The Irish Journal of Education, 2015, *xl*, pp. 45-67.

THE ROLE OF ENGAGEMENT AND TEST-TAKING BEHAVIOUR IN PISA 2012 IN IRELAND

Rachel Perkins
Educational Research Centre
St Patrick's College, Dublin

PISA 2012 data for Ireland were compared with data from previous cycles to explore patterns of non-response as students progressed throughout the test. Students were found to have invested more effort in 2012 than in 2009 in all domains (reading, mathematics, science). Responses to a national questionnaire which asked about their test-taking behaviour indicated that students who kept working on items that they did not know the answers to, who were more interested in the assessment, and whose concentration remained the same throughout the test obtained the highest mean scores across all domains. Significant differences were found between boys and girls and between students attending schools in the School Support Programme (SSP) under DEIS and students in non-SSP schools.

Ireland has participated in the OECD Programme for International Student Assessment (PISA) since it began in 2000. The assessment, which is conducted in three-yearly cycles, assesses the reading, mathematics, and science achievements of 15-year-old students. Sixty-five countries took part in the most recent (fifth) cycle in 2012. In each participating country, representative samples of schools and students are selected for testing. Participation by students in testing is voluntary and the tests are low-stakes for students.

In the first three cycles, results for Ireland were characterized by above average performance for reading, average performance for mathematics, and slightly above average performance for science. However, in the fourth cycle in 2009, results showed statistically significant declines of one-third of an international standard deviation in reading literacy, and of one-sixth of an international standard deviation in mathematics. Average science achievement remained stable. The most recent results, from 2012, show that the reading and mathematics achievements of students in Ireland were at the same levels as prior to 2009, while the average science score increased by about a seventh of a standard deviation.

A number of factors have been suggested as contributing to the declines in reading and mathematics performance observed in 2009, including demographic changes in the PISA cohort in Ireland since 2000, procedural changes in the PISA implementation process, survey fatigue among schools and students, the way the OECD measures change, and a decrease in student engagement with PISA tests over time (Cosgrove & Cartwright, 2014; Cosgrove, Shiel, Archer, & Perkins, 2010; Perkins, Cosgrove, Moran, & Shiel, 2012;). Cosgrove's (2015) paper in this volume also considers the declines in Ireland's reading and mathematics scores observed in 2009.

Other reasons for the decline in achievement have been explored by examining patterns of non-response to items as students progressed through the test. Since 2003, PISA has used a balanced booklet design. Each booklet consists of four half-hour blocks, with each block appearing in each of the four quartiles across booklets to control for the confounding effect of test fatigue on student responses. Cosgrove (2011) identified blocks that contained link items (i.e., items that are common across cycles and are used to estimate trends) for reading, mathematics, and science and described students' percent correct, incorrect, missing and not reached scores for these items at different quartiles in the booklets across cycles. Comparisons were made between 2003 and 2009 for reading, and between 2006 and 2009 for mathematics (since mathematics items appeared in the same combinations only in those years); in the case of science, two blocks were selected from 2006 and 2009 as the same items were not administered in intact blocks across these cycles.

For reading, the percentages of correct, incorrect, and missing responses were found to be stable when items were administered in quartile 1 in 2003 and 2009, indicating that the lower levels of overall performance in reading in 2009 could not be attributed to underperformance on the items appearing in quartile 1 (Cosgrove, 2011). As expected, the percentage of correct responses dropped between quartiles 1 and 4; however, the drop was larger in 2009 than in 2003 and was accompanied by a corresponding increase in the percentages of missing or not reached responses, suggesting that the decline in reading scores in 2009 could be attributed, at least in part, to a decline in effort. For mathematics, a small decline in percent correct scores was observed in quartile 1 between 2006 and 2009, while a larger decline was noted in quartile 4. The percentage of incorrect responses remained stable between 2006 and 2009, while there was an increase in the percentage of missing responses and not reached items. Thus, the findings for mathematics

suggest that declines in proficiency and effort contributed to the overall lower mathematics performance in 2009. Analysis of the science items indicated that the percentages of items answered correctly, incorrectly, that were missing, or were not reached were relatively stable between cycles; science appeared to be less prone to positioning effects than reading and mathematics, unless an actual increase in science proficiency compensated for a positioning effect that might otherwise have occurred.

For reading and mathematics in 2012, the percentage of items answered correctly increased, while the percentage of incorrect items remained fairly stable and the percentage of skipped items decreased (Perkins, Shiel, Merriman, Cosgrove, & Moran, 2013). In the case of science, an increase in the percentage of items answered correctly was noted, as well as a small decrease in the percentage of items answered incorrectly or skipped. This indicates that students invested more effort in the assessment in 2012 than in 2009, in particular in the reading and mathematics tests, reflecting an increase in engagement that probably contributed to the rise in scores observed in 2012.

National and international research has found that genders differ in their tendency to skip multiple-choice questions or to guess answers and has suggested that these differences may at least partially explain observed differences in performance on standardized tests (Baldiga, 2014; Bolger & Kellaghan, 1990). Girls were more likely than boys to skip multiple-choice questions, reflecting a greater tendency for boys to guess answers.

Due to the changes observed in students' response patterns to the assessment in 2009, it was decided to include a national questionnaire in Ireland in 2012 that asked students questions about their test-taking behaviour. The questionnaire was designed to gather information on how students engaged with different items in the assessment, their concentration and interest levels, how easy or difficult they found the assessment and how long it took them to complete it. This paper describes a study that extends the analyses conducted by Cosgrove (2011) by examining the pattern of Irish students' responses as they progressed through the 2012 assessment. It also describes the strategies used to complete the tests as described by students in the test-taking questionnaire.

