams, 2002).

•

- For PISA 2003 the decision was made to report the reading and science scores on these previously developed scales. That is the reading and science reporting scales used for PISA 2000 and PISA 2006 are directly comparable. The value of 500, for example, has the same meaning as it did in PISA 2000 – that is, the mean score in 2000 of the sampled students in the 27 OECD countries that participated in PISA 2000.

- For mathematics this was not the case, however. Mathematics, as the major domain, was the subject of major development work for PISA 2003, and the PISA 2003 mathematics assessment was much more comprehensive than the PISA 2000 mathematics assessment – the PISA 2000 assessment covered just two (*space and shape*, and *change and relationships*) of the four areas that are covered in PISA 2003. Because of this broadening in the assessment it was deemed inappropriate to report the PISA 2003 mathematics scores on the same scale as the PISA 2000 mathematics scores. For mathematics the linear transformation of the logit metric was chosen such that the mean was 500 and standard deviation 100 for the 30 OECD countries that participated in PISA 2003.

- For PISA 2006 the decision was made to report the reading on these previously developed scales. That is, the reading reporting scales used for PISA 2000, PISA 2003 and PISA 2006 are directly comparable. Mathematics reporting scales are directly comparable for PISA 2003 and PISA 2006. For science a new scale was established in 2006. The metric for that scale was set so that the mean was 500 and standard deviation 100 for the 30 OECD countries that participated in PISA 2006.

- To permit a comparison of the PISA 2006 science results with the science results in previous data collections, a science link scale was prepared. The science link scale provides results for 2003 and 2006 using only those items that were common to the two PISA studies. These results were provided in a separate database.

- For PISA 2009, reporting on the reading, mathematics and science was again done on these previously developed scales. Thus the reading reporting scales used for PISA 2000, PISA 2003, PISA 2006 and PISA 2009 are directly comparable. Mathematics reporting scales are directly comparable for PISA 2003, PISA 2006 and PISA 2009. Science scales are directly comparable for PISA 2006 and PISA 2009.

Linking PISA 2006 and PISA 2009 for reading, mathematics and science

- The linking of PISA 2009 reading, mathematics and science to the existing scales was undertaken using standard common item equating methods.

- The steps involved in linking the PISA 2006 and PISA 2009 reading, mathematics and science scales were as follows:

    - *Step 1.* Item parameter estimates for reading and mathematics were obtained from the PISA 2009 calibration sample;

- *Step 2.* The above item parameters estimates where transformed through the addition of constant, so that the mean of the item parameter estimates for the common items was the same in 2009 as it was in 2006;

- *Step 3.* The 2009 student abilities where estimated with item parameters anchored at their 2009 values;

- *Step 4.* The above estimated student abilities were transformed with the shift estimated in step 2.

• Note that this is a much simpler procedure than that employed in linking the reading and science between PISA 2003 and PISA 2000. The simpler procedure could be used on this occasion because the test design was balanced for both PISA 2006 and 2009.

Uncertainty in the Link

• In each case the transformation that equates the 2009 data with previous data depends upon the change in difficulty of each of the individual link items and as a consequence the sample of link items that have been chosen will influence the choice of transformation. This means that if an alternative set of link items had been chosen the resulting transformation would be slightly different. The consequence is an uncertainty in the transformation due to the sampling of the link items, just as there is an uncertainty in values such as country means due to the use of a sample of students.

• The uncertainty that results from the link-item sampling is referred to as linking error and this error must be taken into account when making certain comparisons between the results from different PISA data collection. Just as with the error that is introduced through the process of sampling students, the exact magnitude of this linking error cannot be determined. We can, however, estimate the likely range of magnitudes for this error and take this error into account when interpreting PISA results. As with sampling errors, the likely range of magnitude for the errors is represented as a standard error.

• In PISA 2003 the link error was estimated as follows.

• Let $\hat{\delta}_i^{2000}$ be the estimated difficulty of link *i* in 2000 and let $\hat{\delta}_i^{2003}$ be the estimated difficulty of link *i* in 2003, where the mean of the two sets difficulty estimates for all of the link items for a domain is set at zero. We now define the value:

$$c_i = \hat{\delta}_i^{2003} - \hat{\delta}_i^{2000}.$$

• The value c_i is the amount by which item *i* deviates from the average of all link items in terms of the transformation that is required to align the two scales. If the link items are assumed to be a random sample of all possible link items and each of the items is counted equally then the link error can be estimated as follows:

$$error_{2000,2003} = \sqrt{\frac{1}{L}\sum c_i^2}.$$

• Where the summation is over the link items for the domain and *L* is the number of link items.

• Monseur and Berezner (2007) have shown that this approach to the link error estimation is inadequate in two regards. First, it ignores the fact that the items are sampled a units and therefore a cluster sample rather than a simple random sample of items should be assumed. Secondly, it ignores the fact that partial credit items have a greater influence on students' scores than dichotomously scored items. As such, items should be weighted by their maximum possible score when estimating the equating error.

To improve the estimation of the link error the following improved approach was used in PISA 2006 and has been used for PISA 2009. Suppose we have *L* link items in *K* units. Use *i* to index items in a unit and *j* to index units so that $\hat{\delta}_{ij}^{y}$ is the estimated difficulty of item *i* in unit *j* for year *y*,and let

$$c_{ij} = \hat{\delta}_{ij}^{2006} - \hat{\delta}_{ij}^{2003}$$

The size (total number of score points) of unit *j* is m_j so that:

$$\sum_{j=1}^{K} m_j = L \quad \text{and} \quad \overline{m} = \frac{1}{L}\sum_{j=1}^{K} m_j$$

Further let:

$$c_{\bullet j} = \frac{1}{m_j}\sum_{i=1}^{m_j} c_{ij} \quad \text{, and} \quad \overline{c} = \frac{1}{N}\sum_{i=1}^{K}\sum_{j=1}^{m_j} c_{ij}$$

- and then the link error, taking into account the clustering is as follows:

$$error_{2006,2003} = \sqrt{\frac{\sum_{j=1}^{K} m_j^2 (c_{\bullet j} - \overline{c})^2}{K\left(K-1\right)\overline{m}^2}}$$

- In PISA a common transformation has been estimated, from the link items, and this transformation is applied to all participating countries. It follows that any uncertainty that is introduced through the linking is common to all students and all countries. Thus, for example, suppose the *unknown* linking error (between PISA 2003 and PISA 2006) in reading resulted in an over-estimation of student scores by two points on the PISA 2003 scale. It follows that every student's score will be over-estimated by two score points. This over-estimation will have effects on certain, but not all, summary statistics computed from the PISA 2006 data. For example, consider the following:

  - Each country's mean will be over-estimated by an amount equal to the link error, in our example this is two score points;
  - the mean performance of any subgroup will be over-estimated by an amount equal to the link error, in our example this is two score points;
  - The standard deviation of student scores will not be effected because the over-estimation of each student by a common error does not change the standard deviation;
  - The difference between the mean scores of two countries in PISA 2006 will not be influenced because the over-estimation of each student by a common error will have distorted each country's mean by the same amount;
  - The difference between the mean scores of two groups (e.g. males and females) in PISA 2006 will not be influenced, because the over-estimation of each student by a common error will have distorted each group's mean by the same amount;
  - The difference between the performance of a group of students (e.g. a country) between PISA 2003 and PISA 2006 will be influenced because each student's score in PISA 2003 will be influenced by the error; and finally;
  - A change in the difference in performance between two groups from PISA 2003 to PISA 2006 will not be influenced. This is because neither of the components of this comparison,

which are differences in scores in 2006 and 2003 respectively, is influenced by a common error that is added to all student scores in PISA 2006.

