Large-scale Assessments
in Education
a SpringerOpen Journal

Changes in achievement on PISA: the case of Ireland and implications for international assessment practice

Jude Cosgrove[1*] and Fernando Cartwright[2]

* Correspondence:
jude.cosgrove@erc.ie
[1]Educational Research Centre,
St Patrick's College, Drumcondra,
Dublin, IRELAND
Full list of author information is
available at the end of the article

Abstract

The PISA 2009 results for Ireland indicated a large decline in reading literacy scores since PISA 2000 (the largest of 38 countries). The decline in mathematics scores since PISA 2003 was the second largest of 39 countries. In contrast, there was no change in science achievement since PISA 2006. These results prompted detailed investigations into possible reasons for the declines, particularly in reading. This paper considers the changes in achievement observed for Ireland in PISA 2009 under two themes: implementation of PISA in Ireland and changes in the cohort of students participating in PISA, and response patterns on the PISA test (as measures of student engagement). It is argued that the case of Ireland represents the 'perfect storm', since a range of factors appear to have been in operation to produce the results. The discussion attempts to show how the case of Ireland can be relevant to other countries which may have experienced changes in PISA test scores over time. Some of the findings have relevance to international practice in large-scale surveys of educational achievement more generally.

Keywords: OECD PISA, Ireland, Reading literacy, Trends, International assessment

Review

When first published, the results for PISA 2009 reading, and to a lesser extent mathematics, attracted media attention and commentary in Ireland. For example, the *Irish Times* discussed the results under the headline "shattering the myth of a world-class education system" (December 8, 2010), while the *Irish Independent* noted that "There was shock last year when it emerged there was a fall in reading and maths scores for Irish students in the PISA" (April 3, 2012). *Education Matters* described the results as "an urgent call to action" (December 14, 2010). Ireland's mean reading score on PISA showed the largest decline since 2000 across the 38 countries for which results could be compared (31 score points, or close to one-third of an international standard deviation), and mathematics showed the second-largest decline since 2003 across the 39 countries that could be compared (16 points, or one-sixth of an international standard deviation). In contrast, achievement in science remained stable (OECD, 2010a). These patterns of results were unexpected given the absence of other evidence of a decline in educational standards in Ireland.

Springer

Before the results for PISA 2009 were available for national centres to develop their own national reports, it had been planned to adopt a reporting strategy in Ireland that was similar to previous cycles; i.e. a short report for release at the same time as the initial OECD report in December 2010, followed by a more detailed national report in the following year. However, the unexpected results for 2009 necessitated several other reports and analyses. First, the Department of Education and Skills (Ireland) decided to seek input from independent international experts in explaining the Irish results. These inquiries produced two comprehensive reviews of the results (Cartwright, 2011; LaRoche and Cartwright, 2010). Second, staff at the national centre undertook additional analyses to try to disentangle some of the possible reasons for the Irish PISA 2009 achievement scores, particularly in reading (Cosgrove, 2011; Cosgrove and Moran, 2011; Cosgrove, Shiel, Archer and Perkins, 2010; Shiel, Moran, Cosgrove and Perkins, 2010). These reviews contained analyses relating to four key issues: those relating to demographic changes; features of the PISA test across cycles; response patterns on the PISA test over time; and the methods used in PISA to estimate changes in achievement. Our aim in this paper is to draw on two of these themes and to use Ireland as a case example to illustrate issues in PISA's measurement of change, and the complexities in interpreting change more generally.

This topic has relevance for other countries as well as for other large international assessments of education. PISA's potential use by policy-makers to monitor education systems and effect policy changes on the basis of the results implies that a good understanding of what the results mean is required for appropriate policy interventions, while misinterpretation or partial interpretation could result in misguided and erroneous interventions. Specifically, the paper introduces several additional sources of evidence which should be considered in the context of using PISA results to evaluate the efficacy of policy changes over time.

The paper is organised into two main sections that cover overlapping themes. First, we examine possible direct explanations for the decline in achievement scores: the implementation of PISA in Ireland (e.g. sampling, test administration), and changes in the demographic characteristics of the cohort of students who participated. Second, we present the results of analyses that examine changes in the extent to which students may have engaged in the assessment tasks.

For the sake of brevity, the aims, design and international results of PISA are not considered here; instead, readers are referred to the OECD's reports on PISA 2009 (OECD 2010a, b, c, d, e, 2011a; particularly 2010e), the PISA 2009 technical report (OECD, 2011b), and the PISA 2009 assessment framework (OECD, 2009a). Additional international reports are available at www.oecd.org/pisa/, while national reports on PISA for Ireland are at www.erc.ie/pisa.

Implementation of PISA in Ireland and changes in the PISA cohort over time

PISA implements stringent quality control with rigorous standards on aspects including sampling, translation, printing, test administration, data processing, scaling, and student and school participation rates. Ireland met all technical standards in PISA 2009 (see OECD, 2011b, Chapter 14), as it has in all previous cycles of PISA.

Although the implementation of PISA in Ireland was technically sound, a number of procedural changes were introduced in Ireland in PISA 2009. The possible relevance of these changes to changes in observed performance are documented in several reports

(Cosgrove, Shiel, Archer and Perkins, 2010; LaRoche and Cartwright, 2010; Shiel, Moran, Cosgrove and Perkins, 2010).

The changes implemented in PISA 2009 were first, in order to incentivise student participation, a prize draw was introduced in which three participating students in each school received a 15 euro voucher which good be exchanged in a number of different shops (e.g. for books, music, games). While this could have served to attract a some-what higher number of disengaged students, analyses of the sampling outcomes suggest that this was not the case (Cosgrove et al., 2010). No major issues with testing were identified by the PISA Quality Monitor[a] for Ireland, although some disengagement among students was observed by test administrators (LaRoche and Cartwright, 2010). It is possible that other countries may also have found student engagement to be a problem, though systematic information on this is not available. Further discussion on the disengagement issue can be found in Section 2.

Second, the 'school associate' model of test administration was used for the first time in Ireland; that is, tests were administered by teachers in their own school. About three-quarters of schools in Ireland employed this model, while an external administrator was used in the remaining schools. All individuals administering the assessment instruments in schools received the same training by national centre staff. Schools with internal and external test administrators did not differ significantly in their mean achievement scores or socioeconomic characteristics (Cosgrove et al., 2010).