METHOD

Procedure

Block Position. In PISA, test items are grouped into blocks which are domain specific (i.e., blocks are not made up of items from different domains). Blocks that are made up of link items (i.e., items that are common across cycles of PISA for the purpose of establishing trends) were identified for each domain to allow direct comparisons across cycles. Where possible, blocks that were used in the analyses conducted by Cosgrove (2011) were selected. For mathematics, block M2/M1[1] was selected and comparisons are made for 2006, 2009, and 2012, as this block was administered in the same format across these cycles. For science, the items administered in 2006 were redistributed into different blocks in 2009; thus direct comparisons cannot be made between these two cycles. The science blocks administered in 2009 were also administered in 2012, meaning that direct comparisons can be made between these two cycles. Block S1/S3[2] was selected for 2009 and 2012 and direct comparisons are made across these two cycles. Although direct comparisons cannot be made with 2006 and later cycles, block S4 was also selected for 2006 as this block was used in Cosgrove's analysis. Three new link blocks for reading, which were made up of new items from 2009, were also administered in 2012; comparisons can only be made between these two cycles for reading. One of these new blocks, R6/R2[3], was selected for analysis in this paper.

Since 2003, the assessment has been made up of 13 test booklets and each student has completed one booklet. Each booklet is divided into quartiles and each quartile is made up of one block of items. PISA is designed so that each block appears in each quartile across the test booklets (i.e., in the first quartile in one booklet, the second quartile in another booklet, etc.). Tables 1 to 3 present the PISA test designs for 2006, 2009, and 2012. The analysis in this paper presents students' response patterns for each selected block in different positions across booklets and cycles.

[1]Block M2 in 2006 was renamed M1 in 2009 and 2012
[2]Block S1 in 2009 was renamed S3 in 2012
[3]Block R6 in 2009 was renamed R2 in 2012

Table 1
PISA 2006 Test Design

Booklet	Q1	Q2	Q3	Q4
1	S1	S2	**S4**	S7
2	S2	S3	M3	R1
3	S3	**S4**	M4	M1
4	**S4**	M3	S5	**M2/M1***
5	S5	S6	S7	S3
6	S6	R2	R1	**S4**
7	S7	R1	**M2/M1***	M4
8	M1	**M2/M1***	S2	S6
9	**M2/M1***	S1	S3	R2
10	M3	M4	S6	S1
11	M4	S5	R2	S2
12	R1	M1	S1	S5
13	R2	S7	M1	M3

Note: Q1=quartile 1, Q2=quartile 2 etc. M=mathematics, R=reading and S=science. Blocks marked in bold are those selected for analysis.

*Block M2 was renamed M1 in 2009 and 2012

Table 2
PISA 2009 Test Design

Booklet	Q1	Q2	Q3	Q4
1	**M1**	R1	R3A	M3
2	R1	**S1/S3***	R4A	R7
3	**S1/S3***	R3A	M2	S3
4	R3A	R4A	S2	R2
5	R4A	M2	R5	**M1**
6	R5	**R6/R2***	R7	R3A
7	**R6/R2***	M3	S3	R4A
8	R2	**M1**	**S1/S3***	**R6/R2***
9	M2	S2	**R6/R2***	R1
10	S2	R5	M3	**S1/S3***
11	M3	R7	R2	M2
12	R7	S3	**M1**	S2
13	S3	R2	R1	R5

Note: Q1=quartile 1, Q2=quartile 2 etc. M=mathematics, R=reading and S=science. Blocks marked in bold are those selected for analysis.
*Block S1 in 2009 was renamed S3 in 2012.
*Block R6 in 2009 was renamed R3 in 2012.

Table 3
PISA 2012 Test Design

Booklet	Q1	Q2	Q3	Q4
1	M5	**S3**	M6A	S2
2	**S3**	R3	M7A	**R2**
3	R3	M6A	S1	M3
4	M6A	M7A	R1	M4
5	M7A	S1	**M1**	M5
6	**M1**	M2	**R2**	M6A
7	M2	S2	M3	M7A
8	S2	**R2**	M4	S1
9	**R2**	M3	M5	R1
10	M3	M4	**S3**	**M1**
11	M4	M5	R3	M2
12	S1	R1	M2	**S3**
13	R1	**M1**	S2	R3

Note: Q1=quartile 1, Q2=quartile 2 etc. M=mathematics, R=reading and S=science. Blocks marked in bold are those selected for analysis.

Test-Taking Behaviour Questionnaire. In 2012, after completing the PISA paper-based test, students in Ireland were asked to complete a nationally developed questionnaire of seven multiple-choice questions which asked about strategies for answering questions they did not know the answer to (multiple-choice and written response), if they had skipped questions they felt they had a good chance of getting right and why, how easy or difficult they had found the test, their level of interest in it, their level of concentration during the test, and how long it took them to complete it. All questions related to students' experience of the paper-based assessment and not the computer-based assessment, which a subset of students also completed. Of the 5,016 students who sat the test, 4,946 completed the test-taking behaviour questionnaire, giving a response rate of 98.6 percent.

