Iris Eireannach An
OIDEACHAIS
The Irish Journal of
EDUCATION

Exploring the implications of moving the PISA testing period from Spring to Autumn: A multilevel analysis using achievement data in Ireland

Sylvia Denner[1]*, Michael O'Leary[2], and Gerry Shiel[1]

[1] *Educational Research Centre, Dublin*
[2] *Centre for Assessment Research, Policy and Practice in Education (CARPE), Institute of Education, Dublin City University*

Abstract

The Programme for International Student Assessment (PISA) assesses the performance of 15-year-old students in reading, mathematics, and science. In Ireland, information from PISA is highly valued by the Department of Education as a guide to inform policy-making. Between 2000, when Ireland first participated in PISA, and 2018, testing has taken place in the Spring (March/April). Due to school-based preparation activities for state examinations in June, the springtime is a particularly busy time in schools for nearly three fifths of PISA-eligible students who are in their third year of post-primary education. This issue caused policy makers in Ireland to consider changing PISA testing time from the Spring to the Autumn and, consequently, research examining the outcomes of such a move became a priority. Using multilevel modelling, this paper examines the performances of those 15-year-olds in Ireland who participated in the regular PISA cycle in Spring 2018 and another sample that completed the tests in the Autumn of 2018. Results indicated that, overall, there was no statistically significant difference between the two samples in reading literacy, mathematics, or science across the two time periods. However, while controlling for relevant demographic background variables, performance in mathematics in the Autumn was lower for males than in Spring. Of particular note was the statistically significantly higher percentage of low-achieving males in the Autumn administration of the tests (21.4%) compared to the Spring administration (15.7%). The implications of the findings for policy and practice within the Irish and international educational contexts are discussed.

Keywords: PISA, Ireland, Time of year testing, Multilevel model

Author Note

*Sylvia Denner can be contacted at sylvia.denner@erc.ie or at the Educational Research Centre (ERC), DCU St Patrick's College Campus, Dublin 9, D09AN2F, Ireland

Please cite as: Denner, S., O'Leary, M., & Shiel, G. (2025). Exploring the implications of moving the PISA testing period from Spring to Autumn: A multilevel analysis using achievement data in Ireland. *Irish Journal of Education*, *49*(2), 1-36. www.erc.ie/ije

Introduction

The Programme for International Student Assessment (PISA) is an initiative of the Organisation for Economic Co-Operation and Development (OECD). The primary functions of PISA are to provide international comparative data on students' knowledge and skills near the end of compulsory schooling, a knowledge base for policy analysis and research, and trend indicators showing how study outcomes change over time (OECD, 1999). In Ireland, as in many other countries, results from international large-scale assessments are often used as key performance indicators.

There was an increased focus on PISA results in Ireland when the performance of students in Ireland declined significantly in reading literacy and mathematics in PISA 2009 (Cosgrove & Hislop, 2011; Perkins, 2015; Cosgrove & Cartwright, 2014; Cosgrove, 2015). This helped to spur the implementation of a *National Strategy to Improve Literacy and Numeracy 2011-2020* (Department of Education and Skills [DES], 2011; 2017a). There was a very positive response from the Irish public to the consultation process in developing the Strategy (Hislop, 2011) which included almost 480 written submissions from individuals and representation of over 60 organisations, demonstrating the commitment of people which went beyond the educational system. The Strategy set out an integrated approach to improving standards across all the phases of education from early childhood to the end of the post-primary cycle. The Strategy made specific reference to the PISA 2009 results and the Department of Education (DoE) established nationwide targets in relation to mathematics and reading based on the PISA proficiency levels. These targets were revised in 2017, and additional targets were established for socioeconomically disadvantaged post-primary schools in the DEIS (Delivering Equality of Opportunity in Schools) scheme, with an aim to reduce the achievement gap between these schools and their non-DEIS counterparts based on PISA results (DES, 2017b). Outcomes from PISA surveys have also had an influence on other policy areas such as the development of the *Science, Technology, Engineering and Mathematics (STEM) Education Policy Statement* in 2017 (DES, 2017c) and a *Wellbeing Policy Statement and Framework for Practice 2018–2023 (DES, 2019)*.

The objective of this study is to provide in-depth analysis of students' achievement on the PISA test when testing is administered in Spring compared to Autumn. In doing this, the aim is to go beyond bivariate analysis, as was done in Denner (2023), and examine 'time of testing' while controlling for student and school-level background variables as well as other variables that have a strong association with achievement in PISA.

The first section of the paper provides context for the study. This is done by presenting a brief review of PISA in Ireland and the pressures in schools in the springtime, which led to a Feasibility Study being carried out to examine the possibility of changing the testing period in Ireland from Spring to Autumn. This is followed by a section providing an overview of prior research on changes to the PISA testing period and details of the methodology and data used. A summary of descriptive statistics comparing the Spring and Autumn cohorts for both background variables and variables associated with performance on PISA is then outlined. The main section of the paper is focused on outcomes from multilevel modelling used to examine effects related to 'time of year' testing in PISA while controlling for a variety of variables associated with achievement in PISA. The paper concludes with a discussion of the implications of the findings for policy and practice within the Irish and international contexts.

PISA in Ireland

Ireland has participated in PISA since the first cycle in 2000. Including 2018, the assessment was carried out in the Spring of the relevant year, during a six-week window in the months of March/April. This is a particularly busy time of year in schools, with most of the PISA-eligible students in Third Year of post-primary education (McKeown et al., 2019). Post-primary level consists of a three-year Junior Cycle programme followed by a two-year or three-year Senior Cycle programme, depending on whether an optional Transition Year (TY) is taken by students (Department of Education and Science, 2004). At the end of the Junior Cycle programme (in June), students are assessed on several core (e.g. English, Mathematics) and optional subjects in a final written state exam. Incorporated into the Junior Cycle programme, students complete various practical tests (in subjects such as Home Economics, Music, and Metalwork) and Classroom-Based Assessments (CBAs) in First and Second Year (Grades 8 and 9) in order to complete their Junior Cycle Profile of Achievement (JCPA) (Department of Education, 2023a; National Council for Curriculum and Assessment, 2024). At senior cycle, the Leaving Certificate programme covers two years of post-primary education called Fifth and Sixth Year (Grade 11 and 12), with most students completing the Leaving Certificate written state examination in June of the second year. Similar to the Junior Certificate, there are practical examinations and/or coursework in some subjects including Engineering and Construction Studies which are completed during March-May of the final year. Approximately four months prior to sitting their final exam (i.e. during the month of February), students preparing for the Junior Cycle Profile of Achievement and the Leaving Certificate sit in-house practice exams (known as *mock examinations*) in preparation for the final state exam. Transition Year is a non-academic 'gap' year in Irish post-primary education. It is officially recognised as the first year of a three-year Senior Cycle (Department of Education, 1993), although considered separate from the two-year Leaving Certificate programme. While classes in core examination subjects are provided during TY, there are no final state examinations at the end of the year. A key feature of the TY programme is the out-of-school activities. Students may participate in school exchanges, community service and multi-week placements in the workplace.

In addition to the preparation for state exams, there are pressures and demands on the staff and management of the schools associated with the administration and delivery of various school programmes, the timing of the Easter holidays, national holidays such as St Patrick's Day on 17th March, and the potential for other national/ international surveys to be in the field in the springtime (Denner, 2023).