Third, Ireland participated in two international assessments of education in spring 2009 (PISA and the International Civics and Citizenship Study; ICCS). Both of these drew on samples of post-primary schools, the total number of which is small (around 720). To prevent overlap of sampled schools across the studies, the list of schools was split into equivalent halves and each sample was drawn from half of all schools. Also, a new implicit stratification variable was introduced in PISA 2009: the percentage of students in each school entitled to a fee waiver on the State examinations taken at Grade 9 (an indicator of the percentage of students in families in receipt of social welfare benefits). Analyses conducted by LaRoche and Cartwright (2010) and Cosgrove et al. (2010) confirmed that the changes made to the sampling methodology did not affect the representativeness of the PISA sample, response rates, or sampling weights, in any measurable way.

Fourth, though not a procedural change *per se*, it was found that, while in 2000, all schools that participated in PISA achieved a mean reading score that was within one international standard deviation of the mean (i.e., the mean ± 100 points); in 2009, eight schools each had very low average reading achievement (more than 100 points below the mean score for Ireland (note that these estimates do not take measurement and sampling error into account and should be interpreted in a broad sense). Test adminis-tration records for these schools were examined but failed to reveal any difficulties with test administration. Analyses of the characteristics of these 'outlier' schools (Cosgrove et al., 2010; Cosgrove and Moran, 2011) revealed substantial differences in the charac-teristics of students in outlier and non-outlier schools. For example, students in the eight schools had almost three times as many missing responses on their test booklets as students in other schools; had a mean ESCS (Economic, Social and Cultural Status)[b] score that was 0.6 standard deviations lower than in other schools; and were 1.4 times as likely to be boys (i.e. odds ratio = 1.4), 4 times as likely to speak a language other

than the language of the assessment, and 3 times more likely to be in a vocational school type, than students in the other schools.

Possibly linked with the appearance of outlier schools in 2009, there have been some marked demographic changes in the school-going population in Ireland since 2000, though it is very unlikely that these alone account for the achievement decline. With the exception of Spain, Ireland has experienced the highest increase in the number of immigrant students participating in PISA, from 2.3% in 2000 to 8.3% in 2009 (OECD, 2010e). The percentage of students who spoke a language other than the language of instruction at home increased fourfold during this time, from 0.9% to 3.6%. Immigrant students in Ireland had a significantly lower average ESCS score in 2009 than they did in 2000 (Cosgrove et al., 2010). Between 2000 and 2009, the reading scores of the immigrant student group dropped by 53 score points (about half an international standard deviation), and those of students speaking a language at home other than the language of the assessment dropped by 62 points.

There was also a small decrease in the percentage of PISA-eligible students who had already left the education system (from 2.1% in 2000 to 1.5% in 2009). Higher retention of these students could have contributed to a small portion of the score decline because these students are likely to be lower achievers. Furthermore, greater numbers of children with special educational needs (SEN) have been integrated into mainstream schools since 2000[c]. However, although 3.5% of students who participated in 2009 were classified as having an SEN, corresponding data for 2000 are not available. It is difficult, therefore, to quantify what, if any, effect this may have had on the PISA results.

Another difference between PISA 2000 and 2009 is the change in the distribution of students across grade levels. The percentage of students in Transition Year (Grade 10)[d] increased (from 16.0% to 24.0%), while there was a decrease from 18.6% to 14.4% in the percentage of students in Fifth Year (Grade 11), reflecting greater availability and uptake of the Transition Year programme in schools (Clerkin, 2013). The largest declines in average reading achievement occurred among students in Fifth Year, while the largest decline in mathematics occurred in Transition Year (Table 1). However, these declines cannot be accounted for by changes in the socioeconomic composition of students in different grade levels (Perkins et al., 2012).

To sum up: since the administration of PISA 2000, a number of demographic changes have occurred, chiefly an increase in the immigrant population that took part in PISA. Not only this, the composition of the immigrant population in 2009 is not the same as it was in 2000, being less socioeconomically advantaged than previously. Furthermore, policy changes over the past decade concerning retention rates and the inclusion of students with special educational needs in mainstream schools is having a noticeable impact on the composition of the PISA student samples. The increase of students taking the optional Grade 10 (Transition Year) programme is likely to reflect both the increased availability of this programme and the desire of some students to stay longer in school in the context of shrinking job opportunities. The changes made to the test administration procedures have also been noted. It is easy to see how complex the interpretation of change in PISA is given these factors, since many can expected to overlap and interact. Furthermore, we argue that the non-detection of an empirical effect of such changes does not altogether negate the possibility of such an effect.

Table 1 Comparisons of mean scores in print reading, mathematics and science across grade levels (Ireland, all PISA cycles, and differences in average achievement across cycles)

Domain/Grade	2000		2003		2006		2009		
	Mean	SE	Mean	SE	Mean	SE	Mean	SE	
Print reading									Diff 2009–2000
Second year (G8)	410.7	9.55	406.2	10.01	420.2	13.06	376.0	10.88	**−34.7**
Third year (G9)	516.9	3.60	502.8	3.23	506.9	3.85	487.9	3.43	**−29.0**
Transition year (G10)	568.4	4.52	562.0	4.48	547.8	4.70	525.3	4.42	**−43.1**
Fifth year (G11)	547.9	4.30	530.8	4.36	530.9	4.56	498.2	5.51	**−49.7**
Mathematics									Diff 2009-2003
Second year (G8)	409.1	12.14	406.8	9.48	414.9	9.54	384.8	11.63	−22.0
Third year (G9)	495.4	3.11	492.3	2.97	492.3	2.95	480.1	3.07	**−12.2**
Transition year (G10)	537.3	5.72	542.9	4.56	530.1	4.30	509.5	3.88	**−33.4**
Fifth year (G11)	516.6	4.48	515.1	5.32	511.5	4.18	496.1	4.86	**−19.0**
Science									Diff 2009-2006
Second year (G8)	425.8	10.49	400.5	9.95	408.5	11.0	403.7	10.24	−4.8
Third year (G9)	504.6	3.86	494.1	3.30	499.3	3.5	501.7	3.74	+2.4
Transition year (G10)	550.9	5.61	548.6	4.71	537.1	4.3	532.9	4.93	−4.2
Fifth year (G11)	529.6	5.15	518.8	5.23	519.6	4.3	510.0	5.57	−9.6

Note. Significant differences are in bold. International grade level equivalents are shown in brackets.