Analysis

Block Position. The percentage of items answered correctly, incorrectly, missed, or not reached was computed for each student. Percent correct refers to the number of items answered correctly out of the total number presented to each student. Percent incorrect is the number of items answered incorrectly out of a total number presented to each student. Percent missing refers to the number of questions that were not answered by a student, out of all items

presented to each student, but which have one or more valid responses following the missed item. Finally, percent not reached is the number of questions that were not answered by a student out of the total number of questions presented, which were not followed by any valid responses. Not reached items are generally found at the end of test booklets (Cosgrove, 2011).

The average percent correct, incorrect, missing, and not reached was computed for each block selected for analysis in each quartile and compared across cycles. For ease of analysis, only averages for quartiles 1 and 4 are compared. Analysis was conducted using SPSS 18 and data were unweighted.

Test-Taking Behaviour Questionnaire. Students' responses to the test-taking behaviour questionnaire were linked to their test scores and other background information. Using SPSS 18, data were examined using frequency distributions and the mean reading, mathematics, and science scores of students who selected different responses were compared. Data were weighted and plausible values were used when computing mean achievement scores. The standard errors associated with mean achievement scores were computed using Fay's Balanced Repeated Replication (BRR), which takes into account the blocked nature of the sample (OECD, 2009).

Three indices (perceived easiness of the test, interest level in the test, and time taken to complete the test), each based on one item, were used for analyses, and each was set to have a mean of zero and a standard deviation of one. Correlational analysis was conducted between these indices and achievement scores. Comparisons were made between boys and girls and between students in schools in the School Support Programme (SSP) under DEIS and those in non-SSP schools on these indices. Statistical significance was established through creating 95% confidence intervals around each mean using the following formula: *Mean ± 1.96 standard errors*. The confidence interval is the range within which it is expected the population estimate would fall 95% of the time if many repeated samples were used.

RESULTS

Block Position Analysis

Table 4 presents data on the average percent correct, incorrect, missing, and not reached for students in Ireland and on average across OECD countries on the items that make up block R6/R2 in quartiles 1 and 4 in 2009

and 2012. For Ireland, the average percent correct, incorrect, missing, and not reached for quartile 1 is similar in the two years; however, there are some marked differences at quartile 4. In both cycles, the percentage of items answered correctly is lower in quartile 4 than in quartile 1. However, the drop in percent correct between quartiles 1 and 4 is greater in 2009 (-9.0%) than in 2012 (-5.3%). The percentage of items answered incorrectly is relatively stable between quartiles 1 and 4 in both cycles, with just a slight increase of 0.2 percentage points in 2009 and 0.3 percentage points in 2012. The percentage of items that were not answered (missing) or not reached is similar between cycles for quartile 1. However, the percentage of missing and not reached items in quartile 4 is considerably higher in 2009 (12.5%) than in 2012 (7.7%).

Table 4
Average Percent Correct, Incorrect, Missing, and Not Reached for Students in Ireland and on Average across OECD Countries for Block R2 (Reading) in Quartiles 1 and 4 (2009, 2012)

	Ireland			OECD		
	Quartile 1	Quartile 4	Diff (Q4-Q1)	Quartile 1	Quartile 2	Diff (Q4-Q1)
% correct						
2009 (R6)	66.9	57.9	-9.0	65.4	53.1	-12.3
2012 (R2)	65.8	60.5	-5.3	43.4	38.0	-5.4
% incorrect						
2009 (R6)	29.4	29.6	0.2	29.2	30.7	1.5
2012 (R2)	31.6	31.9	0.3	56.5	56.3	-0.2
% missing						
2009 (R6)	3.3	8.2	4.9	5.4	9.1	3.7
2012 (R2)	2.6	5.1	2.5	0.0	5.7	5.7
% not reached						
2009 (R6)	0.3	4.3	4.0	0.1	7.1	7.0
2012 (R2)	0.0	2.5	2.5	0.0	0.0	0.0
% missing & not reached						
2009 (R6)	3.7	12.5	8.8	5.4	16.2	10.8
2012 (R2)	2.6	7.7	5.1	0.0	5.7	5.7

As link items for mathematics were selected from intact blocks used in 2006, comparisons for block M2/M1 were made across three cycles (2006, 2009, 2012). These revealed a decrease in the percentage of items answered correctly in quartile 1 between 2006 and 2009 (from 47.1% to 44.4%) and a

subsequent increase in 2012 (to 46.7%) in Ireland (Table 5). In each cycle, the percentage of items answered correctly was lower in quartile 4 than in quartile 1; however, the decrease between quartiles was greater in 2009 (-5.0%) than in 2006 (-2.9%) and 2012 (-2.7%). The percentage of items answered incorrectly was similar across all cycles in quartile 1. However, there was a slight decrease in the percentage of items answered incorrectly between quartiles 1 and 4 and that decrease was greater in 2009 (-4.3%) than in 2006 (-1.2%) and 2012 (-1.6%).