One key feature of the PISA assessment is that it uses age-based sampling. This means that the operational definition of the age population in PISA directly depends on the time of testing; hence, age definitions for the Spring and Autumn administrations will differ. In 2018 in Ireland, the Spring testing period was from 1st March - 20th April and so the age definition was 1st January 2002 to 31st December 2002. The Autumn testing period was recorded as 15th October 2018 to 23rd November 2018 and the age definition was 1st August 2002 to 31st July 2003. This means that the age of students in PISA was 15 years and 3 months to 16 years and 2 months (allowing for a one-month deviation) regardless of when testing occurs. This therefore allows for the possibility of moving PISA testing within a given cycle.

In Ireland, moving PISA testing to the Autumn results in a somewhat different distribution of students across grade levels, in that more PISA-eligible students would have completed Third Year and would be in TY and Fifth Year. Also, those students in Third Year (of which there would be fewer) would have more time to focus on a low-stakes assessment such as PISA as they would not be in the process of completing CBAs or other project work. For Spring testing, a majority of students are in Third Year (Grade 9), while for Autumn testing, a majority are in Transition Year (Grade 10) (Denner, 2023; Donohue et al., 2023a; McKeown et al, 2019).

Due to the nature of PISA, which is not a curriculum-based assessment, and which uses an age-based sample, the possibility of moving PISA testing from the Spring to the Autumn (a less busy time of the school year) was an issue discussed informally between the Educational Research Centre and the Department of Education and Skills for some time prior to 2018.

Prior Research on Changes to PISA Testing Time

Similar to Ireland, other countries experience challenges associated with the administration of PISA in the early Spring and some moved their PISA testing from Spring to Autumn in the early years of PISA testing. These included the United States (US) and the United Kingdom (UK). In the US, schools across the nation during springtime are focused on preparing their students for state assessments that have an important bearing on their future. The US had a poor school response rate in the 2003 cycle of PISA (in Spring) and, with permission from the PISA contractors, they supplemented their sample by conducting a second round of testing in the Autumn of that year. An examination of the school average performance at these two time points (after adjusting selection criteria based on age for Autumn administration) revealed that performance on PISA in the Autumn and the Spring was not statistically significantly different (Ferraro et al., 2009). However, Ferraro et al.'s (2009) methodology treated session effects as school-level effects because it was the schools rather than the students that had opted to be tested and so their outcomes reflect differences (or lack of them) across schools rather than students. The United Kingdom did not conduct any analysis on the impact of moving PISA testing from the Spring to the Autumn when they changed their testing window in 2006. A working paper by Jerrim (2011) on the decline of England's PISA mathematics scores refers to the change in season when PISA was administered as one of the reasons for a drop in performance on the PISA test. In a more recent cycle of PISA (2022), the Netherlands moved to the Autumn for testing. The reason was mainly due to the fact that many 15-year-olds in the lower levels of secondary education sit their final school examinations in May and prepare for them in the weeks before that. Another related issue that may have contributed to the move was that, in 2018, several students, parents and schools voiced objections to taking part in PISA during the springtime (Dr. A.M.L. van Langen, personal communication, March 11, 2020).

However, such a move from Spring to Autumn testing is not without its challenges. One concern for policy makers in Ireland considering moving PISA testing to the Autumn was the lack of research on Summer Learning Loss (SLL). The PISA test is not considered to be curriculum based as it is not intended to capture educational attainment or command of a school curriculum per se, but rather provides broad indicators of the knowledge and skills needed in adult life and for full participation in society. However, acknowledging that an element of students' knowledge and skills required in adult life (i.e., their performance on

PISA) is based on what students have learnt in school, there was a concern that students' summer learning or lack of summer learning could bias the results of tests carried out in the Autumn time. While PISA testing in the Autumn can only begin six weeks after the summer holidays (OECD, 2020), there is a possibility that, owing to differential rates of summer learning loss, PISA testing in the Autumn may have a greater negative impact on students in lower socio-economic environments. In a widely-cited meta-analysis by Cooper et al. (1996), SLL was equated to about one-tenth of a standard deviation in test scores on average, was greater for mathematics than for reading, and was greater for children of lower-income than higher-income families. Much of the research on SLL has been undertaken in the United States and is based on performance at primary level. More recent research, involving children in kindergarten to Eighth Grade, which has used longitudinal analysis techniques (Kuhfield, 2020), has shown that the extent of summer learning loss is associated with the size of gains made in the previous school year. While there is very little research in Ireland on SLL, the Department of Education provides summer programmes for students with complex special educational needs and those at greatest risk of educational disadvantage. A key aim of these programmes is to provide a variety of rich educational experiences to the student and support their connection with school (Department of Education, 2023b). The lack of research on SLL in Ireland was a concern in the context of moving to Autumn testing. Hence, in 2018, a Feasibility Study was conducted in Ireland during the Autumn to compare Spring and Autumn cohorts taking PISA (Denner, 2023). The initial results of this feasibility study indicated that there were no statistically significant differences in the overall performance of students on reading literacy, mathematics and science, or on the distribution of performance across proficiency levels, between the two administrations. These results were derived from the use of descriptive statistics comparing mean scores and percentages between Spring and Autumn. This current study goes one step further in the analysis to examine achievement in the Spring and Autumn while controlling not only student and school-level background variables but other variables that are known to have a strong association with achievement in PISA.

Method

Sampling and participants

PISA selects its nationally representative samples based on a two-stage stratified sampling design. Stratification was used in the sampling process in PISA, with the same stratification variables used in both the Autumn and Spring samples (OECD, 2020; Denner, 2022). To manage cost and system burdens, Ireland was treated as an *adjudicated entity* for the Feasibility Study, additional to the primary adjudicated entity of the PISA 2018 Main Study (Spring testing). In total, 5,577 students across 157 schools participated in the PISA test in the Spring, representing a student response rate of 82.5%. There was a similar student response rate for Autumn testing (82.0% or 1,988 students across 57 schools) (Denner, 2023).

Administration of PISA – Spring and Autumn

Identical procedures and standards were applied in both the Spring and Autumn administrations of PISA in Ireland. All these procedures and guidelines were set out by the PISA contractors in their Technical Standards (OECD, 2020) to ensure that all PISA operational objectives are met in a timely and coordinated manner and to ensure consistency, accuracy and comparability of results across schools. Full details of the administration of both Spring and Autumn procedures can be found in Denner (2023) and McKeown et al. (2019).

Measures and variables

The cognitive instruments used for the Spring and Autumn administrations of PISA comprise online tests of reading literacy, mathematical literacy and science literacy, details of which can be found in the PISA 2018 assessment and analytical framework (OECD, 2019a). Scaling of the achievement data in both Spring and Autumn was conducted by the PISA 2018 contractors (Education Testing Service). In Ireland, item parameters used in the scaling of Autumn data were based on those obtained during the scaling of Spring data (these item parameters were based on all the 2018 international data as well as historical data from past cycles from Ireland and from other countries). To generate population-level proficiency estimates, PISA uses the imputation methodology of plausible values. Plausible values constitute random numbers drawn from the distribution of scores that could be reasonably assigned to each individual (Wu, 2005). Ten plausible values were generated in PISA 2018 for each student in each domain.