Response patterns on the PISA test across domains and cycles as measures of student engagement

Variations in student engagement and/or fatigue levels during low-stakes testing also interfere with student performance, which has a confounding effect on the estimation of student ability or proficiency (e.g. Boe, May and Boruch, 2002; Eklöf, 2007). When variance in student engagement is non-zero, any estimate that does not control for the effects of engagement will provide biased estimates for individual students (Wise and DeMars, 2005). In a low-stakes assessment such as PISA, a systematic reduction over time in levels of engagement or effort in equivalent cross-sections of students who have otherwise equivalent levels of proficiency is likely to produce an increase in the proportion of skipped or inadequately attempted responses to test questions (see van Barneveld, Pharand, Ruberto and Haggarty, 2013, pp. 46-48). If these responses are not distinguished from responses that are the product of genuine student effort, the results will inevitably produce declining estimates of achievement. Wise and DeMars (2010) found that even modest amounts of these skipped or poorly attempted questions a have a large impact on estimates of average performance.

The primary mechanism for distinguishing between effort and non-effort is response latency, the time span between the student's initial exposure to an item and the time that he or she either responds or skips to the next item, and, unfortunately, response latency is not available for the paper-based versions of the PISA assessments that were administered from 2000 to 2009 (since this may only be measured on computer-based assessments; van Barneveld et al., 2013). Thus, the evidence on the role of engagement is largely based on changes in strict item non-response over time (and is therefore likely to underestimate the effects of engagement on performance).

Borghans and Schils (2011) have analysed the PISA 2003 and 2006 international datasets to examine the effects of engagement/effort on performance. They showed that although there was a substantial drop in the performance of students as they progressed through the test (an indicator of test fatigue) across all countries, the size of this drop varied substantially. They also found that the magnitude of the drop was generally smaller for girls and students with higher test scores. Interestingly, the relationship between the size of the performance drop was not associated with socioeconomic status in the majority of countries (and only weakly and positively so in the remainder). Furthermore, the drop in performance was correlated across cycles, but only weakly related to achievement scores within cycles. Borghans and Schils argued that the performance drop may be taken as a proxy for test motivation, which is related to characteristics other than cognitive ones. The magnitude of the performance drop, which they term the 'motivation effect', explained 34% of the variation in PISA scores between countries. In Ireland, the magnitude of the performance drop in PISA 2006 (when science was a major domain) was small relative to a majority of countries, while the gender difference in the size of the performance drop was the third largest across the 38 countries in their analysis. It should be noted that Borghans and Schils did not examine the motivation effect by domain or item format.

The focus of the remainder of this section is on patterns of students' responses to the PISA tests over successive cycles with respect to the position of items in a booklet. A key observation that drove these analyses is the substantial increase in the percentages of missing responses displayed by students in Ireland in PISA 2009 relative to previous cycles (Cosgrove et al., 2010). These analyses were necessarily conducted on sub-samples of students. Sampling weights have not been applied, and sampling and measurement error (see OECD, 2009b) are not taken into account. As such, results should be treated as being broadly descriptive of response patterns over time.

It was necessary to identify a common set of items administered in a manner (sequence) similar enough to allow comparisons of responses across cycles. The PISA test design (see Table 2) is such that each student attempts a booklet consisting of four half-hour blocks, and, since 2003, the test design has been balanced, meaning that each block appears in each of the four positions.

In PISA 2000, the test design was not balanced: not all blocks appeared in all positions (see OECD, 2002, Chapter 2), which makes comparisons of position effects between 2000 and other cycles inherently problematic. Hence, comparisons for reading are confined to data from 2003 and 2009. For mathematics, comparisons are made between 2006 and 2009, since no intact mathematics blocks from 2003 were administered in 2006 or 2009. In science, intact blocks were not selected from 2006 to form the blocks used in 2009, so analyses consisted of comparing the same block *within* a cycle in positions 1 and 4.

Two caveats should be borne in mind. First, as already noted, it is difficult to disentangle the influences of proficiency (ability) and of effort or engagement in any analysis of student responses to a test without also measuring the latency of each response. Second, analyses are based on whether or not students responded to questions on PISA: we do not have a direct measure of the level of motivation or effort invested during the test. However, the PISA test administration is explicitly designed to allow sufficient time for students to respond to all or most questions

Table 2 PISA test design–2003, 2006, and 2009

Booklet	PISA 2003				PISA 2006				PISA 2009			
	P1	P2	P3	P4	P1	P2	P3	P4	P1	P2	P3	P4
1	M1	M2	M4	R1	S1	S2	S4	S7	M1	R1	R3A	M3
2	M2	M3	M5	R2	S2	S3	M3	R1	R1	S1	R4A	R7
3	M3	M4	M6	PS1	S3	S4	M4	M1	S1	R3A	M2	S3
4	M4	M5	M7	PS2	S4	M3	S5	M2	R3A	R4A	S2	R2
5	M5	M6	S1	M1	S5	S6	S7	S3	R4A	M2	R5	M1
6	M6	M7	S2	M2	S6	R2	R1	S4	R5	R6	R7	R3A
7	M7	S1	R1	M3	S7	R1	M2	M4	R6	M3	S3	R4A
8	S1	S2	R2	M4	M1	M2	S2	S6	R2	M1	S1	R6
9	S2	R1	PS1	M5	M2	S1	S3	R2	M2	S2	R6	R1
10	R1	R2	PS2	M6	M3	M4	S6	S1	S2	R5	M3	S1
11	R2	PS1	M1	M7	M4	S5	R2	S2	M3	R7	R2	M2
12	PS1	PS2	M2	S1	R1	M1	S1	S5	R7	S3	M1	S2
13	PS2	M1	M3	S2	R2	S7	M1	M3	S3	R2	R1	R5

Note. P1 = position 1, P2 = position 2, etc. M = mathematics, R = reading, S = science, PS = problem solving.

presented to them (e.g., OECD, 2011b). Hence, it is unlikely that students would have skipped items due to lack of time.

The analyses represent an attempt to examine two (possibly overlapping) potential explanations for the changes in achievement observed in Ireland: (i) the decline in PISA scores is due to a decrease in engagement (ii) the decline in PISA scores is due to a decrease in proficiency.

One would expect that, because of test fatigue, percent correct would generally be lower and the percent missing and not reached higher in position 4 relative to position 1 (cf. Borghans and Schils, 2011). One would also expect the response patterns for items in position 1 to be stable across cycles, all other things being equal. However, if the hypothesis about a decline in proficiency is to be supported, one would expect to see a decline in percent correct and a corresponding increase in percent missing/not reached in *both* positions across cycles. If the disengagement hypothesis is to be supported, one would expect stable percent correct and missing/not reached in position 1, but a decrease in percent correct (and an increase in missing responses) in position 4 across cycles[e]. The response patterns associated with the possibilities that are of interest are illustrated in Table 3.