Table 5
Average Percent Correct, Incorrect, Missing, and Not Reached for Students in Ireland and on Average across OECD Countries for Block M2/M1 (Mathematics) in Quartiles 1 and 4 (2006, 2009, 2012)

	Ireland			OECD		
	Quartile 1	Quartile 4	Diff (Q4-Q1)	Quartile 1	Quartile 2	Diff (Q4-Q1)
% correct						
2006 (M2)	47.1	44.2	-2.9	44.3	39.2	-5.1
2009 (M1)	44.4	39.4	-5.0	47.7	42.5	-5.2
2012 (M1)	46.7	44.0	-2.7	21.6	20.5	-1.1
% incorrect						
2006 (M2)	44.8	43.6	-1.2	44.8	42.4	-2.4
2009 (M1)	45.4	41.1	-4.3	42.0	40.7	-1.3
2012 (M1)	45.9	44.3	-1.6	78.3	78.6	0.3
% missing						
2006 (M2)	8.2	9.8	1.6	10.9	13.1	2.2
2009 (M1)	9.9	13.0	3.1	10.2	12.7	2.5
2012 (M1)	7.4	9.5	2.1	0.1	0.9	0.8
% not reached						
2006 (M2)	0.0	2.4	2.4	0	5.4	5.4
2009 (M1)	0.3	6.4	6.1	0.1	4.1	4.0
2012 (M1)	0.0	2.2	2.2	0.0	0.0	0.0
% missing & not reached						
2006 (M2)	8.2	12.2	4.0	10.9	18.4	7.5
2009 (M1)	10.2	19.5	9.3	10.3	16.8	6.5
2012 (M1)	7.4	11.7	4.3	0.1	0.9	0.8

The percentage of missing items in quartile 1 was somewhat greater in 2009 (9.9%) than in 2006 (8.2%) and 2012 (7.4%), suggesting somewhat higher levels of disengagement among students in 2009, even from the beginning of the assessment. There was an increase in the percentage of missed items between quartiles 1 and 4 across all cycles, but this increase was greater in 2009 (+3.1%) than in 2006 (+1.6%) and 2012 (+2.1%). Only a

very small percentage of items was not reached in quartile 1 in 2009 (0.3%), while all items were considered to have been 'reached' by students in quartile 1 in 2006 and 2012. There was also a marked increase in the percentage of not reached items between quartiles 1 and 4 in 2009 (+6.1%) compared to 2006 (+2.4%) and 2012 (+2.2%). Consequently, almost 20% of items were either missing or not reached in quartile 4 in 2009, compared to about 12% in 2006 and 2012.

Table 6 presents the average percent correct, incorrect, missing, and not reached for students in Ireland and on average across OECD countries on the items that made up block S1/S3[4] in quartiles 1 and 4 in 2009 and 2012, and for general comparison, on the items that made up block S4 in 2006. There was a small increase in the average percentage of items answered correctly in quartile 1 between 2009 and 2012 (from 63.0% to 65.6%). As expected, the percentage of items answered correctly decreased between quartile 1 and 4 in both cycles, although the decrease was greater in 2009 (-7.8%) than in 2012 (-6.2%). The average percentages of items answered correctly in quartiles 1 and 4 in 2006 were lower than in 2009 and 2012. There was a slight decrease (from 32.2% to 31.3%) in the average percentage of items answered incorrectly in quartile 1 between 2009 and 2012. The average percentage of items answered incorrectly remained relatively stable between quartiles 1 and 4 in 2009 and 2012, with just a small increase in 2009 (+1.3%) and a slight decrease in 2012 (-0.6%). The average percentages of items answered incorrectly in quartiles 1 and 4 in 2006 were considerably higher than in other cycles.

The average percentage of missing items at quartile 1 was somewhat higher in 2009 (4.8%) than in 2012 (3.1%). The average percentage of missing responses increased between quartiles 1 and 4 for both cycles; the increase was slightly greater in 2009 (4.2%) than in 2012 (3.6%). All items were 'reached' in quartile 1 in 2009 and 2012 and the increase in the average percentage of not reached items between quartiles 1 and 4 was greater in 2012 (+3.2%) than in 2009 (+2.3%). In 2006, the percentage of items that were missing or not reached was lower for both quartiles 1 and 4 than in 2009 or 2012.

[4] Block S1 in 2009 was renamed to S3 in 2012

Table 6
Average Percent Correct, Incorrect, Missing, and Not Reached for Students in Ireland and on Average across OECD Countries for Block S1/S3 (Science) in Quartiles 1 and 4 (2009, 2012) and S4 (Science) in Quartiles 1 and 4 in 2006

	Ireland			OECD		
	Quartile 1	Quartile 4	Diff (Q4-Q1)	Quartile 1	Quartile 2	Diff (Q4-Q1)
% correct						
2006 (S4)	60.9	54.0	-6.9	57.3	48.3	-9.0
2009 (S1)	63.0	55.2	-7.8	66.2	55.4	-10.8
2012 (S3)	65.6	59.4	-6.2	21.1	15.9	-5.2
% incorrect						
2006 (S4)	37.5	38.3	0.8	39.2	38.0	-1.2
2009 (S1)	32.2	33.5	1.3	27.9	29.7	1.8
2012 (S3)	31.3	30.7	-0.6	78.9	76.6	-2.3
% missing						
2006 (S4)	1.7	5.4	3.7	3.5	13.7	10.2
2009 (S1)	4.8	9.0	4.2	5.9	14.9	9.0
2012 (S3)	3.1	6.7	3.6	0.1	7.5	7.4
% not reached						
2006 (S4)	0.0	2.3	2.3	3.4	7.5	4.1
2009 (S1)	0.0	2.3	2.3	5.8	9.4	3.6
2012 (S3)	0.0	3.2	3.2	0.1	7.5	7.4
% missing & not reached						
2006 (S4)	1.7	7.7	6.0	0.0	6.2	6.2
2009 (S1)	4.8	11.3	6.5	0.1	5.4	5.3
2012 (S3)	3.1	9.9	6.8	0.0	0.0	0.0

Test-Taking Behaviour Analysis

There was a clear relationship between student achievement and test-taking behaviour. When students were asked to indicate the strategy they usually used if they did not know the answer to a multiple-choice question or a question to which they had to write an answer, those who reported that they kept on working until they thought they had the right answer achieved the highest scores in mathematics, reading, and science, significantly outperforming all other students (Table 7).