- In general terms, the linking error need only be considered when comparisons are being made between results from different PISA data collections, and then usually only when group means are being compared.

- The most obvious example of a situation where there is a need to use linking error is in the comparison of the mean performance for a country between two PISA data collections. For example, let us consider a comparison between 2003 and 2006 of the performance of Canada in mathematics. The mean performance of Canada in 2003 was 532 with a standard error of 1.8, while in 2006 the mean was 527 with a standard error of 2.0. The standardised difference in the Canadian mean is -1.82, which is computed as follows: $-1.82 = (527 - 532) \Big/ \sqrt{2.0^2 + 1.8^2 + 1.4^2}$, and is not statistically significant.

Link Error

Link errors estimated using the methodology for the following eleven links; PISA mathematics scales 2003 to 2006, 2006 to 2009 and 2003 to 2009, PISA reading scales 2000 to 2003, 2000 to 2006, 2000 to 2009, 2003 to 2006, 2003 to 2009 and 2006 to 2009, PISA science scale 2006 to 2009 and science trend scale 2003 to 2006, are given in Table B1. Note that the value of 4.474 given for the PISA Reading scale 2000 to 2003 link is a little larger than the value of 3.744, as reported in OECD (2005). Similarly for the interim science scale the new estimate of 3.112 is a little larger than the previously reported value of 2.959. The differences in these values is due to the improved link error estimation method used for PISA 2006 and PISA 2009.

Table B1. Link Error Estimates

	Link Error on PISA Scale
PISA Reading scale 2000 to 2003	4.474
PISA Reading scale 2000 to 2006	4.976
PISA Reading scale 2000 to 2009	6.698
PISA Reading scale 2003 to 2006	5.307
PISA Reading scale 2003 to 2009	4.088
PISA Reading scale 2006 to 2009	4.069
PISA Mathematics scale 2003 to 2009	1.990
PISA Mathematics scale 2006 to 2009	1.333
PISA Mathematics scale 2003 to 2006	1.382
PISA Science scale 2006 to 2009	2.566
Interim Science scale 2000 to 2003	3.112
Science trend scale 2003 to 2006	4.963

Appendix C: Analysis of PISA Test Design for Reading[22]

PISA 2000 Test Design

The test design for PISA 2000 (Table C1) was not balanced in that it did not have each cluster in each position.

In fact 7 of the 9 clusters only appear in the first three positions – and also for these 7 clusters in the first position is during the first and second half hours of the test. In contrast, the last two clusters appear only during the second half of the booklets. Another notable aspect of this test design is that for the first seven booklets, students do 90 uninterrupted minutes of reading followed by a half hour block of mathematics or science.

Table C1. PISA 2000 Test Design

Booklet	P1	P2	P3	P4
1	R_1	R_2	R_4	M_1/M_2
2	R_2	R_3	R_5	S_1/S_2
3	R_3	R_4	R_6	M_3/M_4
4	R_4	R_5	R_7	S_3/S_4
5	R_5	R_6	R_1	M_2/M_3
6	R_6	R_7	R_2	S_2/S_3
7	R_7	R_1	R_3	R_8
8	M_4/M_2	S_1/S_3	R_8	R_9
9	S_4/S_2	M_1/M_3	R_9	R_8

Link units positioned within clusters are as follows:
- R055 position 2/5, R5
- R067 position 4/5, R5
- R083 position 2/4, R6
- R101 position 3/4, R1
- R102 position 3/4, R4
- R104 position 4/4 R6
- R111 position 4/4, R4
- R219 position 1/4, R1
- R220 position 4/4, R7
- R227 position 1/4, R4
- R245 position 1/5, R2.

In all cases the link items were taken from clusters that never appeared at the end of the booklet and that were not preceded by mathematics or science items. Also about half of the units (5/11) appeared in the first half of the cluster. This implies that students found these items easier to attempt than if the test design had been different.

Also, items were taken from 6 clusters, rather than taking intact clusters.

[22] Prepared by Jude Cosgrove.

Test Designs of Subsequent Cycles

The test design of subsequent cycles was changed since the 2000 design was not balanced and gave rise to booklet effects which were corrected by applying a linear transformation by booklet and domain; however it is not clear from the 2000 technical report the extent to which this actually fixed the problem of over-estimating student ability in reading in the first seven booklets.

In each subsequent PISA cycle the booklet design placed each 30-minute cluster in each of the four possible positions and, generally, each booklet contained two or more test domains. This resulted in an increase in the number of booklets. The PISA 2009 test design is shown in Table C2 in order to compare where the link items are with PISA 2000. Clusters containing link items are marked in bold.

Table C2. PISA 2009 Test Design

Booklet	P1	P2	P3	P4
1	M1	**R1**	R3A	M3
2	**R1**	S1	**R4A**	R7
3	S1	R3A	M2	S3
4	R3A	**R4A**	S2	**R2**
5	**R4A**	M2	R5	M1
6	R5	R6	R7	R3A
7	R6	M3	S3	**R4A**
8	**R2**	M1	S1	R6
9	M2	S2	R6	**R1**
10	S2	R5	M3	S1
11	M3	R7	**R2**	M2
12	R7	S3	M1	S2
13	S3	**R2**	**R1**	R5

The link between 2000 and 2009 was established by again using two intact clusters of eight units – R1 and R2 – that had been used in 2003 and 2006. In an attempt to increase the stability of the trends, three additional units from 2000 were put into cluster 4A (along with another non-trend unit in position three of that cluster).

As can be seen from the table above, each trend cluster appears in all four positions.
Of these 12 cases, seven clusters were either in first position or preceded by another reading cluster; one was preceded by a mathematics cluster, and 4 preceded by science clusters. So even though the design is balanced in terms of positioning, it is not balanced in terms of preceding and succeeding domains.

It should also be noted that in this design, two booklets contain no trend items; one of these booklets consists entirely of reading items and the other consists of all three domains.

The positioning of the link units within clusters are as follows:
- R055 position 3/4, R2
- R067 position 2/4, R1
- R083 position 1/4, R4A
- R101 position 4/4, R4A
- R102 position 3/4, R4A
- R104 position 4/4 R2
- R111 position 2/4, R2

- R219 position 1/1, R1
- R220 position 4/4, R1
- R227 position 1/4, R2
- R245 position 3/4, R2.