Table 4 shows percent correct, incorrect, missing and not reached for block R2 in positions 1 and 4 in 2003 and 2009 for Ireland and the OECD[f]. (Results of comparisons between Ireland and the OECD averages only are presented here; comparisons with specific countries are described in Cosgrove [2011].) As would be expected due to test fatigue, the percent of correct responses is lower in position 4 than in position 1 in both cycles and in Ireland and across the OECD on average. The percentage of correct responses remained stable in position 1 both in Ireland and on average across the OECD. However, there is a marked decline in the percentage of correct responses for Ireland in position 4. This decline is not accompanied by an increase in incorrect responses. Rather, there has been an increase in both missing and not reached responses in this position. In contrast, percent incorrect, missing,

Table 3 Hypothesised response patterns associated with stable proficiency, a decline in engagement, a decline in proficiency, and test fatigue (example for reading, 2003 and 2009)

Stable achievement	P1 2009–P1 2003	P4 2009–P4 2003
Percent correct	No change	No change
Percent missing	No change	No change
Decline in engagement	**P1 2009–P1 2003**	**P4 2009–P4 2003**
Percent correct	No change	Decrease
Percent missing	No change	Increase
Decline in proficiency	**P1 2009–P1 2003**	**P4 2009–P4 2003**
Percent correct	Decrease	Decrease
Percent missing	Increase	Increase
Test fatigue	**P1 2003–P4 2003**	**P1 2009–P4 2009**
Percent correct	Decrease	Decrease
Percent missing	Increase	Increase

Note. P1 = position 1, P4 = position 4.

and not reached responses remained stable in 2003 and 2009 in position 4 across the OECD on average.

Cosgrove (2011) found that the decrease in percent correct in position 4 in Ireland was more marked for open response and multiple-choice items than for short response items. The decrease in percent correct on multiple-choice items was accompanied by an increase in the percentage of incorrect responses, while in the case of short and open response items, percent incorrect remained stable. In other words, fewer written response items were answered correctly in 2009 due to students skipping them, whereas fewer multiple-choice items were answered correctly in 2009 due to students responding to them incorrectly. This pattern suggests that students in Ireland were

Table 4 Average percent correct, incorrect, missing and not reached for block R2 (reading), positions 1 and 4, 2003 and 2009–Ireland and OECD averages[9]

	P1 2003	P1 2009	P4 2003	P4 2009
% correct				
Ireland	65.1	64.4	59.9	46.5
OECD	65.6	65.9	54.4	52.4
% incorrect				
Ireland	31.4	30.7	32.1	33.9
OECD	28.2	28.3	28.2	29.2
% missing				
Ireland	3.5	4.9	6.1	10.2
OECD	6.2	5.8	10.4	10.8
% not reached				
Ireland	0.0	0.0	1.9	9.4
OECD	0.0	0.0	7.0	7.6
% missing + not reached				
Ireland	3.5	4.9	7.9	19.6
OECD	6.3	5.8	17.5	18.4

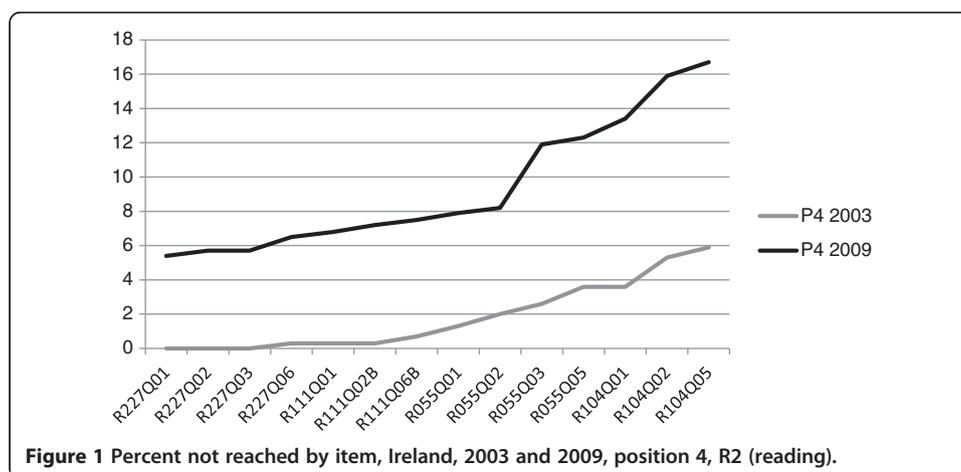Source: Cosgrove, 2011, Table twelve.

guessing the answers to multiple-choice items in position 4 to a greater degree in 2009 than in 2000.
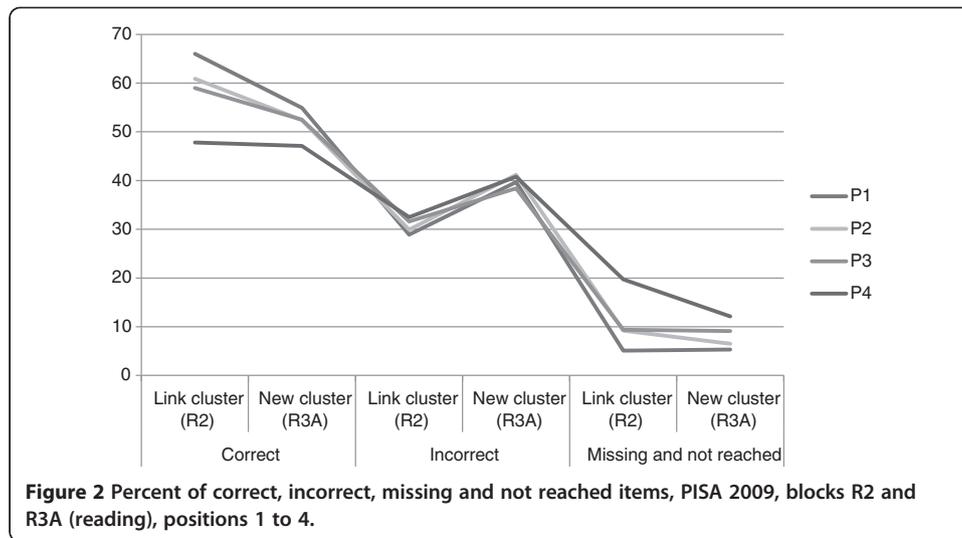
Figure 1 shows the percentage of not reached items in Ireland in both cycles for block R2, position 4 only. The rate of not reached items was much lower in 2003 than in 2009: about 6% of items at the end of the block were not reached in 2003, which is much lower than the equivalent figure for 2009 (about 17%).