Table 7
Mean Mathematics, Reading, and Science Scores of Students, by Reported Test-Taking Behaviour (2012)

What did you usually do if you did not know the answer to a multiple-choice question on this test?		Mathematics		Reading		Science	
	%	Mean	SE	Mean	SE	Mean	SE
I skipped it and did not go back to it.	3.1	**407.3**	**9.24**	**420.5**	**10.5**	**412.4**	**10.21**
I skipped it and went back to it later.	20.0	**498.3**	**3.62**	**516.6**	**3.84**	**518.6**	**3.73**
I guessed the answer.	29.9	**478.2**	**2.86**	**501.3**	**3.34**	**496.1**	**3.28**
I kept working on it until I thought I had the right answer. (Reference category)	28.7	540.7	2.51	560.0	2.4	563.4	2.61
I did a mixture of these things.	18.3	**500.0**	**3.05**	**527.4**	**3.38**	**523.7**	**3.54**
What did you usually do if you did not know the answer to a question on this test where you had to write an answer?							
		Mathematics		Reading		Science	
	%	Mean	SE	Mean	SE	Mean	SE
I skipped it and did not go back to it.	10.6	**458.0**	**4.05**	**477.4**	**4.52**	**473.5**	**4.45**
I skipped it and went back to it later.	26.1	**505.8**	**2.95**	**527.0**	**3.05**	**528.8**	**3.03**
I guessed the answer.	24.3	**482.9**	**3.36**	**506.9**	**3.49**	**500.2**	**3.59**
I kept working on it until I thought I had the right answer. (Reference category)	24.6	536.5	2.67	555.8	2.79	559.5	3.14
I did a mixture of these things.	14.4	**501.6**	**3.51**	**524.9**	**4.22**	**522.3**	**4.12**

Note: significant differences are in bold.

Students who said that they skipped the question and did not go back obtained the lowest mean scores for each domain (for both multiple-choice and written response questions), followed by those who said that they guessed the answer. For written response questions, students who reported going back to questions that they had skipped had the second highest mean scores for each domain, while for multiple-choice questions; students who performed a mixture of things had the second highest achievement scores.

Significantly more boys (30.8%) than girls (26.5%) reported that they kept working until they thought they had the right answer for multiple-choice

questions (Table 8). For written response questions, the percentages were 26.7 for boys and 22.5 for girls. On the other hand, significantly more girls (26.7%) than boys (22.0%) reported guessing the answer to written response questions and doing a mixture of things for multiple-choice questions (20.5% and 16.1%, respectively).

Table 8
Percentages of Boys and Girls and of Students in SSP and non-SSP Schools who Reported Engaging in Various Strategies when they Did Not Know the Answer to Multiple-Choice or Written Response Questions (2012)

What did you usually do if you did not know the answer to a multiple-choice question on this test?	Boys (Ref Category)		Girls		Non-SSP (Ref Category)		SSP	
	%	SE	%	SE	%	SE	%	SE
I skipped it and did not go back to it.	3.6	.46	2.6	.35	2.3	.27	**6.2**	**1.09**
I skipped it and went back to it later.	20.5	.95	19.6	.78	19.8	.65	20.9	1.61
I guessed the answer.	29.1	1.17	30.8	.86	28.9	.87	**33.8**	**1.71**
I kept working on it until I thought I had the right answer.	30.8	1.15	**26.5**	**1.10**	30.7	.93	**20.9**	**1.51**
I did a mixture of these things.	16.1	.84	**20.5**	**.75**	18.3	.68	18.3	1.24
What did you usually do if you did not know the answer to a question on this test where you had to write an answer?	Boys (Ref Category)		Girls		Non-SSP (Ref Category)		SSP	
	%	SE	%	SE	%	SE	%	SE
I skipped it and did not go back to it.	11.1	.73	10.0	.67	9.3	.50	**15.5**	**1.36**
I skipped it and went back to it later.	26.7	.97	25.4	.85	25.5	.71	28.2	1.89
I guessed the answer.	22.0	.91	**26.7**	**.85**	24.4	.66	23.9	1.65
I kept working on it until I thought I had the right answer.	26.7	1.02	**22.5**	**.80**	26.4	.65	**18.0**	**1.33**
I did a mixture of these things.	13.4	.9	15.4	.78	14.4	.59	14.3	1.26

Note: significant differences are in bold.

About a fifth (20.9%) of students in schools in the School Support Program (SSP) under DEIS reported that they kept working on multiple-choice questions until they felt they had the right answer (Table 8). This is

significantly lower than the percentage for non-SSP schools (30.7%). On the other hand, significantly more students in SSP schools (6.2%) than in non-SSP schools (2.3%) reported that they skipped questions they did not know the answer to and did not go back to them or that they guessed the answer to questions (33.8% and 28.9%, respectively).

Just over 15% of students nationally said that they skipped questions that they felt they had a good chance of getting right; these students performed significantly less well on all domains than students who did not skip such questions (Table 9). Of those that skipped these questions, 28.2% reported that they did so because they ran out of time. These students obtained significantly higher mean scores across all domains than students who reported that they felt it was not worth the effort (23.2%) and those who said they skipped questions because of a mixture of these two things (25.7%). Almost 23% of students indicated that they skipped questions for another reason (but did not cite these reasons). The performance of these students, although lower in all three domains, did not differ significantly from that of students who said they ran out of time.