PISA collects background information using questionnaires (McKeown et al., 2019; OECD, 2019a). Selection of student- and school-level background variables in the Autumn study was based on their use in the stratification during the sampling process, as well as some items of national interest. Student background variables included: gender (male, female) and immigrant language status (native students, immigrant students who normally spoke English/Irish at home, and immigrant students who normally spoke another language at home). Also included at the student level is socio-economic status (ESCS score, referring to the PISA index of students' Economic, Social, and Cultural Status). The distribution of ESCS score at the student level is split into four roughly equal groups: lowest ESCS quartile, low-medium ESCS quartile, medium-high ESCS quartile, and highest ESCS quartile. The selected school background variables used in this study were: school sector and gender composition (girls' secondary, boys' secondary, mixed secondary, community/comprehensive, and ETB vocational) and DEIS (Delivering Equality of Opportunity in Schools) status (Non-DEIS, DEIS). DEIS status is an indicator of school educational disadvantage (DES, 2021). In comparing achievement at Spring and Autumn testing, a new binary variable called 'session' was created, with schools tested in the Spring given a code '0' and those in Autumn a code '1'.

Included with the student and school background variables are other learning outcome variables that have associations with achievement (OECD, 2019b, 2019c, 2019d) and have been shown to fluctuate across time. For example, Opdenakker et al. (2012) found that students' autonomous motivation (i.e. intrinsic motivation) declined over the course of a school year as their teachers "focused too much on keeping students on task and (unconsciously) neglected the interpersonal relationships" (p. 113). Wise et al. (2014) found that the use of different learning strategies, in particular how students approach a test by implementing learnt strategies, can fluctuate at different times of the year. Ouweneel et al. (2013) found that changes in self-efficacy did appear to occur over the course of the school year, although they did note that their study was experimental. There was also a concern that there may be a loss in 'Opportunity to Learn' key skills ahead of testing in the Autumn. Schmidt et al. (2013) found that there was a significant relationship with the overall PISA measure of mathematics literacy at the student level in Ireland and constructed 'Opportunity to Learn' indices, in particular exposure to word problems, formal mathematics topics, and applied mathematics problems. As reading was the main domain in PISA 2018, variables such as exposure to content, various reading activities in English classes and related teacher variables were included in the analysis. In general, the prioritisation of these constructs is supported by their relationships with performance and the potential for such constructs to fluctuate throughout the school year. Their selection in this paper is not intended to represent an exclusive set of variables associated with achievement; however, for the most part, they are PISA related, in that they are often examined in PISA. Details of the variables can be found in OECD (2019b, 2019c, 2019d).

Statistical analysis

The initial analysis to test for significant differences between the Autumn and Spring testing involved applying independent samples t-tests. The IEA (International Association for the Evaluation of Educational Achievement) International Database Analyzer (IDB Analyzer; IEA, 2016) was used to supply the standard errors. It generates

SPSS syntax that considers information from the sampling design in the computation of sampling variance and is capable of processing plausible values. The main difference between the Spring and Autumn data was the difference in the sample sizes (the Autumn sample had 57 schools (1,988 participating students) and the Spring sample had 157 schools (5,577 participating students)). This difference was addressed in the computation of standard errors (which considers sampling error) and, where relevant, in the adjustment of the alpha level associated with confidence intervals. The 95% confidence level is the most used level when evaluating the difference between two means. This is the level used in this study; however, given the smaller sample size from the Autumn testing, examination of the results at the 99% (or alpha of .01) may also be referred to, to allow a greater degree of certainty in reporting identified differences.

The second stage of the analysis involved multilevel modelling which was conducted using *Mplus* (Muthén & Muthén, 2017). *Mplus* is a statistical modelling programme capable of handling various types of data (e.g. cross-sectional, longitudinal data, single-level and multilevel data). The purpose of modelling data is to describe the structure of data in a simple way so that it is easier to understand and interpret (Muthén & Muthén, 2017). This specialised software takes appropriate account of the clustered nature of the data. The two-stage sampling in PISA means that schools are the primary sampling unit, with students within the sampled schools as the secondary sampling units. The multilevel modelling used students as level one and schools at level two (applying a two-level model). To take account of the ten plausible values, imputation techniques were applied by *Mplus* to involve all plausible values in the analysis (i.e., *type = imputation*). To account for the complex unit sampling design, weights were used. In multilevel models, the design weights need to be disentangled, and weight adjustment factors introduced at the multiple sampling stages. In this study the final student weight was separated into a "within-school weight" and a "between-school weight" for the appropriate estimation (i.e. by dividing the final student weight by the final school weight). For the multilevel models in this study maximum likelihood estimation was used and *grand mean centering* was considered appropriate for the variables used. All the models presented in this paper are 'random intercept models' which means that the intercepts were allowed to vary, and therefore, the score on the dependent variable (performance on the PISA test) for each individual observation was predicted by the intercept that varies across groups/clusters.

The following steps were employed when conducting the multilevel modelling:

- **Model 1**: **Reading, mathematics and science performance on PISA with school-level background variables.** This involved the examination of a model similar to that constructed by Ferraro et al. (2009) using PISA 2003 data.
- **Model 2: Reading, mathematics and science performance on PISA with student and school-level background variables.** Student background variables, as examined in this study (gender, ESCS and immigrant and language

status), were then added to Model 1. At this stage, interactions between 'Session' and the student background variables were also examined.

- **Analysis of other learning outcome variables.** Other learning outcome variables were examined one-by-one for their associations with achievement in reading, mathematics and science. The significance of a variable was determined by adding that variable to Model 2 (i.e., where student and school background variables were controlled) and, if $p < .05$, the variable was retained.
- **Model 3: Reading, mathematics and science performance on PISA with student and school-level background variables and other learning outcome variables significantly associated with achievement**. The model was finalised by retaining only those variables that were statistically significant when all variables were added to the model together. Interactions between 'Session' and each other learning outcome variable in the model were also examined.

Results

The first part of this section presents a summary of the overall performance for PISA 2018 Spring and Autumn testing in Ireland including performance by specific sub-groups. Also included is a summary of other factors related to PISA performance that have been shown in other studies to fluctuate over time during the academic year. The second part presents the results from the three multilevel models.

Summary of performance in PISA 2018, Spring and Autumn

The overall differences in mean scores between Spring and Autumn were 2.3, 2.9 and 1.0 score points respectively for reading, mathematics, and science (Table 1). None of these were statistically significant at either the .05 or .01 levels. Among subgroups, there was a 9.3-point difference for males on the mathematics scale between PISA 2018 Spring and Autumn testing (502.6 and 493.4, respectively); although larger compared to the corresponding difference in reading, it was not statistically significant. The score difference for females (3.4 score points higher in Autumn), was not statistically significant either.