Cosgrove (2011) also compared the response patterns for link and new items (i.e. items included in the assessment for the first time in 2009) (though she looked at just one of the four 'new' reading blocks). The hypothesised decline in student engagement, indicated by the position effects on item responses, was greater for the link items than for the new ones. Figure 2 shows the percentage of correct, incorrect, missing and not reached items for block R2 (a linking block) and R3A (a new block) across the four positions in the PISA 2009 test. While the percentage of incorrect responses remained stable for both blocks across all four positions, there is a more marked decline in percent correct, and a sharper increase in the rate of missing and not reached responses, for the block of link items relative to the new block.

Table 5 shows percentages correct, incorrect, missing and not reached for the same mathematics block (of two blocks in total) administered in 2006 and 2009 for Ireland and on average across the OECD. Similar to the results for reading, the percent correct is lower, and percent missing is higher, in position 4 relative to position 1 both in Ireland and across the OECD. However, in contrast to reading, there is a decline in percent correct in Ireland between 2006 and 2009 in both position 1 *and* position 4. Percent correct across the OECD on average is stable within position. The percentages of missing and not reached responses increased in position 4 in Ireland in 2009 relative to 2006 while the OECD averages remained stable. There is also a small increase in the percentage of missing responses for Ireland in position 1.

With respect to response type, in position 1, the largest decrease in percent correct in mathematics in Ireland was associated with short response items, then multiple-choice items, while there was no change in the percent correct in position 1 for longer written response items. In position 4, the change in percent correct by item type followed a slightly different pattern. Percent correct on all item types in this position

Figure 1 Percent not reached by item, Ireland, 2003 and 2009, position 4, R2 (reading).

Figure 2 Percent of correct, incorrect, missing and not reached items, PISA 2009, blocks R2 and R3A (reading), positions 1 to 4.

decreased from 2006, but was greatest for multiple-choice items, then short response items, followed by longer written response items (Cosgrove, 2011).
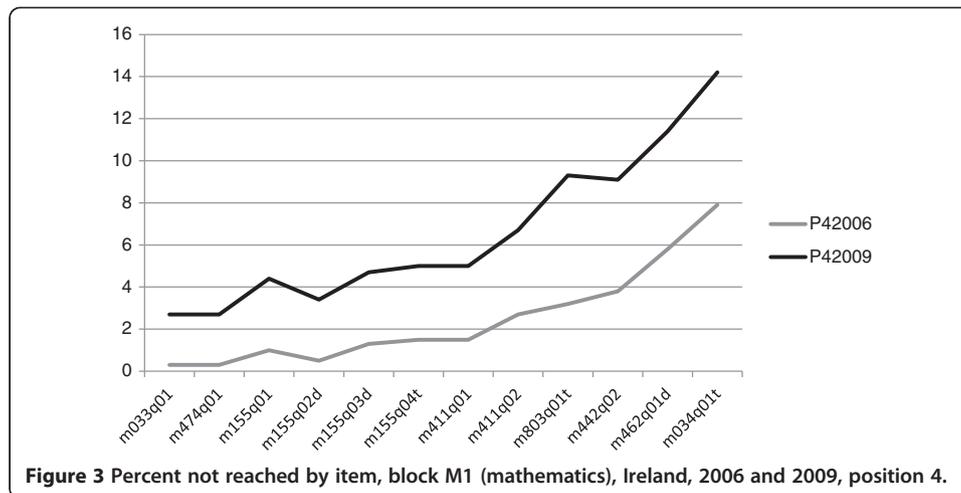
Figure 3 shows the percentage of not reached items in Ireland in both cycles for mathematics, in position 4 only. The data reveal a steady increase since 2006 in not reached items as students progressed through the block, but the differences between cycles are not as marked for mathematics as for reading (cf. Figure 1).

Table 6 shows the percent of correct, missing, and not reached responses for science blocks S1 (2009) and S4 (2006). It should be recalled that, unlike the previous analyses of mathematics and science, it was not possible to compare the same block across cycles.

Table 5 Average percent correct, incorrect, missing and not reached for block M1 (mathematics), positions 1 and 4, 2006 and 2009–Ireland and OECD averages

	P1 2003	P1 2009	P4 2003	P4 2009
% correct				
Ireland	52.7	49.5	49.2	44.5
OECD	51.1	51.5	45.7	46.1
% incorrect				
Ireland	40.4	41.5	39.4	36.9
OECD	38.8	38.9	37.9	37.4
% missing				
Ireland	6.9	8.7	8.9	12.0
OECD	10.1	9.6	12.2	12.2
% not reached				
Ireland	0.0	0.3	2.5	6.6
OECD	0.0	0.0	4.2	4.3
% missing + not reached				
Ireland	6.9	8.9	11.4	18.6
OECD	10.1	9.6	16.4	16.5

Source: Cosgrove, 2011, Table twenty-three.

Figure 3 Percent not reached by item, block M1 (mathematics), Ireland, 2006 and 2009, position 4.

In 2006, there was a 6.5% decline in percent correct across positions 1 and 4 in Ireland. In 2009, this decline was 8%. Across the OECD on average, the decline in percent correct in 2006 across positions 1 and 4 was 10%, and it was 11% in 2009. Thus, Ireland is not unusual in its decline in percent correct across positions. Similarly, the changes in the percentages of incorrect and missing responses across positions 1 and 4 in Ireland are comparable to the OECD averages in both years.

The percentage of not reached items in position 4 in Ireland in both 2006 and 2009 remained low, at about 2% in both cycles. This pattern contrasts with the percentages of not reached items in position 4 in reading and mathematics. The results suggest that students in Ireland remained more engaged in the science part of the assessment in 2009 when science items appeared at the end of the test booklet, compared to reading and mathematics.