Table 9
Mean Mathematics, Reading, and Science Scores of Students who Reported Skipping a Question that they Felt they had a Good Chance of Getting Right (2012)

Did you skip any questions that you felt you had a good chance of getting right?							
		Mathematics		Reading		Science	
	%	Mean	SE	Mean	SE	Mean	SE
Yes	15.3	**450.6**	**3.77**	**468.0**	**4.08**	**464.4**	**4.02**
No (Reference category)	84.7	512.2	2.06	534.5	2.34	533.9	2.29
If you answered 'Yes', was this because:							
		Mathematics		Reading		Science	
	%	Mean	SE	Mean	SE	Mean	SE
You ran out of time. (Reference category)	28.2	473.2	7.34	491.1	8.4	486.9	8.52
You felt it was not worth the effort.	23.2	**442.9**	**7.8**	**453.6**	**9.28**	**452.5**	**9.58**
A mixture of these two things.	25.7	**435.2**	**5.44**	**454.5**	**5.96**	**451.6**	**6.17**
Another reason.	22.9	448.5	7.41	469.8	7.24	464.6	7.92

Note: significant differences are in bold.

A significantly greater percentage of boys (17.4%) than of girls (13.2%) in the national sample and of students in SSP schools (24.0%) than of students in non-SSP schools (13.0%) reported that they skipped questions that they felt they had a good chance of getting right (Table 10). Significantly more boys (27.1%) than girls (17.8%) reported that they skipped these questions because they felt it was not worth the effort, while a significantly greater percentage of girls (28.3%) than of boys (19.0%) said they did so for 'another reason'. There are no statistically significant differences in achievement scores across domains between students in SSP and non-SSP schools in the reasons they gave for skipping questions they felt they had a good chance of getting right.

Table 10
Percentages of Boys and Girls and of Students in SSP and non-SSP schools who Reported that they Skipped a Question they Felt they had a Good Chance of Getting Right (2012)

Did you skip any questions that you felt you had a good chance of getting right?								
	Boys (Ref Category)		Girls		Non-SSP (Ref Category)		SSP	
	%	SE	%	SE	%	SE	%	SE
Yes	17.4	.83	13.2	.74	13.0	.64	24.0	1.72
No	82.6	.83	86.8	.74	87.0	.64	76.0	1.72
If you answered 'Yes', was this because:								
	Boys (Ref Category)		Girls		Non-SSP (Ref Category)		SSP	
	%	SE	%	SE	%	SE	%	SE
You ran out of time.	26.7	2.52	30.1	2.68	30.4	2.13	23.4	3.50
You felt it was not worth the effort.	27.1	2.43	17.8	1.99	23.5	1.74	22.5	3.23
A mixture of these two things.	27.1	2.16	23.8	2.99	24.6	1.98	28.0	3.63
Another reason.	19.0	2.13	28.3	3.22	21.5	2.15	26.1	3.77

Note: significant differences are in bold.

Table 11 presents the average percentage of items that were not answered by students (missed or not reached) for each domain and the reasons given for skipping questions that they felt they had a good chance of getting right.

Table 11
Mean Percentage of Items that were Missed or Skipped by Students for Each Domain, by the Reasons Given for Skipping Questions that Students felt they had a Good Chance of Getting Right

	Mathematics Mean %	Reading Mean %	Science Mean %
You ran out of time.	12.5	10.3	17.2
You felt it was not worth the effort.	23.9	13.6	15.5
A mixture of these two things.	17.3	10.9	10.9
Another reason.	17.5	9.1	14.6

For mathematics and reading, students who felt it was not worth the effort had the highest average percentage of missing and not reached items (23.9% for mathematics and 13.6% for reading), while for science the highest average percentage of missing or not reached items was found for students who skipped such items because they ran out of time (17.2%). This suggests that students were less engaged with the mathematics and reading items than with the science items.

When students were asked to rate their concentration during the test, just over 40% indicated that their concentration levels remained the same throughout the test. These students had significantly higher mathematics, reading, and science scores than students who said their concentration was better at the start than at the end, and students who said it varied a lot (Table 12).

Table 12
Mean Mathematics, Reading and Science Scores of Students, by Students' Reported Concentration Levels During the Test

How was your concentration during the test:	%	Mathematics Mean	Mathematics SE	Reading Mean	Reading SE	Science Mean	Science SE
It was the same throughout the test. (Reference category)	43.1	514.1	3.03	535.2	3.2	535.9	3.23
It was better at the start than at the end.	33.0	**493.9**	**2.77**	**515.7**	**3.03**	**513.5**	**3.01**
It varied a lot.	23.9	**493.7**	**2.74**	**516.7**	**3.2**	**512.8**	**3.08**

Note: significant differences are in bold.

Students who said their concentration was better at the start obtained achievement scores across all domains that were similar to those of students who said their concentration varied a lot throughout the test. There are no significant differences between boys and girls in the national sample, or between students in SSP and non-SSP schools, in their ratings of their concentration levels throughout the test (Table 13).