TABLE 1

Overall mean scores and standard errors for reading, mathematics and science, including student background variables, Spring and Autumn testing

	Spring		**Autumn**		**Spring - Autumn**	
	Mean	**(SE)**	**Mean**	**(SE)**	**Mean Diff**	**(SE Diff)**
Reading						
All students	518.1	(2.2)	515.8	(4.1)	2.3	(4.7)
Male	506.4	(3.0)	503.1	(5.7)	3.3	(6.5)
Female	529.6	(2.5)	528.7	(4.0)	0.9	(4.7)
Lowest ESCS	482.3	(3.0)	479.3	(6.1)	3.0	(6.8)
Low-Medium ESCS	510.0	(3.0)	500.6	(5.1)	9.4	(5.9)
Medium-High ESCS	527.4	(2.9)	531.4	(5.2)	-4.1	(5.9)
Highest ESCS	556.7	(3.0)	554.1	(5.1)	2.6	(5.9)
Native	522.1	(2.3)	521.0	(4.2)	1.1	(4.8)
Immigrant & Eng./Irish	518.2	(5.2)	506.6	(6.6)	11.7	(8.4)
Immigrant & other lang.	498.5	(5.3)	495.9	(7.2)	2.6	(9.0)
Mathematics						
All students	499.6	(2.2)	496.7	(3.6)	2.9	(4.2)
Male	502.6	(2.9)	493.4	(4.9)	9.3	(5.7)
Female	496.7	(2.7)	500.1	(3.8)	-3.4	(4.6)
Lowest ESCS	466.9	(3.1)	463.4	(5.3)	3.5	(6.1)
Low-Medium ESCS	491.4	(2.5)	483.5	(4.7)	7.8	(5.3)
Medium-High ESCS	509.0	(2.8)	509.1	(5.0)	-0.2	(5.7)
Highest ESCS	533.9	(3.2)	532.2	(4.3)	1.7	(5.4)
Native	501.9	(2.3)	500.2	(3.6)	1.6	(4.2)
Immigrant & Eng./Irish	494.3	(4.8)	493.6	(6.5)	0.7	(8.1)
Immigrant & other lang.	498.5	(5.6)	486.1	(7.1)	12.4	(9.1)
Science						
All students	496.1	(2.2)	495.1	(4.0)	1.0	(4.5)
Male	495.4	(3.0)	492.7	(5.6)	2.7	(6.3)
Female	496.9	(2.6)	497.6	(4.0)	-0.8	(4.7)
Lowest ESCS	460.0	(3.3)	458.3	(6.3)	1.7	(7.1)
Low-Medium ESCS	487.4	(2.9)	480.1	(5.1)	7.3	(5.9)
Medium-High ESCS	506.8	(2.8)	509.2	(5.3)	-2.4	(6.0)
Highest ESCS	533.6	(3.4)	534.4	(4.8)	-0.8	(5.9)
Native	498.1	(2.3)	497.7	(4.0)	0.4	(4.6)
Immigrant & Eng./Irish	499.7	(5.2)	496.4	(7.1)	3.3	(8.8)
Immigrant & other lang.	496.0	(5.4)	487.2	(6.8)	8.8	(8.7)

Note: **Bold** indicates statistically significant difference with *indicating .05 level and **indicating .01 level.

Examining school type by sector-gender composition, four of the five school types had very similar mean scores in Spring and Autumn (Table 2). However, in the case of mixed secondary schools, mean scores were higher in Spring than in Autumn by some 23-25 score points in each domain (though these differences did not reach statistical significance, the corresponding effect sizes ranged from d=0.25 to d=0.28). According to the What Works Clearinghouse (2014), an effect size of 0.25 or higher on school-based research can be considered 'substantively important' (that is, the association with achievement is strong), irrespective of whether or not the underlying difference is statistically significant. Students in DEIS schools did not score statistically significantly differently between Spring testing and Autumn testing on any of the three domains (with score differences of 0.3 in reading, 0.9 in mathematics and 4.8 in science – all higher in Spring).

TABLE 2

Mean scores and standard errors for reading, mathematics, and science by school background variables (percentages included), Spring and Autumn testing

	Spring Testing			Autumn Testing			Spring - Autumn	
	%	**Mean**	**(SE)**	**%**	**Mean**	**(SE)**	**Mean Diff**	**(SE Diff)**
Reading								
Girls' secondary	21.2	541.7	(3.5)	19.2	538.4	(7.5)	3.3	(8.3)
Boys' secondary	15.0	511.1	(6.0)	15.9	515.3	(11.3)	-4.2	(12.8)
Community/comprehensive	16.9	507.6	(3.8)	17.5	516.6	(6.5)	-9.0	(7.6)
Mixed secondary	17.4	536.0	(7.7)	19.0	511.5	(13.4)	24.5	(15.5)
ETB vocational	29.5	500.0	(4.2)	28.3	503.1	(7.0)	-3.1	(8.1)
Non-DEIS	75.9	530.4	(2.5)	74.4	528.5	(3.7)	1.9	(4.5)
DEIS	24.1	479.2	(4.8)	25.6	478.9	(6.5)	0.3	(8.0)
Mathematics								
Girls' secondary	21.2	506.6	(3.9)	19.2	507.7	(7.9)	-1.1	(8.8)
Boys' secondary	15.0	510.0	(5.6)	15.9	508.2	(10.0)	1.7	(11.5)
Community/comprehensive	16.9	492.2	(4.2)	17.5	494.4	(4.8)	-2.3	(6.4)
Mixed secondary	17.4	517.4	(6.5)	19.0	494.4	(11.4)	23.0	(13.2)
ETB vocational	29.5	483.2	(3.9)	28.3	485.8	(5.5)	-2.6	(6.7)
Non-DEIS	75.9	510.2	(2.4)	74.4	507.5	(3.4)	2.7	(4.2)
DEIS	24.1	466.4	(4.5)	25.6	465.4	(5.0)	0.9	(6.8)
Science								
Girls' secondary	21.2	506.6	(3.8)	19.2	504.4	(8.4)	2.2	(9.2)
Boys' secondary	15.0	498.8	(5.5)	15.9	503.6	(10.6)	-4.8	(11.9)
Community/comprehensive	16.9	488.3	(4.0)	17.5	497.6	(5.1)	-9.3	(6.5)
Mixed secondary	17.4	515.0	(7.7)	19.0	492.2	(13.5)	22.9	(15.5)
ETB vocational	29.5	480.5	(4.0)	28.3	484.6	(6.8)	-4.1	(7.9)
Non-DEIS	75.9	506.0	(2.6)	74.4	507.1	(3.5)	-1.2	(4.3)
DEIS	24.1	465.0	(4.7)	25.6	460.3	(6.3)	4.8	(7.9)

Note: **Bold** indicates statistically significant difference with *indicating .05 level and **indicating .01 level.

In PISA, the reading, mathematics and science scales are divided into a range of proficiency levels. Students performing below Level 2 are often referred to as lower-achieving students or low performers in the PISA literature and are considered to be below a baseline level of proficiency in literacy required to enable successful participation in education and work (OECD, 2019b). In contrast, the term 'high-achievers' is often used to describe the combined percentages of students at Levels 5 and 6. While there was a larger percentage point difference between Spring and Autumn on PISA mathematics for low-achievers (2.4% more in Autumn), the difference was not statistically significant (Table 3). There were no statistically significant differences between the proportions of low-achieving students on the reading literacy or science scale for any of the student groupings examined (final two columns). Examining mathematics, there was a statistically significant difference at the .05 level between the Spring and Autumn testing for low-achieving males (15.7% achieved below Level 2 in the Spring compared to 21.4% in the Autumn; effect size h=0.15). Additional analysis in relation to the high-achieving students found that there were no statistically significant differences between the proportions of high-achieving students on the three domains examined for any of the student groupings, with a difference of less than two percentage points for the majority of the groupings.

Results of other learning outcomes when measured in Spring and Autumn

Several PISA measures related to engagement, motivation and other learning outcomes were examined in this study and those that were statistically significant were considered for inclusion in the final multilevel model. Outcomes for nine categorical variables were statistically significantly different (at least at the .05 level) across the two time periods (Table 4). A variable in the reading domain measuring an element of 'Opportunity to Learn' is exposure to content; students were asked how about the longest piece of text they had read for English class during the school year, with reference to the number of pages completed. There was a statistically significant difference on the length of text read for school as reported by students in the Autumn compared to the Spring. Approximately 42.4% of students reported reading *101 or more pages* in the Spring compared to 24.1% of the students tested in the Autumn (effect size h=0.39). The descriptive reading activity, to 'select a passage you liked or disliked and explain why', was completed by a higher percentage of students in the Spring compared to the Autumn (percentage score difference of 7.3%, effect size h=0.15).