Possible Cluster Positioning Effect
Comparing cluster positioning across 2000 and 2009 and hypothesizing that within-cluster position could have an impact on student performance, it could be argued that, in 2009, the clusters would behave as follows:

- R055 – harder
- R067 – easier
- R083 – easier
- R101 – harder
- R102 – no change
- R104 – no change
- R111 – easer
- R219 – no change
- R220 – no change
- R227 – no change
- R245 – harder

Of course, this would not take the effect of whether a preceding cluster is the same domain or not. Percent correct, percent not reached and percent missing by cluster is shown in Table C3 for the link clusters used in PISA 2000 and PISA 2009. The results do not, generally speaking, support the hypothesis that within-cluster position is related to difficulty. In fact the average percent correct on all clusters has decreased by between 5 to 10 percentage points with the exception of R067.

Item Deletion Within Clusters Used to Measure Trends
Six items were deleted from five of the trend clusters as follows (with Irish and OECD average percent correct, respectively, in brackets):

R102 (Q01; 89.0 vs 81.5, Q06; 37.3 vs 33.7)
R108 (Q08; 62.5 vs 55.8)
R111 (Q04; 88.2 vs 78.1)
R227 (Q04; 74.8 vs 80.3)
R104 (Q06; 81.5 vs 79.7)

These deletions have an unknown consequence for the measurement of trends from 2003 onwards. We examined unit R102 in particular as this had the first item removed from the unit – which was multiple choice – so that the first item in the unit became a written response item from 2003 onwards. This appears to have put students off the new ‘first’ item in that unit, which had percentages correct of 38.3, 30.4, 31.9 and 28.0 in 2000, 2003, 2006 and 2009, respectively.

Table C3. Percent Correct, Not Reached, and Missing by Cluster, PISA 2000 and PISA 2009 – Ireland

Cluster	Name	Percent Correct 2000	Percent Correct 2009	Not reached 2000	Not reached 2009	Missing 2000	Missing 2009	Expected change 2000-2009
R055	DS	75.5	68.7	0.7	2.6	4.0	8.3	Decrease
R067	Aes	72.6	73.5	4.5	1.7	4.3	5.3	Increase
R083	House	78.1	70.2	0.8	1.1	0.8	3.0	Increase
R101	Rhi	70.3	60.4	0.5	3.4	2.1	2.7	Decrease
R102	Shirts	60.4	55.1	0.9	2.4	5.6	9.3	No change
R104	Tel	54.2	48.4	4.9	3.9	2.6	3.7	No change
R111	Exch	53.6	48.8	2.3	1.9	8.7	12.1	Increase
R219	Employ	81.5	75.5	0.1	1.2	3.8	7.4	No change
R220	Pole	63.2	56.3	2.6	3.2	4.9	8.5	No change
R245	Opt	72.1	63.3	0.2	2.2	1.2	3.9	Decrease

Appendix D: International and National Item Parameters, all Reading Items, PISA 2009

Table D1 shows international and national item parameters for all reading items used in 2009. Parameters have an international mean of 0.0 and standard deviation of 1.0 and minus values indicate easier items while plus values indicate harder ones.

Table D1. International and National Item Parameters, all Reading Items, PISA 2009

Item	International	National
R456Q01	-3.396	-3.400
R446Q03	-2.484	-2.712
R067Q01	-2.135	-2.405
R412Q01	-1.650	-2.217
R420Q02	-1.583	-0.884
R432Q01	-1.572	-1.944
R102Q07	-1.547	-2.420
R055Q01	-1.459	-1.427
R101Q02	-1.457	-1.466
R219Q02	-1.423	-1.704
R083Q02	-1.382	-1.333
R456Q02	-1.380	-2.478
R460Q05	-1.369	-1.758
R456Q06	-1.365	-1.611
R220Q05	-1.347	-1.098
R453Q01	-1.307	-1.989
R104Q01	-1.255	-1.252
R466Q06	-1.171	-0.787
R083Q03	-1.049	-0.819
R446Q06	-1.045	-1.264
R455Q03	-1.041	-1.204
R447Q05	-1.037	-0.851
R420Q09	-1.037	-1.060
R101Q04	-1.025	-1.342
R424Q07	-0.912	-1.111
R227Q06	-0.888	-0.802
R055Q05	-0.771	-0.805
R406Q05	-0.763	-1.255
R404Q03	-0.757	-0.958
R432Q05	-0.660	-1.090
R453Q06	-0.611	-1.111
R442Q03	-0.592	-1.082
R442Q02	-0.501	-1.323
R220Q06	-0.435	0.209
R245Q02	-0.415	-0.295

R111Q01	-0.377	-0.076
R245Q01	-0.377	-0.005
R420Q10	-0.354	-0.319
R406Q01	-0.339	-0.203
R424Q03	-0.325	0.030
R083Q04	-0.307	-0.135
R460Q01	-0.295	-0.562
R452Q04	-0.266	0.041
R067Q05	-0.264	-0.458
R455Q04	-0.251	-0.032
R447Q01T	-0.251	-0.650
R220Q02B	-0.179	-0.072
R453Q05T	-0.170	-0.461
R453Q04	-0.159	-0.961
R414Q09	-0.092	-0.488
R220Q04	-0.005	0.474
R460Q06	0.003	0.266
R055Q03	0.012	-0.238
R101Q03	0.028	0.274
R083Q01	0.072	0.019
R067Q04	0.080	0.011
R227Q02T	0.093	0.337
R458Q04	0.095	0.388
R412Q05	0.111	0.420
R227Q01	0.172	0.920
R227Q03	0.225	0.386
R458Q07	0.237	0.800
R447Q04	0.330	0.877
R437Q06	0.333	0.489
R437Q01	0.335	0.830
R414Q06	0.379	0.130
R055Q02	0.474	0.276
R101Q01	0.486	0.993
R458Q01	0.519	0.528
R452Q06	0.633	0.419
R404Q06	0.654	0.676
R102Q05	0.668	0.596
R447Q06	0.687	0.762
R452Q07	0.760	0.810
R111Q06B	0.762	0.909
R420Q06	0.768	0.698
R101Q05	0.808	0.863
R414Q02	0.817	0.812
R466Q02	0.855	1.118

R220Q01	0.901	1.030
R404Q10A	0.940	0.907
R424Q02T	0.958	0.948
R412Q08	1.052	0.629
R412Q06T	1.114	1.121
R111Q02B	1.171	1.051
R404Q10B	1.189	1.276
R442Q07	1.196	1.634
R455Q02	1.245	0.739
R104Q02	1.330	1.268
R442Q05	1.345	1.895
R414Q11	1.355	1.722
R102Q04A	1.422	1.734
R406Q02	1.426	1.281
R404Q07T	1.436	1.867
R455Q05T	1.960	1.784
R442Q06	1.969	2.028
R104Q05	2.475	2.866
R437Q07	2.483	2.562
R466Q03T	2.662	3.246
R432Q06T	2.890	
R452Q03	2.918	3.112

Appendix E: Reading Literacy Framework Components 2000-09

The reading literacy framework used in PISA 2000 and PISA 2009 is intended to be comprehensive, coverall all aspects reading and allowing for the publication performance on reading subscales as well as performance on overall reading. In the intermediate years (2003 and 2006), the smaller sets of items used (the ‘link’ items) were intended to mirror the full framework used in 2000. Table 2c illustrates aspects of the framework in 2000 and 2009 (2003 and 2006 were the same).   It can be seen that the item set (link items) used in 2003 and 2006 is broadly representative of the full item sets used in 2000 and 2003.