Table 13
Percentages of Boys and Girls, and of Students in SSP and non-SSP schools by Perceived Level of Concentration During the Test (2012)

How was your concentration during the test?	Boys (Ref Category)		Girls		Non-SSP (Ref Category)		SSP	
	%	SE	%	SE	%	SE	%	SE
It was the same throughout the test.	43.5	.99	42.7	.97	43.0	.85	43.4	1.65
It was better at the start than at the end.	33.5	1.07	32.4	1.05	32.5	.83	34.7	1.67
It varied a lot.	23.0	1.02	25.0	.91	24.5	.82	21.9	1.28

Just over a fifth (21.6%) of students in the national sample reported that they found the PISA test easy, while 62.9% said they found it to be 'ok' (Table 14). The remaining students found it to be either difficult (12.8%) or very difficult (2.6%). Boys (27.9%) were significantly more likely than girls (15.2%) to rate the test as easy, as were students in non-SSP schools (23.2%) compared to those in SSP schools (15.6%).

Table 14
Percentages of Boys and Girls, and of Students in SSP and non-SSP schools by Perceived Easiness of Test (2012)

How easy or difficult did you find this test?	All		Boys (Ref Category)		Girls		Non-SSP (Ref Category)		SSP	
	%	SE	%	SE	%	SE	%	SE	%	SE
Easy	21.6	.72	27.9	1.09	**15.2**	**.92**	23.2	.82	**15.6**	**1.91**
Ok	62.9	.77	59.5	1.15	**66.3**	**1.05**	63.8	.81	**59.6**	**1.97**
Difficult	12.8	.57	10.4	.76	**15.4**	**.87**	11.1	.60	**19.6**	**1.46**
Very difficult	2.6	.33	2.2	.47	3.1	.42	2.0	.24	**5.1**	**1.21**

Note: significant differences are in bold.

Correlations between perceived easiness of the test and achievement in all three domains were higher than correlations between achievement and interest level or time taken to complete the test (Table 15). All the correlations are statistically significant.

Table 15
Correlations between Perceived Easiness of the Test, Interest Level in the Test, Time Taken to Complete the Test, and Mathematics, Reading and Science Scores

	Perceived easiness		Interest level		Time taken	
	r	SE	r	SE	r	SE
Mathematics	**0.460**	**0.01**	**0.213**	**0.02**	**0.136**	**0.02**
Reading	**0.400**	**0.02**	**0.213**	**0.02**	**0.161**	**0.02**
Science	**0.443**	**0.01**	**0.220**	**0.02**	**0.152**	**0.02**

Most students indicated they were very interested (14.1%) or quite interested (45.8%) in the test, while a substantial minority were not very interested (31.8%) or not interested at all (8.2%) (Table 16). Boys and girls did not differ significantly from each other in terms of their interest level, while students in SSP schools reported significantly lower levels of interest than students in non-SSP schools.

Table 16
Percentages of Boys and Girls in National Sample, and of Students in SSP and non-SSP schools by Perceived Level of Interest in the Test (2012)

How did you feel when you were taking this test?	All		Boys (Ref Category)		Girls		Non-SSP (Ref Category)		SSP	
	%	SE	%	SE	%	SE	%	SE	%	SE
Very interested	14.1	.60	14.7	.81	13.5	.88	14.6	.63	12.4	1.34
Quite interested	45.8	1.01	44.2	1.47	47.6	1.22	47.6	1.02	**39.3**	**2.51**
Not very interested	31.8	.92	31.3	1.30	32.3	1.15	30.6	.96	**36.5**	**2.10**
Not interested at all	8.2	.49	9.8	.78	**6.6**	**.54**	7.3	.47	**11.8**	**1.39**

Note: significant differences are in bold.

Students were allocated two hours in which to complete the PISA paper-based assessment. The vast majority (84.1%) of students in the national sample indicated that they had finished the test within one-and-a-half hours, while 16% said that it took them the full two hours (Table 17). Students in SSP schools did not differ significantly from students in non-SSP schools in the average length of time it took them to complete the test. There was, however, a significant gender difference, with girls significantly more likely than boys to report finishing the test in around an hour and a half (57.8% and 54.0%, respectively), or around two hours (18.1% and 13.9%, respectively).

Table 17
Percentages of Boys and Girls, and of Students in SSP and non-SSP schools by Length of Time Taken to Complete the Test (2012)

About how long did it take you to finish this test?	All		Boys (Ref Category)		Girls		Non-SSP (Ref Category)		SSP	
	%	SE	%	SE	%	SE	%	SE	%	SE
Half an hour or less	2.4	.33	2.9	.50	1.9	.34	1.9	.23	4.4	1.28
Around an hour	25.8	1.07	29.1	1.54	**22.3**	**1.23**	25.4	1.14	27.3	2.60
Around an hour and a half	55.9	.91	54.0	1.42	**57.8**	**1.00**	56.7	1.00	52.6	2.42
Around two hours	16.0	.83	13.9	1.10	**18.1**	**1.11**	16.0	.96	15.7	1.59

Note: significant differences are in bold.

Students who found the test to be more difficult were more likely to miss or skip items (Table 18). Students who indicated that they had less interest in the test were also more likely to skip items (Table 19), indicating that students' motivation to take part is linked to their achievement levels. Students who indicated that their concentration was better at the beginning of the test than at the end were also more likely to skip items than students who indicated that their concentration was the same throughout the test (Table 20).