School absenteeism appeared to be more prevalent in the Autumn compared to the Spring. Almost 29% of students involved in Autumn testing reported *skipping 1 or 2 days* compared to 24.2% in Spring testing (effect size h=0.10).

TABLE 3

Percentages and standard errors of low-achievers for reading, mathematics and science for student-level variables, Spring and Autumn testing

	Spring Testing		Autumn Testing		Spring - Autumn	
	%	**(SE)**	**%**	**(SE)**	**% Diff**	**(SE Diff)**
Reading						
All low-achievers	11.8	(0.7)	12.4	(1.4)	-0.6	(1.5)
Male	15.1	(1.0)	17.6	(2.2)	-2.5	(2.4)
Female	8.5	(0.7)	7.2	(1.1)	1.4	(1.3)
Lowest ESCS	20.5	(1.3)	20.2	(2.5)	0.3	(2.8)
Low-Medium ESCS	11.5	(1.1)	14.9	(2.2)	-3.4	(2.5)
Medium-High ESCS	8.8	(0.8)	9.0	(1.4)	-0.3	(1.7)
Highest ESCS	5.3	(0.8)	5.4	(1.4)	-0.1	(1.7)
Native	10.9	(0.7)	11.3	(1.3)	-0.5	(1.5)
Immigrant & Eng./Irish	12.7	(1.7)	13.4	(3.2)	-0.6	(3.7)
Immigrant & other lang.	14.9	(2.1)	16.6	(3.0)	-1.8	(3.7)
Mathematics						
All low-achievers	15.7	(0.8)	18.1	(1.7)	-2.4	(1.9)
Male	15.7	(1.1)	21.4	(2.5)	**-5.7***	(2.8)
Female	15.7	(1.1)	14.7	(1.6)	1.0	(1.9)
Lowest ESCS	27.3	(1.7)	27.8	(3.2)	-0.5	(3.7)
Low-Medium ESCS	16.5	(1.3)	22.1	(2.7)	-5.6	(3.0)
Medium-High ESCS	11.5	(1.2)	13.9	(2.1)	-2.4	(2.4)
Highest ESCS	6.4	(1.0)	8.2	(1.7)	-1.8	(1.9)
Native	14.9	(0.8)	16.7	(1.8)	-1.8	(1.9)
Immigrant & Eng./Irish	19.2	(2.4)	19.4	(3.8)	-0.2	(4.5)
Immigrant & other lang.	14.6	(2.7)	20.8	(3.5)	-6.2	(4.4)
Science						
All low-achievers	17.0	(0.8)	17.3	(1.5)	-0.3	(1.7)
Male	18.1	(1.2)	19.9	(2.1)	-1.8	(2.4)
Female	16.0	(1.1)	14.7	(1.6)	1.3	(2.0)
Lowest ESCS	29.0	(1.7)	28.1	(3.3)	0.9	(3.7)
Low-Medium ESCS	17.5	(1.4)	20.8	(2.4)	-3.3	(2.7)
Medium-High ESCS	12.5	(1.1)	13.2	(1.6)	-0.7	(1.9)
Highest ESCS	7.8	(1.0)	7.0	(1.3)	0.8	(1.7)
Native	16.1	(0.9)	16.5	(1.5)	-0.4	(1.7)
Immigrant & Eng./Irish	18.1	(2.1)	15.7	(3.6)	2.4	(4.2)
Immigrant & other lang.	17.0	(2.4)	20.2	(3.3)	-3.3	(4.0)

Note: **Bold** indicates statistically significant difference with *indicating .05 level and **indicating .01 level.

TABLE 4

Percentages and standard errors of categorical variables considered for inclusion in the final multilevel model, Spring and Autumn testing

	Spring			Autumn			Spring-Autumn	
	N	**%**	**(SE)**	**N**	**%**	**(SE)**	**% Diff**	**(SE Diff)**
Longest piece of text read-up to 10 pages)	1,980	36.4	(0.8)	1,059	54.2	(1.9)	**-17.9***	(2.1)
Longest piece of text read-10-100 pages	1,177	21.2	(0.6)	415	21.6	(1.3)	-0.4	(1.4)
Longest piece of text read-101 or more pages	2,353	42.4	(0.9)	471	24.1	(1.5)	**18.3***	(1.7)
Write a summary of the book/ chapter -Yes, activity completed	4,300	78.0	(0.8)	1,385	71.5	(2.1)	**6.5***	(2.2)
Compare the content of the book or the chapter with your own experience -Yes, activity completed	1,861	33.8	(0.8)	666	33.9	(1.1)	-0.1	(1.4)
Select a passage you liked or disliked and explain why -Yes, activity completed	3,702	67.3	(0.8)	1,164	60.0	(1.4)	**7.3***	(1.6)
How to use keywords when using a search engine -Yes, was taught it	2,436	44.3	(0.9)	958	48.5	(1.3)	**-4.1***	(1.6)
To understand the consequences of making information publicly available online -Yes, was taught it	4,567	83.1	(0.7)	1,607	82.5	(1.3)	0.5	(1.5)
How to use the short description below the links in the list of search results -Yes, was taught it	1,924	35.1	(0.8)	775	39.5	(1.4)	**-4.4***	(1.6)
How to detect whether the information is subjective or biased -Yes, was taught it	1,543	28.0	(0.9)	662	33.4	(1.5)	**-5.4***	(1.8)
Skipped a whole day in 2 weeks before PISA testing -No whole days skipped	3,249	70.1	(0.8)	1,092	65.4	(1.3)	**4.7***	(1.5)
Skipped a whole day in 2 weeks before PISA testing -1-2 whole days skipped	1,116	24.2	(0.7)	472	28.5	(1.3)	**-4.3***	(1.5)
Skipped a whole day in 2 weeks before PISA testing -3 or more whole days skipped	264	5.7	(0.3)	107	6.1	(0.5)	-0.4	(0.6)
Arrived late to school in 2 weeks before PISA testing -No days late	3,058	66.4	(1.0)	1,109	66.3	(1.8)	0.1	(2.1)

	Spring			Autumn			Spring-Autumn	
	N	**%**	**(SE)**	**N**	**%**	**(SE)**	**% Diff**	**(SE Diff)**
Arrived late to school in 2 weeks before PISA testing -1-2 late days	1,136	24.5	(0.8)	394	23.8	(1.3)	0.7	(1.5)
Arrived late to school in 2 weeks before PISA testing -3 or more late days	426	9.2	(0.5)	165	9.9	(1.0)	-0.8	(1.1)

Note: **Bold** indicates statistically significant difference with *indicating .05 level and **indicating .01 level.