Table E1. PISA Reading Literacy Framework (2000-2009)

		Items (%) 2000	Items (%) 2003/6	Items (%) 2009
Text structure	Continuous Texts	89 (63%)	18 (64%)	81 (61.8%)
	Non-continuous	52 (37%)	28 (36%)	38 (29%)
	Mixed/Multiple	-	-	12 (9.1%)
Process	Interpret*	70 (50%)	14 (50%)	67 (51.1%)
	Reflect/Evaluate	29 (21%)	7 (25%)	33 (25.2%)
	Retrieve	42 (29%)	7 (25%)	31 (23.7%)
	Complex	-	-	
Context	Educational	39 (28%)	7 (25%)	38 (29%)
	Occupational	22 (16%)	7 (25%)	21 (16%)
	Personal	26 (18%)	6 (21%)	37 (28.2%)
	Public	54 (38%)	7 (25%)	35 (26.7%)

*Integrate and interpret in 2009; **Access and retrieve in 2009. Not all items were administered in all countries; not all administered items were used in scaling.

Appendix F: Additional Analyses

Table F1 compares the very low achieving schools with others on a range of characteristics.

Table F1. Characteristics of outlier and non-outlier schools in PISA 2009 (continuous variables)

Characteristic	Not outlier	Outlier	Total
Mean % missing	5.70	16.43	6.30
SD % missing	7.50	16.74	8.02
Proportion > 25% missing	0.05	0.28	0.06
Reading mean*	501.51	371.02	494.26
SD reading*	81.09	97.20	81.99
Mathematics mean*	490.59	390.26	485.02
SD mathematics*	74.44	89.85	75.29
Science mean*	512.23	397.54	505.86
SD science*	84.48	95.89	85.12
SES mean	0.01	-0.66	-0.02
SD SES	0.90	0.81	0.89
Fee waiver mean	26.90	50.66	28.43
Proportion female	0.51	0.32	0.50
Proportion other language	0.03	0.14	0.04
Average number of participants	27.75	20.38	27.34

*Computed on the basis of the first plausible value only

Table F2 compares the characteristics of categorical variables for outlier and non-outlier schools.

Table F2. Characteristics of outlier and non-outlier schools in PISA 2009 (categorical variables)

Characteristic	Not outlier	Outlier
Location of TA		
Internal TA	77.2	75.0
External TA	22.8	25.0
Gender of TA		
Male TA	47.4	37.5
Female TA	52.6	62.5
Position of TA		
Principal/Deputy	33.3	25.0
Teacher	30.5	37.5
Careers guidance/SEN	36.2	37.5
School sector		
Comm/Comp	16.9	12.5
Secondary	62.5	25.0
Vocational	20.6	62.5

Table F3 shows some of the characteristics of the individual outlier schools.

Table F3. Characteristics of individual outlier schools in PISA 2009

	School ID							
Characteristic	5025	7006	7007	8010	6012	7004	3019	8011
Mean % missing	12.12	19.84	11.56	18.63	14.17	18.24	20.23	16.69
SD % missing	13.34	17.65	10.08	15.37	15.84	19.9	21.66	20.11
Proportion > 25% missing	0.17	0.27	0.11	0.47	0.27	0.33	0.39	0.26
Reading mean*	393.7	344.7	396.7	393.9	378.4	367.4	395.3	298
SD reading*	89.92	101.7	80.81	40.15	122.2	108.4	110.5	124
Mathematics mean*	422.8	337.9	442.3	381.3	410.7	402.2	381.2	343.7
SD mathematics*	81.39	97.11	76.77	53.42	113	83.72	91.82	121.6
Science mean*	410.4	339.5	503.1	394.5	440.5	363.2	401.3	327.9
SD science*	89.04	96.24	90.28	51.69	112.5	94.27	103.6	129.5
SES mean	-1.1	-0.87	-0.38	NA	0.18	-0.56	-1.16	-0.74
SD SES	0.51	0.67	1.04	NA	1.12	0.81	NA	0.69
Fee waiver mean	60.49	70.12	47.4	57.64	14.98	56.72	53.41	44.52
Poportion female	0	0.27	0.11	0.8	0	0.57	0.43	0.35
Proportion other language	0.18	0.31	0.13	NA	0.04	0.17	NA	0
Average number of participants	23	15	9	15	26	21	23	31
TA location	Internal	Internal	Internal	Internal	External	External	Internal	Internal
TA gender	Male	Female	Male	Female	Male	Female	Female	Female
TA position	Teacher	Careers Guidance/SEN	Careers Guidance/SEN	Teacher	Principal/Deputy	Principal/Deputy	Teacher	Careers Guidance/SEN
Sector	Comm/Comp	Secondary	Secondary	Vocational	Comm/Comp	Vocational	Vocational	Secondary

*Computed on the basis of the first plausible value only

**Questionnaire data available for just one student

Table F4 compares ‘large drop’ and other schools on a number of categorical characteristics.

Table F4. Characteristics of schools common to PISA 2000 and 2009 showing an achievement drop of more than 100 scale points (categorical variables)

Characteristic	Drop < 100 points	Drop > 100 points
Location of TA		
Internal TA	76.3	100.0
External TA	23.7	0.0
Gender of TA		
Male TA	47.8	20.0
Female TA	52.2	80.0
Position of TA		
Principal/Deputy	34.3	0.0
Teacher	30.6	40.0
Careers guidance/SEN	35.2	60.0
School sector		
Comm/Comp	15.1	60.0
Secondary	61.9	20.0
Vocational	23.0	20.0

Table F5 compares country average performance on PISA 2009 and ICCS 2009.

Table F5. Comparison of Country Mean Performance on PISA 2009 and ICCS 2009 (N=33)

Country	ICCS 2009	PISA reading 2009	PISA rank	ICCS rank	Score Diff
Denmark	576	495	15	2	81
Chinese Taipei	559	495	14	4	63
Slovak Republic	529	477	23	12	51
Colombia	462	413	32	30	49
Russian Federation	506	459	27	20	47
Italy	531	486	17	11	45
Finland	576	536	2	1	41
Sweden	537	497	12	6	40
Ireland	534	496	13	8	38
Bulgaria	466	429	29	29	37
Lithuania	505	468	26	21	37
Poland	536	500	10	7	36
Chile	483	449	28	25	34
Slovenia	516	483	19	16	33
Austria	503	470	25	23	33
Liechtenstein	531	499	11	10	32
Czech Republic	510	478	22	19	32
Switzerland	531	501	9	9	31
Indonesia	433	402	33	33	31
Thailand	452	421	31	32	30
Mexico	452	425	30	31	26
Korea	565	539	1	3	26
United Kingdom	519	494	16	14	25
Estonia	525	501	8	13	24
Spain	505	481	21	22	24
Hong Kong-China	554	533	3	5	21
Norway	515	503	7	17	11
Belgium	514	506	6	18	8
Luxembourg	473	472	24	28	1
Latvia	482	484	18	26	-2
New Zealand	517	521	4	15	-4
Greece	476	483	20	27	-7
Netherlands	494	508	5	24	-15

Table F6 shows the characteristics of individual 'large drop' schools.