Table 6 Average percent correct, missing and not reached for block S1/S4 (science), positions 1 and 4, 2006 and 2009–Ireland and OECD averages
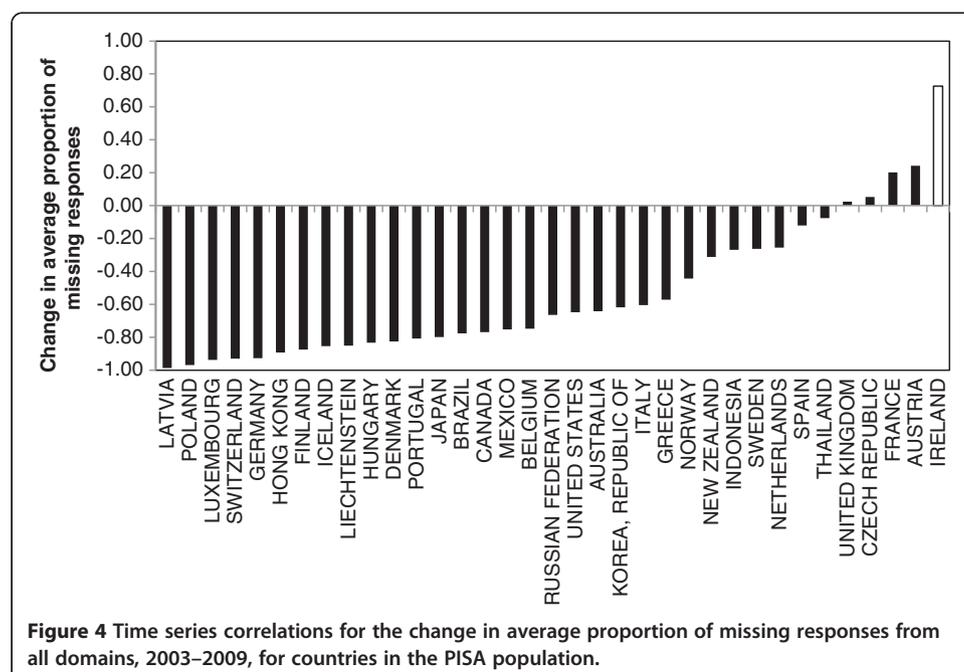
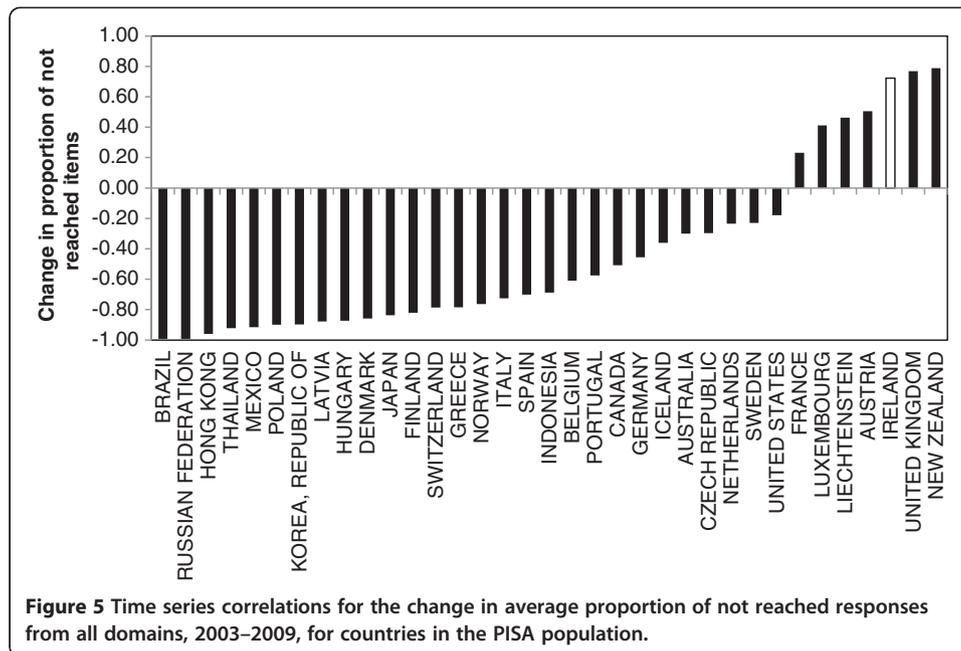|	P1 2003	P1 2009	P4 2003	P4 2009
% correct				
Ireland	63.8	57.3	62.0	54.0
OECD	59.8	50.9	64.3	53.1
% incorrect				
Ireland	34.6	38.1	35.3	39.0
OECD	37.0	41.5	31.7	39.3
% missing				
Ireland	1.6	4.6	2.7	7.0
OECD	3.2	7.6	4.0	7.6
% not reached				
Ireland	0.0	2.2	0.0	2.1
OECD	0.0	5.2	0.1	5.7
% missing + not reached				
Ireland	1.6	6.8	2.7	9.1
OECD	3.2	12.8	4.1	13.3

Source: Cosgrove, 2011, Table twenty-six.

Cartwright (2011) conducted an analysis of relationships between achievement and response patterns across countries and PISA cycles. He found, first, that (i) country-level correlations between missing (as opposed to not reached) responses are stronger for adjacent cycles and decrease with time, and (ii) correlations between the percentages of missing and not-reached responses at the country level are stronger between adjacent PISA cycles than they are with achievement within the same year. He commented: 'not only are non-response and test incompletion in PISA distinct from proficiency, they are also nationally distinctive characteristics that change over time' (p. 33). He argued that this strongly implies that test-taking behaviour in PISA is affected by country-specific features of the way in which PISA is implemented, which in turn is related to the amount of effort elicited from students. Second, on the basis of changes in the percentages of not reached items and percent correct scores at the country level across cycles, he concluded that 'changes in student effort have a large influence on changes in student performance' (p. 33).

These two findings, based on an analysis of response patterns internationally, are relevant to the case of Ireland, since changes in the average percentages of not reached items and missing responses are highly nationally idiosyncratic. While other countries, on average, have tended to show decreases in the percentages of missing and not-reached items in successive PISA cycles, percentages in Ireland have either remained stable or increased (Cartwright, 2011).

Figures 4 and 5 illustrate the extent to which Ireland may be considered idiosyncratic in this respect by displaying the results of time-series correlations that represent changes in average proportions of missing and not reached items, respectively. Ireland is unique among the countries examined in the consistency in the increase in missing responses over time, and is one of a small number of countries (along with Austria, France, Liechtenstein, Luxembourg, New Zealand, and the United Kingdom) that show consistent increases in not reached responses. These findings are of key importance to

Figure 4 Time series correlations for the change in average proportion of missing responses from all domains, 2003–2009, for countries in the PISA population.