Several continuous variables were considered for inclusion in the final multilevel model. A composite index of assessing credibility of texts was constructed by the OECD with a mean score of 0.0 and a standard deviation of 1. The mean score in Ireland for the Spring testing was 0.21 compared to 0.13 in Autumn testing, with a statistically significant difference at the .05 level and a small effect size (d=0.08) (Table 5). Three indicators in relation to wellbeing were presented to students in PISA 2018: 'life satisfaction'; eudemonia (a sense of meaning and purpose in life); and subjective wellbeing (positive emotions). Only the affective element of subjective wellbeing (positive feelings) showed a statistically significant decrease (albeit very small effect size, d=0.07) from Spring testing (-0.09) to Autumn testing (-0.16). This suggests that students in the Spring reported slightly more positive feelings than students in the Autumn. Students at Spring testing also reported a more positive 'attitude toward school/learning activities' than those tested in the Autumn, with a statistically significant difference (based on mean scores of 0.12 and 0.00, respectively).

TABLE 5

Means and standard errors of continuous variables considered for inclusion included in the final multilevel model, Spring and Autumn testing

	Spring			Autumn			Spring-Autumn	
	N	**Mean**	**(SE)**	**N**	**Mean**	**(SE)**	**Mean Diff**	**(SE Diff)**
Teachers' stimulation of reading engagement	5,508	0.06	(0.02)	1,946	0.05	(0.03)	0.02	(0.03)
Teacher-directed instruction	5,530	-0.20	(0.02)	1,953	-0.21	(0.04)	0.01	(0.04)
Understanding and remembering	5,287	0.05	(0.02)	1,860	0.04	(0.03)	0.01	(0.03)
Summarising strategies	5,326	0.10	(0.02)	1,868	0.10	(0.03)	0.00	(0.03)
Assessing credibility	5,256	0.21	(0.01)	1,854	0.13	(0.03)	**0.08***	(0.04)
Perceived competence in reading	5,454	0.12	(0.01)	1,923	0.07	(0.04)	0.05	(0.04)
Perceived difficulty in reading	5,450	0.00	(0.01)	1,922	0.04	(0.03)	-0.03	(0.04)
Enjoyment of reading	5,519	-0.07	(0.02)	1,951	-0.13	(0.03)	0.06	(0.04)
General self-efficacy (resilience)	5,391	-0.04	(0.01)	1,903	-0.03	(0.02)	-0.01	(0.03)
Attitudes towards competition	5,443	0.16	(0.01)	1,919	0.16	(0.02)	0.00	(0.03)
Motivation to master tasks	5,362	-0.09	(0.01)	1,899	-0.11	(0.03)	0.02	(0.03)
Life satisfaction	5,459	6.74	(0.05)	1,932	6.69	(0.08)	0.04	(0.09)
Eudemonia: meaning in life	5,355	-0.18	(0.01)	1,893	-0.22	(0.03)	0.04	(0.03)
Subjective wellbeing: positive affect	5,339	-0.09	(0.01)	1,893	-0.16	(0.03)	**0.07***	(0.03)
Attitude towards school/learning activities	5,476	0.12	(0.02)	1,928	0.00	(0.03)	**0.11***	(0.03)

Note: **Bold** indicates statistically significant difference with *indicating .05 level and **indicating .01 level.

Multilevel analysis of performance in PISA 2018, Spring and Autumn

Table 6 presents outcomes from the multilevel models for reading, mathematics and science performance on PISA with 'session' and two school-level background variables (DEIS status and sector-gender composition). The main coefficient of interest, that of 'session', did not reach statistical significance; i.e., controlling for all other school background variables in the models, the time of year was not associated with the achievement of students in PISA reading, mathematics, or science. For example, in the case of reading, performance was 1.23 points higher in Spring compared with Autumn, but this outcome was not statistically significant when school disadvantaged status (DEIS) and sector-gender composition were accounted for.

Building on Model 1, the models in Table 7 adds level 1 student background variables: gender, ESCS and student immigrant and language status. Across reading, mathematics and science, performance was 3.34, 9.95 and 2.69 points higher (respectively) in Spring compared with Autumn. Session was not statistically significant on reading and science and significant on mathematics (*coefficient: 9.95; p-value: 0.049)* when the student level 1 background variables were added to the model.

TABLE 6

Model Set 1: Reading, mathematics and science performance on PISA with school-level background variables

	Reading			Mathematics			Science		
	Coefficient	**SE**	**p-value**	**Coefficient**	**SE**	**p-value**	**Coefficient**	**SE**	**p-value**
Intercept									
	519.36**	8.15	<.001	**499.84****	7.05	<.001	**499.01****	7.32	<.001
Level 2 (Between School)									
DEIS school status (ref.=Non-DEIS)	**-51.30****	6.34	<.001	**-43.48****	5.48	<.001	**-46.95****	6.16	<.001
School Sector Gender (ref.=ETB Vocational)									
Girls' secondary	**19.23***	8.05	0.017	5.46	7.05	0.439	4.75	7.79	0.542
Boys' secondary	1.42	8.70	0.871	**14.94***	7.54	0.048	8.73	8.37	0.297
Community/Comprehensive	4.11	7.18	0.567	3.24	6.31	0.607	3.59	6.63	0.589
Mixed secondary	13.01	8.46	0.124	12.34	7.34	0.093	10.34	8.73	0.237
Session (ref.=Autumn)	1.23	5.06	0.808	1.16	4.48	0.795	0.95	5.11	0.852
R-Squared									
Between-schools	0.542			0.511			0.457		

Note: Each model based on 7,565 students, no missingness. (ref.) = reference group.

Bold indicates statistically significant difference with *indicating .05 level and **indicating .01 level.

TABLE 7

Model Set 2: Reading, mathematics and science performance on PISA with student and school-level background variables

	Reading			Mathematics			Science		
	Coefficient	**SE**	**p-value**	**Coefficient**	**SE**	**p-value**	**Coefficient**	**SE**	**p-value**
Intercept									
	509.14**	7.09	<.001	**494.77****	6.63	<.001	**496.45****	6.43	<.001
Level 1 (Within School)									
Gender (ref.= Male)	**18.64****	3.52	<.001	4.88	5.07	0.335	-1.85	3.79	0.626
Immigrant Language Status (ref.=Native)									
Immigrant English/Irish	-7.71	4.13	0.062	-6.20	3.68	0.092	0.32	3.93	0.934
Immigrant other language	**-11.34***	5.44	0.037	4.22	5.07	0.405	5.80	5.43	0.285
ESCS	**27.63****	1.65	<.001	**25.35****	1.44	<.001	**28.17****	1.59	<.001
Session by Gender	a	a	a	**-15.35****	5.88	0.009	a	a	a
Level 2 (Between School)									
DEIS status (ref.=Non-DEIS)	**-33.27****	5.69	<.001	**-28.22****	5.03	<.001	**-29.73****	5.48	<.001
School Sector Gender (ref.=Vocational)									
Girls' secondary	9.77	6.89	0.156	7.23	6.18	0.242	5.54	6.93	0.424
Boys' secondary	8.21	7.67	0.284	12.23	7.02	0.081	6.64	7.31	0.364
Community/Comprehensive	2.24	6.18	0.717	2.52	5.38	0.639	2.42	5.72	0.673
Mixed secondary	9.48	7.66	0.216	9.53	6.56	0.146	6.80	7.88	0.388
Session (ref.=Autumn)	3.34	4.26	0.433	**9.95***	5.06	0.049	2.69	4.32	0.534
R-Squared									
Within-schools	0.092			0.090			0.085		
Between-schools	0.460			0.473			0.374		

Note: Each model based on 7,271 students, overall missingness=3.9%. No missingness on Gender, DEIS status, School Sector Gender, missingness on Immigrant Language Status (3.3%) and ESCS (1.3%).