Table F6. Characteristics of individual 'large drop' schools in PISA 2009

	School ID				
Characteristic	3007	3017	3019**	6053	8011
Mean % missing	10.34	7.59	20.23	6.12	16.69
SD % missing	13.9	10.44	21.66	9.23	20.11
Proportion > 25% missing	0.14	0.08	0.39	0.08	0.26
Reading mean*	420.4	463.51	395.33	453.33	297.95
SD reading*	106.68	109.79	110.51	87.85	123.97
Mathematics mean*	432.45	492.53	381.18	506.37	343.7
SD mathematics*	82.85	90.94	91.82	76.93	121.59
Science mean*	430.94	504.34	401.32	494.92	327.86
SD science*	105.34	105.99	103.56	90.94	129.53
SES mean	0	-0.15	-1.16	0.13	-0.74
SD SES	0.79	1.03	NA	1.14	0.69
Fee waiver mean	21.48	30.95	53.41	13.07	44.52
Poportion female	0.48	0.54	0.43	0	0.35
Proportion other language	0.04	0.04	NA	0	0
Average number of participants	29	26	23	26	31
TA location	Internal	Internal	Internal	Internal	Internal
TA gender	Male	Female	Female	Female	Female
TA position	Careers Guidance/SEN	Careers Guidance/SEN	Teacher	Teacher	Careers Guidance/SEN
Sector	Comm/Comp	Comm/Comp	Comm/Comp	Secondary	Vocational

*Computed on the basis of the first plausible value only

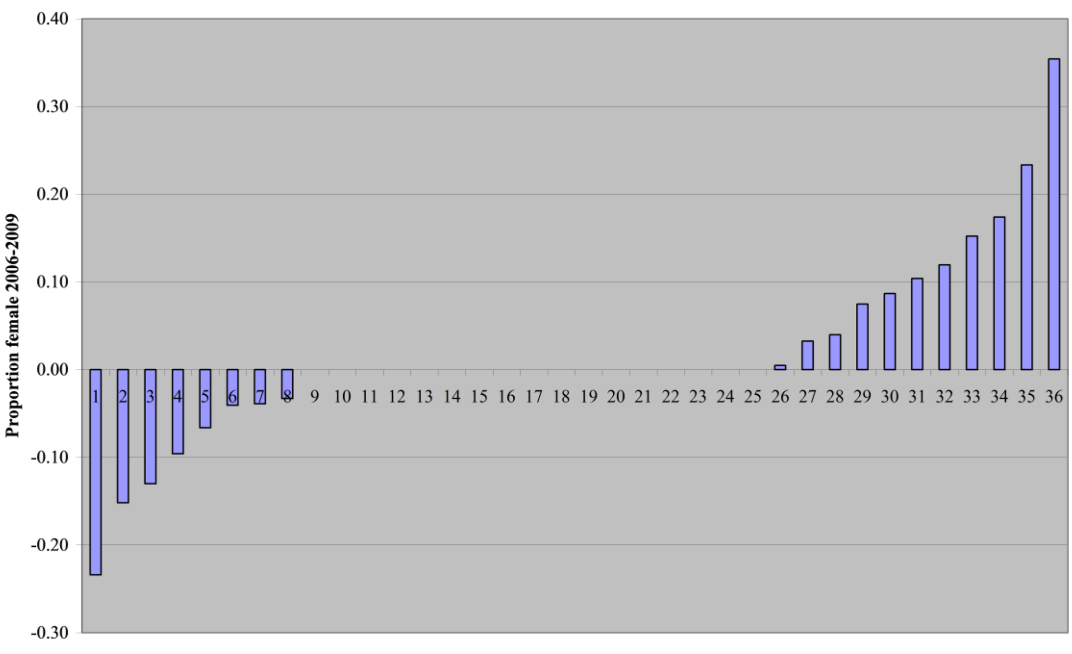
**Questionnaire data available for just one student

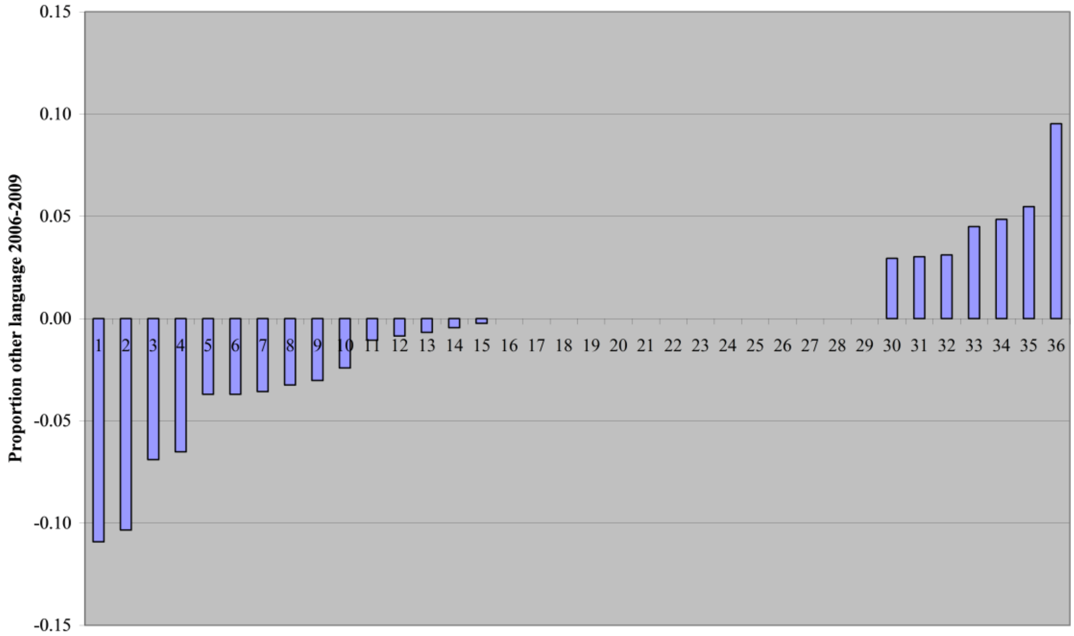
Table F7 shows various characteristics of schools that participated in both PISA 2000 and 2009.