Figure 5 Time series correlations for the change in average proportion of not reached responses from all domains, 2003–2009, for countries in the PISA population.

understanding the distinctive changes in the Irish PISA achievement scores, since they show that Ireland's response patterns are not only relatively unique among PISA countries, they are also related to changes in achievement over time. It is also relevant to note that Cosgrove and Moran (2011) found that students in Ireland in 2009 appear to have engaged much more on the digital than on the print reading assessment, though it is not clear from the analysis whether this is due to the mode of the assessment, assessment length, differences in response formats, and/or some other reason.

The analyses presented in this section provoke a question that is difficult to answer: if there had been a generalised decline in student engagement, why was it not evident in similar degrees in all three domains? This suggests that there are aspects of the PISA test questions that may be inherently more prone to declines in student engagement than others. It could equally be the case that there is something about the second-level curriculum in Ireland that is related to these differing patterns.

Discussion

This paper considered some potential reasons for the reported declines in Ireland's reading and mathematics scores in PISA 2009. It should be borne in mind that the benchmark against which achievement in 2009 was compared is itself problematic in that the booklet design for PISA 2000 was not balanced. Not all possible reasons have been explored in this review: for example, changes in curriculum; instruction time; the education system at primary level.

The task of disentangling methodological issues from ones which indicate substantive changes in proficiency is complex and, in a sense, the circumstances in which the Irish results for PISA 2009 have emerged represent the 'perfect storm'. It is unlikely that any country would experience such a confluence of confounding factors affecting the interpretation of performance trends (including those raised elsewhere, e.g. Gebhardt and Adams, 2007; Monseur and Berezner, 2007; Monseur, 2009; Mazzeo and von

Davier, 2008). In this respect, the case study of Ireland is useful in that the overall magnitude of the combined effect of these factors stimulated a detailed examination that might not have been undertaken if their random effects had reached a zero sum.

A sound policy response to PISA 2009 requires a proper understanding of what the results mean. Other than PISA itself, though, there are no other data against which to benchmark the 2009 results since, at present, there are no national standardised assessment data for second-level schools in Ireland (Department of Education and Skills, 2011; Perkins et al., 2012). While it is true that students sit national examinations at the end of lower and upper secondary education (the Junior and Leaving Certificates at Grades 9 and 12, respectively), results are not directly comparable from year to year.

No issues were detected in the national implementation procedures for PISA 2009, and Ireland met the requirements of the PISA 2009 technical standards (OECD, 2011b), as it has in all previous cycles. In contrast, some demographic changes were identified as potentially contributing to the decline in achievement. These include a substantial increase in the number of immigrant students (and changes in the socioeconomic composition of this group), slightly lower rates of early school leaving, and possible increases in the number of SEN students who have been integrated into mainstream education. However, while it is clear that changes in the demography of the school-going population have had some impact, it is difficult to quantify that impact, since these changes overlap in complex ways. Also, the emergence of the eight outlier (very low-achieving) schools in PISA 2009 remains difficult to explain. It may be that these schools existed in the system in 2000, but were not sampled due to chance, or that demographic and socioeconomic characteristics *and* very low engagement of students in these schools on the print assessment contributed to some of the decline in achievement in Ireland in 2009. The changes in the distribution of students across Grades 10 and 11 and the differential patterns of achievement decline across grades have prompted a review of teaching and learning, particularly in mathematics, at Grade 10 (Perkins et al., 2012).

Changes in demographics have occurred alongside changes in response patterns on the PISA tests. Results presented here indicate a general decline in engagement in the test in 2009, particularly reading, and this was most evident in items that required more effort. Thus, as Cartwright (2011) has pointed out:

> Even if there are true changes in student proficiency in Ireland, the role of student effort on changes in student performance is likely greater than that in other countries (p. 35).... Given the evidence suggesting that student effort does play a strong role in the PISA results for Ireland, particularly compared to other participating countries, any statements that interpret the PISA results beyond the context of the PISA test itself should be regarded with appropriate scientific scepticism (p. 40).

Students in Ireland showed a large fall-off in engagement with the PISA test in 2009, particularly on link reading items (Cosgrove, 2011). Existing research (e.g. Borghans and Schils, 2011) supports the view that engagement with a test such as PISA reflects individual characteristics that are distinguishable from cognitive proficiency and which also explains substantial variation in achievement across countries. It is suggested,

therefore, that student engagement in the PISA tests be researched further. One potentially fruitful avenue for this work could be through a comparison of students' responses on the paper-based and digital components of PISA. Unlike the paper and pencil tests, response latencies can be easily captured in the digital format and incorporated into data analysis. It may well be the case that cognitive proficiency and engagement with the tasks in PISA are as important as each other in understanding differences between countries and developing policies. At present, the OECD reports on and interprets the achievement results of PISA solely as estimates of cognitive proficiency. If the role of PISA as a benchmark for monitoring the relative progress of education systems continues to grow in prominence while at the same time becoming more sensitive to variations in student engagement, accurate reporting should incorporate response latencies in the estimation of either individual student estimates or aggregate performance.

Looking forward, these issues suggest that expanding the use of computer-based assessments, which can facilitate use of complete item response data including latency, has strong potential to improve the relevance of PISA and other international assessments to the policy-making process. This is certainly the case with PISA, where it is planned to transition, insofar as possible, to an entirely computer-based assessment of student achievement for the 2015 assessment.

In conclusion, we have attempted in earlier reviews as well as here to demonstrate that the PISA results for Ireland in 2009 represent a potentially valuable case study for other countries seeking to interpret unexpected changes in achievement in PISA or other international assessments over time. We have also raised some issues that merit consideration in the design and analysis of trend estimates in future cycles of PISA (as well as other large-scale assessments). The approach that is needed to understand change over time within a country, we suggest, is one which is multifaceted, taking into account changes within the design and administration of the assessment, the wider context in which that assessment is implemented, and possible interactions between the two.

It is tempting at times to go beyond the data to raise provocative possible reasons for Ireland's decline in 2009. Were students in this cohort in Ireland, who grew up in the midst of the "Celtic Tiger", complacent about the importance of investing effort in their education? Were they assuming that the jobs are there for the taking, regardless of whether or not they "flunked" in school? Obtaining scientific evidence for this possibility would seem an impossible task.

Endnotes

[a]PISA Quality Monitors observe the assessment in a subset of schools to verify that test administration and other procedures adhere to the international guidelines.

[b]This is a combined internationally-derived measure of socioeconomic status (see OECD, 2010b).

[c]Significant changes in this respect have been partially implemented in accordance with the Education of Persons with Special Educational Needs (EPSEN) Act (Government of Ireland, 2004).