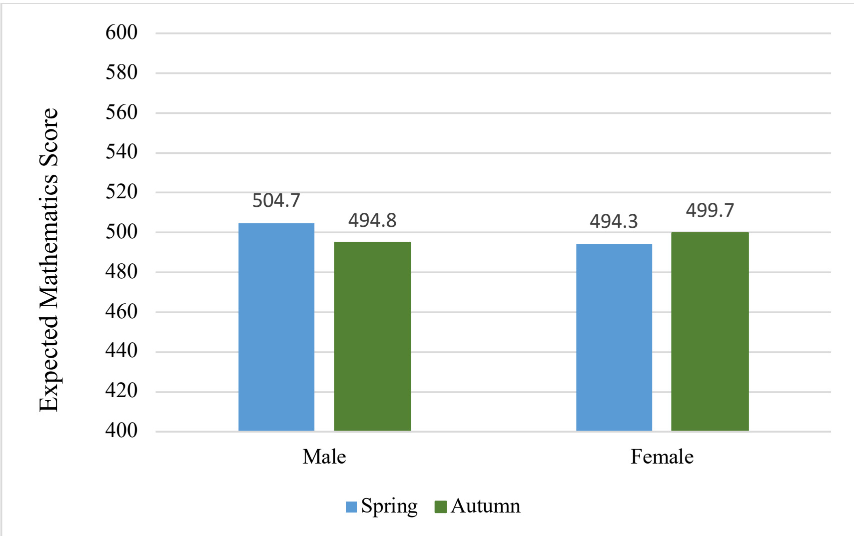
[a]:not included in the model for reading or science as it was not statistically significant. (ref.) = reference group.

Bold indicates statistically significant difference with *indicating .05 level and **indicating .01 level.

In developing these models, the interaction between 'session' and each level 1 background variable was examined. Controlling for student and school-level background variables, the coefficient for 'session' did not reach statistical significance on PISA reading or science performance. However, the interaction between 'session' and 'gender' was statistically significant on the mean mathematics score. This interaction can be seen in Figure 1, which shows that being tested in the Spring was more strongly associated with a lower mathematics score for females compared to males, and being tested in Autumn was associated with a lower mathematics score for males. Based on the model coefficients, and holding all other variables constant, a male's mathematics score in the Spring was predicted to be 504.7 compared to 494.8 in the Autumn. In contrast, a female's score in the Spring was 494.3 compared to 499.7 in the Autumn.

FIGURE 1

Interaction: Gender and Session on mathematics performance



The final models (one each for reading, mathematics, and science) which include only those variables that were statistically significant when all variables were added to each model together are presented in Table 8. Included in the model on reading performance are variables measuring elements of 'Opportunity to Learn' such as exposure to content (e.g., length of text read, writing a summary of the book/chapter and using a number of strategies for online reading) and the learning environment. Variables measuring engagement, motivation and wellbeing are also included in the reading model, indicating that, controlling for all other variables in the model, these indicators have a statistically significant relationship with reading performance. However, the main coefficient of interest is that of 'session' and the data indicate that when all other variables of interest were controlled, it was not statistically significantly related to reading performance. The interaction of this variable with each of the other level 1 learning outcome variables in the model was examined and the coefficient on the interaction between 'perceived difficulty in reading' by 'session' indicated statistical significance (*coefficient: - 4.54; p-value: 0.046,* see Appendix Table A1).

Examining the model on mathematical performance, the coefficient related to the interaction between 'session' and 'gender' that was seen as significant when only the level 1 and level 2 background variables were included no longer reached statistical significance when variables measuring engagement, motivation and other variables associated with performance were added to the model. There were no statistically significant interactions with session and the other level 1 learning variables in the model of mathematical performance.

Similar to the reading model, variables measuring engagement, motivation and other learning outcomes were statistically significant on science performance when all variables were added to the model. Again, controlling for these variables, 'session' did not reach statistical significance in the model on science performance. The coefficient on the interaction of the variable '3 or more whole days skipped' by 'session' indicated a statistical significance (*coefficient: - 21.54; p-value: 0.046*, see Appendix Table A1).

TABLE 8

Model Set 3. Reading, mathematics, and science performance on PISA with student- and school-level background variables, and other factors associated with performance

	Reading			Mathematics			Science		
	Coefficient	**SE**	**p-value**	**Coefficient**	**SE**	**p-value**	**Coefficient**	**SE**	**p-value**
Intercept									
	541.33**	6.01	<.001	**511.16****	6.78	<.001	**513.26****	6.41	<.001
Level 1 (Within School)									
Gender (ref. = Male)	0.69	2.80	0.804	-0.21	4.85	0.965	-4.51	3.63	0.214
Immigrant Language Status (ref.=Native)									
Immigrant with English	-4.11	3.51	0.242	-1.89	3.87	0.625	6.11	4.07	0.133
Immigrant with other language	**-11.21****	3.92	0.004	4.72	4.87	0.332	6.85	4.68	0.143
ESCS	**11.41****	1.50	<.001	**21.52****	1.55	<.001	**23.51****	1.61	<.001
Session by Gender	[a]	[a]	[a]	-10.71	5.71	0.061	[a]	[a]	[a]
Level 2 (Between School)									
DEIS school status (ref.=Non-DEIS)	**-16.92****	4.53	<.001	**-26.63****	5.08	<.001	**-27.92****	5.69	<.001
School Sector Gender (ref.=Vocational)									
Girls' secondary	2.06	4.69	0.66	3.63	6.39	0.57	0.88	7.37	0.905
Boys' secondary	2.74	4.93	0.578	9.36	6.64	0.159	2.47	6.84	0.718
Community/Comprehensive	-1.42	4.24	0.738	1.84	5.47	0.736	2.10	5.64	0.709
Mixed secondary	3.32	6.05	0.584	8.33	6.51	0.201	5.97	7.73	0.44
Session (ref.=Autumn)	-2.71	3.03	0.370	8.53	5.03	0.090	4.98	4.20	0.236

	Reading			Mathematics			Science		
	Coefficient	SE	p-value	Coefficient	SE	p-value	Coefficient	SE	p-value
Other Level 1 (Within School)									
Exposure to content									
Longest piece of text (ref.=up to 10 pages)									
10-100 pages	-4.42	3.16	0.162	a	a	a	a	a	a
101 or more pages	**10.58****	2.54	<.001	a	a	a	a	a	a
Write a summary of the book/ chapter (ref.=No)	**-7.02****	2.42	0.004	a	a	a	a	a	a
Compare the content of the book or the chapter with your own experience (ref.=No)	**-12.78****	2.29	<.001	a	a	a	a	a	a
Select a passage you liked or disliked and explain why (ref.=No)	**-7.57****	2.35	0.001	a	a	a	a	a	a
How to use keywords when using a search engine (ref.=No)	**-9.08****	2.42	<.001	a	a	a	a	a	a
To understand the consequences of making information publicly available online (ref.=No)	**15.25****	2.93	<.001	a	a	a	a	a	a
How to use the short description below the links in the list of search results (ref.=No)	**-19.38****	2.84	<.001	a	a	a	a	a	a
How to detect whether the information is subjective or biased (ref.=No)	**5.62****	2.16	0.009	a	a	a	a	a	a