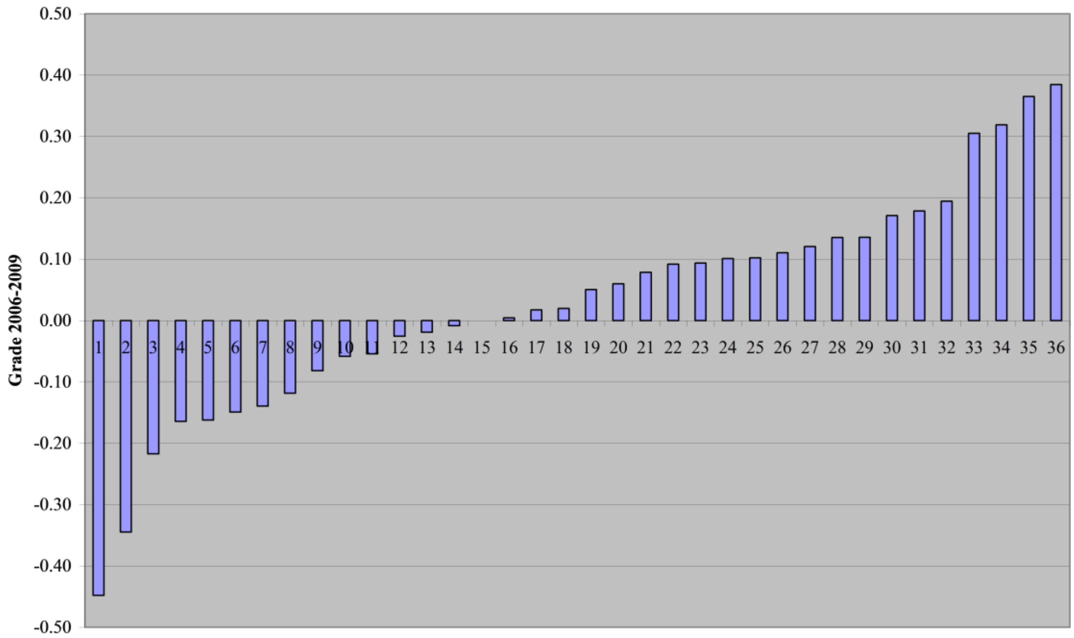
Table F7. Characteristics of schools participating in both PISA 2006 and 2009 (N=36)

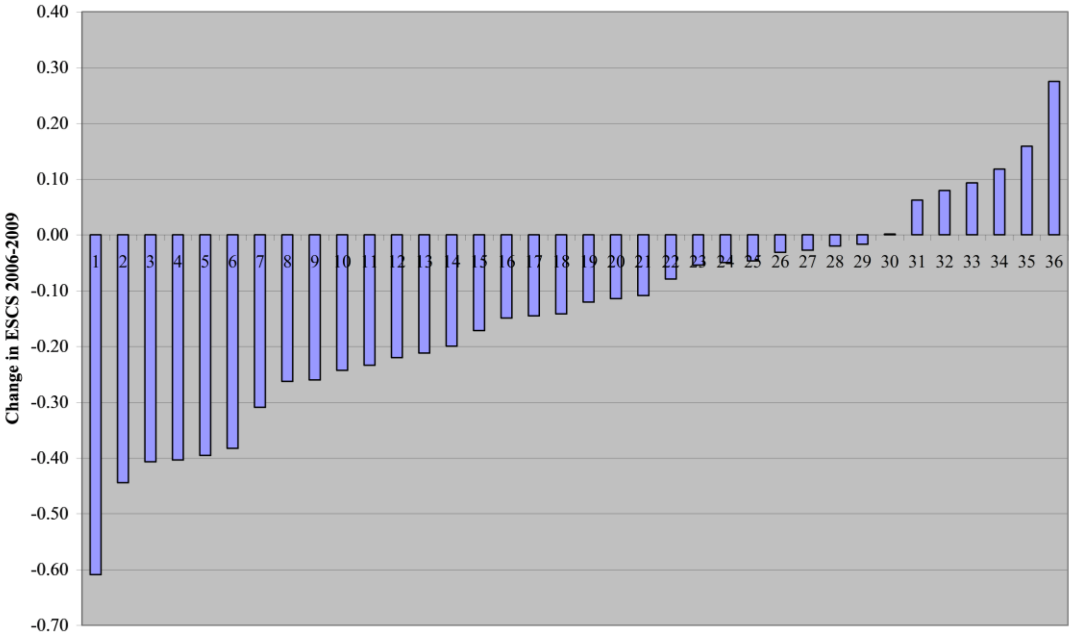
	2006		2009	
Characteristic	Mean	SD	Mean	SD
Proportion female	0.472	0.365	0.456	0.362
Proportion other language	0.029	0.062	0.036	0.060
Average grade level	9.502	0.205	9.473	0.171
Average ESCS	-0.014	0.479	0.127	0.447
Average PISA reading	510.849	54.648	503.561	57.104
Average PISA mathematics	498.937	41.070	504.102	54.269
Average PISA science	505.647	43.297	505.886	54.596
Average JCE English	9.327	0.868	9.498	0.839
Average JCE mathematics	8.609	1.096	8.714	0.900
Average JCE science	9.330	0.825	9.567	0.750

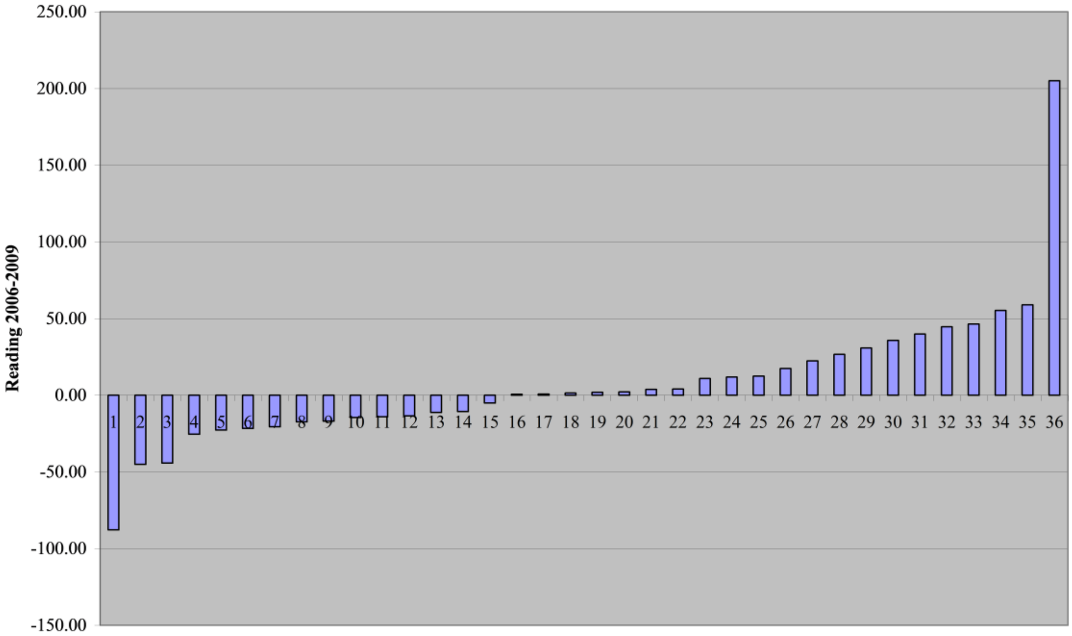
Figures F1-F10: Frequency distributions for characteristics shown in Table F6 (N=36).