Table 18
Mean Percentage of Items that were Missed or Skipped by Students for Each Domain, by Perceived Difficulty Level of the Test

	Mathematics Mean %	Reading Mean %	Science Mean %
Easy	4.0	1.9	2.4
OK	8.1	4.6	5.5
Difficult	15.6	9.6	10.0
Very difficult	21.4	18.7	21.5

Table 19
Mean Percentage of Items that were Missed or Skipped by Students for Each Domain, by Interest Level in the Test

	Mathematics Mean %	Reading Mean %	Science Mean %
Very interested	5.1	2.7	4.4
Quite interested	7.4	3.6	5.4
Not very interested	10.8	6.3	6.1
Not interested at all	15.5	11.0	9.9

Table 20
Mean Percentage of Items that were Missed or Skipped by Students for Each Domain, by Perceived Concentration Level During the Test

	Mathematics Mean %	Reading Mean %	Science Mean %
It was the same throughout the test.	6.9	4.1	5.3
It was better at the start than at the end.	9.5	5.0	6.3
It varied a lot.	10.9	5.9	6.4

Students who finished the test within half-an-hour or less had the highest average percentage of skipped items, while students who finished the test in about an hour-and-a-half were least likely to skip items (Table 21).

Table 21
Mean Percentage of Items that were Missed or Skipped by Students for Each Domain, by Time Taken to Complete the Test

	Mathematics Mean %	Reading Mean %	Science Mean %
Half an hour or less	23.7	12.4	24.3
About an hour	10.3	5.5	6.5
About an hour and a half	7.3	4.0	4.0
About two hours	8.8	6.0	7.4

CONCLUSION

The results of the block position analysis show that Irish students displayed greater levels of engagement with the PISA assessment in 2012 than in 2009 and that, in the case of mathematics, student engagement in 2012 was at the same level as in 2006. Thus, it would seem that students invested less effort in the 2009 assessment than in other cycles and that this contributed to the decline in reading and mathematics scores observed in that year. The improvement observed in the mean science scores for Ireland between 2009 and 2012 seems to be down to both an improvement in proficiency and greater levels of effort invested by students in the assessment.

Findings from the test-taking behaviour questionnaire demonstrated the relationship between students' interaction with the test and their achievement scores. Students who kept working on questions they did not know the answer to obtained the highest scores, while those who skipped these questions and did not go back to them had the lowest scores. Furthermore, students who indicated that they skipped questions they felt they had a good chance of getting right had significantly lower achievement levels across domains and, of these students, those who did so because they felt it was not worth the effort performed least well. The large percentage of students who indicated that they skipped mathematics items because they felt they were not worth the effort (24%) compared to reading (14%) and science (16%) is of particular concern. Students' level of concentration and of interest and time taken to complete the test were also found to be significantly related to achievement. These findings point towards an important relationship between the efforts invested in the test by students and their achievement scores.

Significant differences were observed between boys and girls and between students in SSP and non-SSP schools. For example, boys were more likely to report that they kept working on questions that they didn't know the answer to until they thought they had the right answer for all question types, while girls were more likely to report guessing the answer for written response questions. Boys were also more likely to report skipping questions that they felt they had a good chance of getting right, and were more likely to do so because they felt it was not worth the effort. Students in SSP schools were more likely to report skipping questions that they felt they had a good chance of getting right and to report not returning to questions that they skipped because they did not know the answer; however, these students

reported significantly lower levels of interest in the test. These findings indicate the need for different approaches to increasing engagement among different groups of students.

Finally, we may consider other reasons why student engagement with the PISA study might have varied across cycles. One such factor is survey fatigue. A number of other large-scale studies were also carried out in post-primary schools at around the same time as PISA 2009, including the International Civics and Citizenship Study (ICCS; in 2009) and the Teaching and Learning International Survey (TALIS; in 2008). Given the limited population of post-primary schools in Ireland, it is likely that some schools were selected to participate in more than one survey. Other factors, such as a change from using the Department of Education and Skills inspectorate to administer the PISA tests to using teachers in students' schools in 2009 could have compounded the effect of survey fatigue. The fact that PISA 2009 results received widespread attention in Ireland may have led to greater awareness of the study among schools, which, in turn, may have increased student effort in 2012.

The findings of our study point towards a need for further examination of the effect of motivation or effort invested by students on test scores in low-stakes situations, especially in the context of measuring trends when changes in achievement levels need to be explained. Future research could draw on the work of Boe, May, and Boruch (2002), by examining the relationship between how students respond to background questionnaire items in PISA and student achievement. Levels of student engagement in assessments such as PISA should also be monitored.

REFERENCES

Baldiga, K. (2014). Gender differences in willingness to guess. *Management Science, 60*, 434-448.

Boe, E.E., May, H., & Boruch, R.F. (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, PA: University of Pennsylvania, Graduate School of Education, Center for Research and Evaluation in Social Policy.

Bolger, N., & Kellaghan, T. (1990). Method of measurement and gender differences in scholastic achievement. *Journal of Educational Measurement, 27*, 165-174.

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. (2015). Changes in achievement in PISA from 2000 to 2009 in Ireland: Beyond the test scores. *Irish Journal of Education, 40*, 29-44.

Cosgrove, J., & Cartwright, F. (2014). Changes in achievement on PISA: The case of Ireland and implications for international assessment practice. *Large-scale Assessments in Education, 2,* 2. Retrieved September 19, 2014, from *http://www.largescaleassessmentsineducation.com/content/2/1/2*

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.

OECD (Organisation for Economic Co-operation and Development). (2009). *PISA data analysis manual: SPSS (2nd ed.)*. Paris: Author.

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

Perkins, R., Shiel, G., Merriman, B., Cosgrove, J., & Moran, G. (2013). *Learning for life: The achievements of 15-year-olds in Ireland on mathematics, reading literacy and science in PISA 2012*. Dublin: Educational Research Centre.