[d]Transition Year is an optional one-year programme that can be taken after completing lower second-level (Grade 9) in Ireland. Approximately three-quarters of schools offer this programme, whose content and assessment is not centrally prescribed.

[e]Percent correct is the number of questions answered correctly out of the total number presented; percent incorrect is the number of questions answered incorrectly out of the total number presented; percent missing is the number of questions that were not answered by a student out of all items presented, but which have one or more valid responses (whether correct or incorrect) subsequent to the missed item; and percent not reached is the number of questions that were not answered by a student out of the total number presented, which were not followed by any subsequent valid responses.

[f]Block R2 is one of the two reading blocks (along with R1) that has been used to estimate trends in PISA since 2003. Responses to block R1 are described in Cosgrove (2011).

[g]The OECD average does not comprise the same set of countries across all cycles. Since 2006, Chile, Estonia, Israel, the Slovak Republic and Slovenia have joined. Average percentages were computed on the pooled OECD datasets which were weighted such that each country contributes equally to the averages.

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
JC drafted the parts of the manuscript concerning background information, analyses of response patterns concerning positioning effects, and conclusions. FC drafted the sections concerning international response patterns. The work as a whole represents extensive collaboration and discussion between JC and FC. Both authors read and approved the final manuscript.

Author details
[1]Educational Research Centre, St Patrick's College, Drumcondra, Dublin, IRELAND. [2]Polymetrika Inc., Ottowa, Ontario, CANADA.

References
Van Barneveld, C, Pharand, SL, Ruberto, L, & Haggarty, D. (2013). Student motivation in large-scale assessments. In M Simon,
    K Ercikan, & M Rousseau (Eds.), *Improving large-scale assessment in education: theory, issues and practice* (pp. 43–61). Oxfordshire: Routledge.
Boe, EE, May, H, & Boruch, RF. (2002). *Student task persistence in the Third International Mathematics and Science Study: a major source of achievement differences at the national, classroom, and student levels (Research Report No. 2002-TIMSS)*. Philadelphia: University of Pennsylvania, Graduate School of Education, Center for Research and Evaluation in Social Policy.
Borghans, L, & Schils, T. (2011). *The leaning tower of PISA: the effect of test motivation on scores in the international student assessment*. Paphos, Cyprus: Paper presented at the EALE annual conference. September 22–24.
Cartwright, F. (2011). *PISA in Ireland, 2000-2009: Factors affecting inferences about changes in student proficiency over time*. Dublin: Educational Research Centre.
Clerkin, A. (2013). Growth of the 'Transition Year' programme nationally and in schools serving disadvantaged students, 1992-2011. *Irish Educational Studies*. doi: 10.1080/03323315.2013.770663.
Cosgrove, J. (2011). *Does student engagement explain performance on PISA? Comparisons of response patterns on the PISA tests across time*. Dublin: Educational Research Centre.
Cosgrove, J, & Moran, G. (2011). *Taking the PISA 2009 test in Ireland: students' response patterns on the print and digital assessments*. Dublin: Educational Research Centre.
Cosgrove, J, Shiel, G, Archer, P, & Perkins, R. (2010). *Comparisons of performance in PISA 2000 to PISA 2009: A preliminary report to the Department of Education and Skills*. Dublin: Educational Research Centre.
Department of Education and Skills. (2011). *Literacy and numeracy for learning and life: the national strategy to improve literacy and numeracy among children and young people 2011-2020*. Dublin. Department of Education and Skills.
Eklöf, H. (2007). Test-taking motivation and mathematics performance in TIMSS 2003. *International Journal of Testing, 7*, 311–332.
Gebhardt, E, & Adams, RJ. (2007). The influence of equating methodology on reported trends in PISA. *Journal of Applied Measurement, 8*, 305–322.
Government of Ireland. (2004). *Education for persons with special educational needs act*. Dublin: Stationery Office.
LaRoche, S, & Cartwright, F. (2010). *Independent review of the PISA 2009 results for Ireland: Report prepared by the Educational Research Centre at the request of the Department of Education and Skills*. Dublin: Department of Education and Skills.

Mazzeo, J, & Von Davier, M. (2008). *Review of the Programme for International Student Assessment (PISA) test design: Recommendations for fostering stability in assessment results.* Paris: OECD Education Working Papers (EDU/PISA/GB(2008)28).

Monseur, C. (2009). *Item dependency in PISA.* Kiel, Germany: Paper presented at the PISA research conference. September 14-16.

Monseur, C, & Berezner, A. (2007). The computation of equating errors in international surveys of education. *Journal of Applied Measurement, 8*, 323–335.

OECD. (2002). *PISA 2000 technical report.* Paris. OECD.

OECD. (2009a). *PISA 2009 assessment framework: Key competencies in reading, mathematics and science.* Paris. OECD.

OECD. (2009b). *PISA data analysis manual: SPSS* (2nd ed.). Paris. OECD.

OECD. (2010a). *PISA 2009 results: what students know and can do–Student performance in reading, mathematics and science (Volume I).* Paris. OECD.

OECD. (2010b). *PISA 2009 results: Overcoming social background–Equity in learning opportunities and outcomes (Volume II).* Paris. OECD.

OECD. (2010c). *PISA 2009 results: Learning to learn–Student engagement, strategies and practices (Volume III).* Paris. OECD.

OECD. (2010d). *PISA 2009 results: Resources, policies and practices (Volume IV).* Paris. OECD.

OECD. (2010e). *PISA 2009 results: Learning trends–Changes in student performance since 2000 (Volume V).* Paris. OECD.

OECD. (2011a). *PISA 2009 results: Students on line–Digital technologies and performance (Volume VI).* Paris. OECD.

OECD. (2011b). *PISA 2009 technical report.* Paris. OECD.

Perkins, R, Cosgrove, J, Moran, G, & Shiel, G. (2012). *PISA 2009: Results for Ireland and changes since 2000.* Dublin: Educational Research Centre.

Shiel, G, Moran, G, Cosgrove, J, & Perkins, R. (2010). *A summary of the performance of students in Ireland on the PISA 2009 test of mathematical literacy and a comparison with performance in 2003: Report requested by the Department of Education and Skills.* Dublin: Educational Research Centre.

Wise, S, & DeMars, C. (2005). Low examinee effort in low-stakes assessment: problems and potential solutions. *Educational Assessment, 10*, 1–17.

Wise, S, & DeMars, C. (2010). Examinee noneffort and the validity of program assessment results. *Educational Assessment, 15*, 27–41.