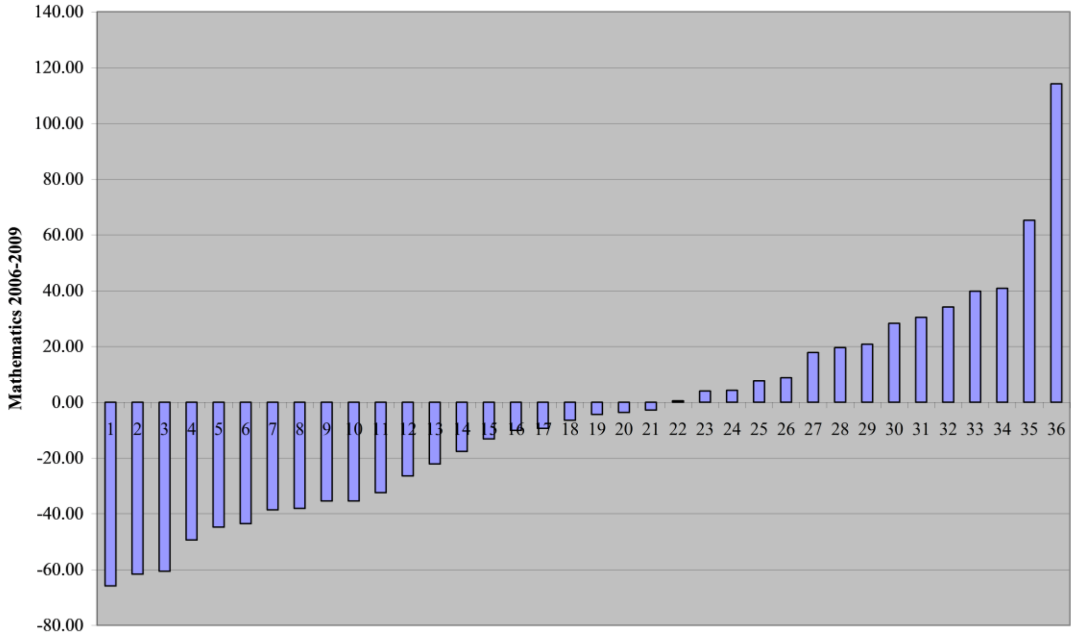



Proportion other language 2006-2009
0.15
0.10
0.05
0.00
-0.05
-0.10
-0.15
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36


Grade 2006-2009
0.50
0.40
0.30
0.20
0.10
0.00
-0.10
-0.20
-0.30
-0.40
-0.50
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36

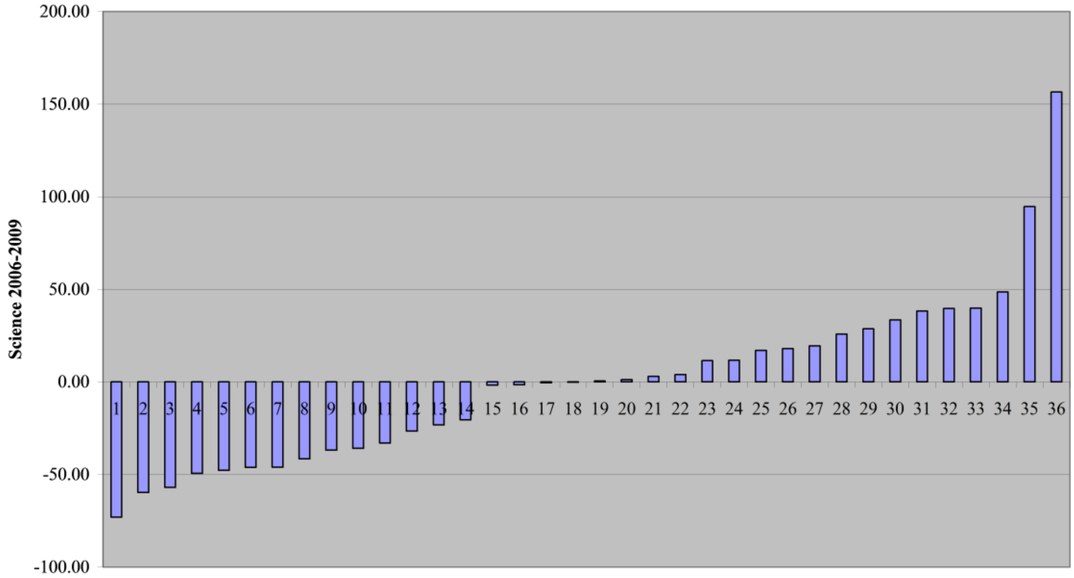

Change in ESCS 2006-2009
0.40
0.30
0.20
0.10
0.00
-0.10
-0.20
-0.30
-0.40
-0.50
-0.60
-0.70
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36


Reading 2006-2009
250.00
200.00
150.00
100.00
50.00
0.00
-50.00
-100.00
-150.00
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36

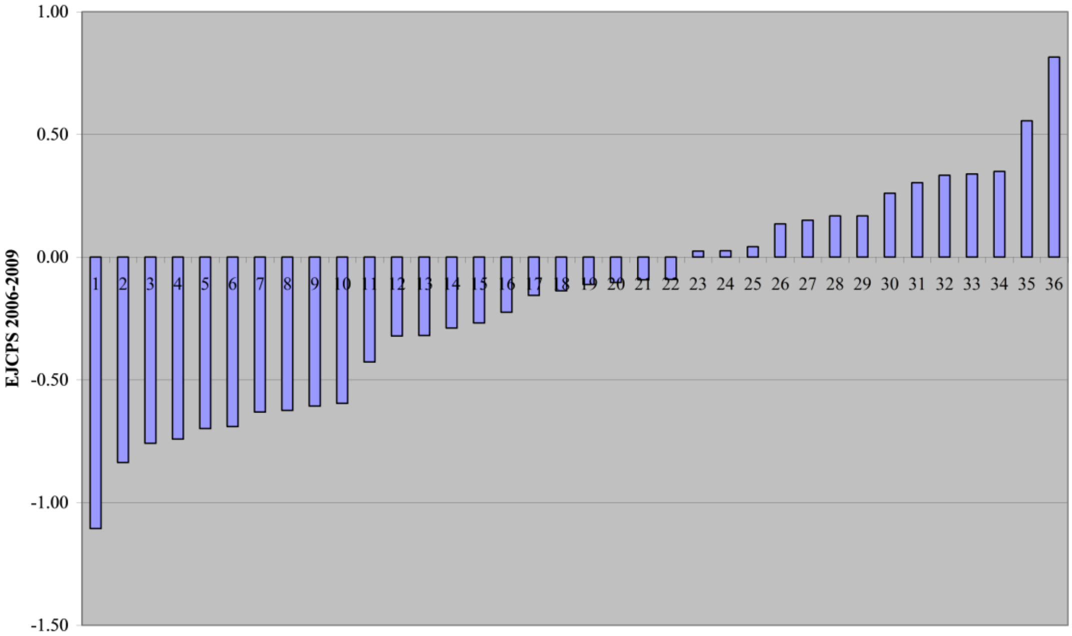


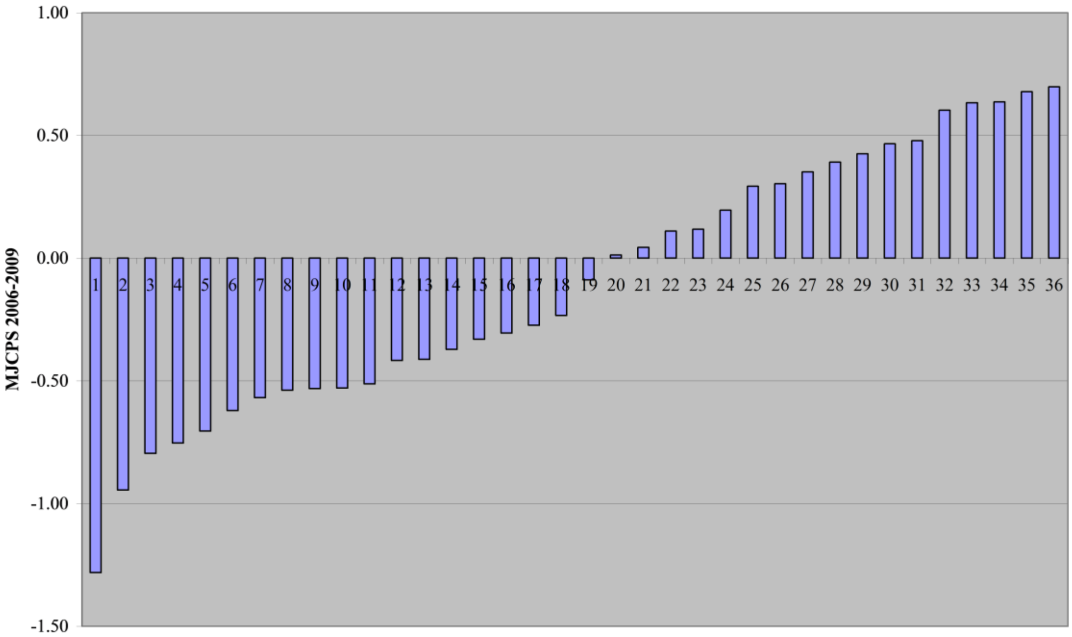
Mathematics 2006-2009
140.00
120.00
100.00
80.00
60.00
40.00
20.00
0.00
-20.00
-40.00
-60.00
-80.00
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36

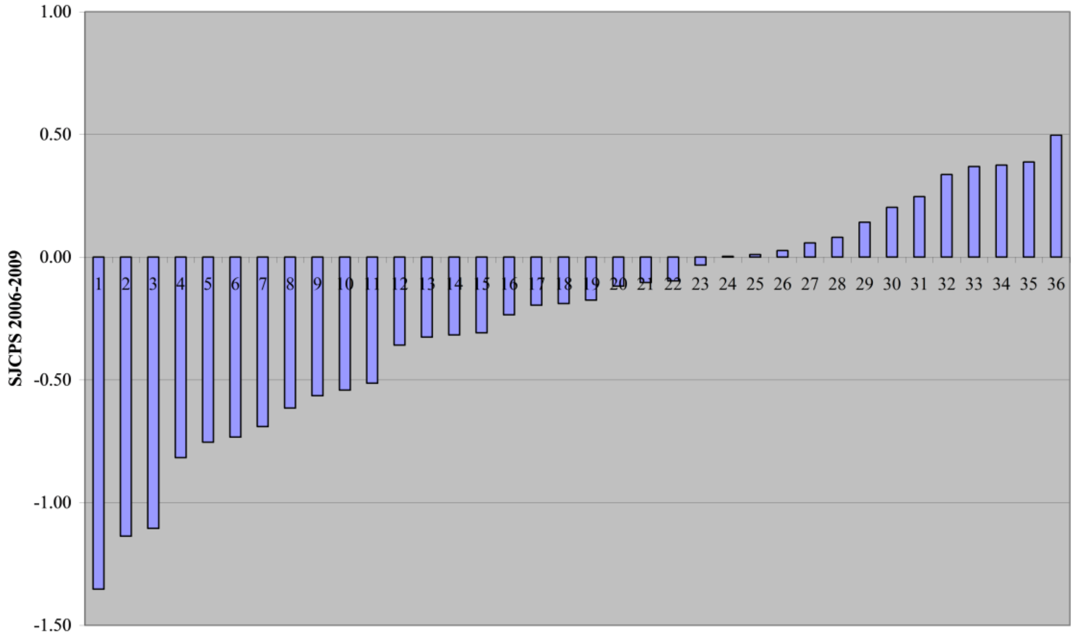
sciediff



Science 2006-2009
200.00
150.00
100.00
50.00
0.00
-50.00
-100.00
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36


1.00
0.50
0.00
-0.50
-1.00
-1.50
EJCPS 2006-2009
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36


1.00
0.50
0.00
-0.50
-1.00
-1.50
MJCPS 2006-2009
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36



Tables F8 and F9 compare characteristics of schools according to whether they had an internal or external test administrator.

Table F8. Characteristics of schools with internal and external test administrators (continuous variables)

Characteristic	Internal TA	External TA	Total
Mean % missing	6.3	6.2	6.3
SD % missing	8.1	7.8	8.0
Proportion > 25% missing	0.1	0.1	0.1
Reading mean*	492.1	501.7	494.3
SD reading*	81.3	84.5	82.0
Mathematics mean*	482.7	492.9	485.0
SD mathematics*	74.4	78.3	75.3
Science mean*	504.2	511.6	505.9
SD science*	84.3	87.8	85.1
SES mean	-0.1	0.2	0.0
SD SES	0.9	0.9	0.9
Fee waiver mean	30.1	22.2	28.4
Proportion female	0.5	0.5	0.5
Proportion other language	0.0	0.0	0.0
Average number of participants	27.1	28.1	27.3

*Computed on the basis of the first plausible value only

Table F9. Characteristics of schools with internal and external test administrators (categorical variables)

Characteristic	Internal TA	External TA
Gender of TA		
Male TA	40.9	66.7
Female TA	59.1	33.3
Position of TA		
Principal/Deputy	25.6	
Teacher	35.6	
Careers guidance/SEN	38.9	
School sector		
Comm/Comp	19.8	6.1
Secondary	55.9	75.8
Vocational	24.3	18.2