	Reading			Mathematics			Science		
	Coefficient	SE	p-value	Coefficient	SE	p-value	Coefficient	SE	p-value
English Learning Environment									
Teachers' stimulation of reading engagement	**9.50****	1.61	<.001	a	a	a	a	a	a
Teacher-directed instruction	**-6.18****	1.24	<.001	a	a	a	a	a	a
Engagement									
Skipped a whole day in 2 weeks before PISA testing (ref.=No whole days skipped)^									
1-2 whole days skipped	a	a	a	**-7.87***	3.20	0.014	**-7.69***	3.50	0.028
3 or more whole days skipped	a	a	a	**-22.97****	5.47	<.001	**-26.57****	5.60	<.001
Arrived late to school in 2 weeks before PISA testing (ref.=No days late)^									
1-2 late days	**-8.10****	2.83	0.004	**-13.49****	3.09	<.001	**-15.13****	3.25	<.001
3 or more late days	**-12.26****	4.39	0.005	**-26.86****	4.94	<.001	**-33.00****	5.66	<.001
Understanding and remembering	**5.88****	1.26	<.001	**a**	a	a	**a**	a	a
Summarising strategies	**10.99****	1.33	<.001	**a**	a	a	**a**	a	a
Assessing credibility	**17.14****	1.09	<.001	**a**	a	a	**a**	a	a
Motivation									
Perceived competence in reading	**18.99****	1.50	<.001	**a**	a	a	**a**	a	a
Perceived difficulty in reading	**-4.57****	1.66	0.006	**a**	a	a	**a**	a	a
Enjoyment of reading	**10.06****	1.14	<.001	**a**	a	a	**a**	a	a
General self-efficacy (resilience)	**a**	a	a	**8.44****	1.75	<.001	**11.60****	1.98	<.001
Attitudes towards competition	**a**	a	a	**4.54****	1.38	0.001	**3.84****	1.50	0.01
Motivation to master tasks	**a**	a	a	**5.51****	1.47	<.001	**6.70****	1.62	<.001

	Reading			Mathematics			Science		
	Coefficient	**SE**	**p-value**	**Coefficient**	**SE**	**p-value**	**Coefficient**	**SE**	**p-value**
Other School Outcomes									
Life satisfaction	**1.71****	0.58	0.003	[a]	[a]	[a]	**1.66***	0.67	0.013
Eudemonia: meaning in life	**-10.60****	1.30	<.001	**-15.18****	1.50	<.001	**-20.32****	1.69	<.001
Subjective wellbeing: positive affect	**-4.45****	1.27	<.001	**-5.56****	1.48	<.001	**-9.01****	1.54	<.001
Attitude towards school/learning activities	[a]	[a]	[a]	**4.40****	1.29	0.001	**4.51****	1.47	0.002
R-Squared									
Within-schools	0.504			0.165			0.179		
Between-schools	0.363			0.468			0.371		

Note. [a] not included in the final model. ^Missing indicator included to retain data.

Final reading model based on 6,018 students; final mathematics model based on 6,711 students; final science model based on 6,676 students.

Bold indicates statistically significant difference with *indicating .05 level and **indicating .01 level.

Summary

Overall, student performance at Spring testing was not statistically significantly different to Autumn PISA testing across the three domains, reading, mathematics and science. Examining performance in terms of the proportions of low achievers (proficiency level below Level 2), percentages were similar across Spring and Autumn for each domain. However, there was a statistically significantly lower percentage of low-achieving males on the mathematics proficiency scale in the Spring (15.7%) compared to the Autumn (21.4%). In contrast, for each domain tested, the proportions of high achievers (proficiency at Level 5 and above) were the same for both testing periods.

Consideration of the outcomes of other factors associated with achievement and the time of testing showed some statistically significant differences when testing occurred in the spring compared to testing in the autumn. It was found that fewer students reported '*skipping a whole day*' in the two weeks before PISA testing in the Spring compared to the Autumn. A variable related to the construct 'Opportunity to Learn' is exposure to content and, in the reading domain, students were asked how many pages was the longest piece of text that they had read for English class during the school year. Across all students, a statistically significantly higher percentage of students reported reading '*101 or more pages of a text*' in the Spring compared to the Autumn. Also related to the reading domain, a statistically significantly higher percentage of students reported being asked by their teacher to carry out reading activities that require more cognitive capabilities in Spring compared to Autumn.

While considering other factors associated with achievement along with school and student background variables, multilevel modelling was carried out to take into account the clustered nature of the data. The first multilevel model examined was similar to that carried out by Ferraro et al. (2009) on the United States PISA 2003 data which controlled for level 2 school background variables. Similar to Ferraro et al.'s results, the coefficient for 'session' (i.e., time of year when testing took place) in this study did not reach statistical significance on any domain (reading, mathematics or science) while controlling for level 2 school background variables. The second model set in this study introduced a number of student-level background variables. While the coefficient for 'session' on the mathematics scale was not statistically significant, the interaction between 'session' and 'gender' was statistically significant on the mean mathematics score. This indicates that being tested in the Spring was more strongly associated with lower mathematics scores for females compared to males, and lower mathematics scores for males compared with females in the Autumn. The final model sought to control for all other PISA-related variables that were significantly associated with achievement while examining the coefficient for time of testing (session). The results from the final multilevel models (for each domain) indicate that time of year when PISA takes places did not have a statistically significant association with reading, mathematics or science performance while controlling for the school and student background variables and other level 1 variables that are associated with performance and examined in this study. In summary, many variables correlate with one another and re-affirm the associations between DEIS status, student ESCS, gender, and, for example, endorsement of reading comprehension strategies in explaining performance on PISA and this study shows that time of testing is not a major factor when these other variables are considered.

Conclusions

Initial findings from bivariate analysis (Denner, 2023) indicated that there was no significant difference in achievement in PISA when conducted in Spring or Autumn. This study, in providing more in-depth analysis, consolidates the findings of the initial report. The eighth cycle of PISA was administered in Autumn 2022 in Ireland. While other challenges were experienced with PISA 2022 (such as schools re-adjusting to issues surrounding COVID) data collected in the Autumn can be reliably used to evaluate policies such as Ireland's recent *Literacy, Numeracy and Digital Literacy Strategy 2024-2033* (Department of Education, 2024) and *STEM Education Policy Statement 2017-2026* (Department of Education and Skills, 2017c). However, for PISA 2025, the international testing window is between March and September, when it was previously March to December. Ireland, along with a number of other countries who have tested for the most part in the Autumn, will be reverting back to administering PISA earlier in the year. Nevertheless, the results of the current study will alleviate any concerns for countries moving their testing period to between March and September, where they have not conducted any prior research on the relationship between time of testing and achievement on PISA. Future cycles of PISA are moving to a four-year cycle (previously a three-year cycle) and whether this will allow for a further change in the international testing window is uncertain. Countries moving their testing time will undoubtedly face administrative challenges as noted in Denner (2023) and the possibility of poor student response rates which has been a concern with springtime testing.

A key focus of this study was to examine, in depth, achievement on PISA between students tested in the Spring compared to the Autumn. The finding of a statistically significant lower percentage of low-achieving males on the mathematics scale in the Spring compared to the Autumn stands out as one with potentially important implications for interpreting the PISA results for Ireland. A key concern with conducting PISA testing in the Autumn was the potential impact of SLL. Very little research has been conducted in Ireland on the topic and the data from this study suggest that testing six weeks into the academic year appears not to be associated with change in overall performance on PISA. That said, the higher proportion of low-achieving males on the mathematics scale in the Autumn compared to the Spring merits further investigation. Of note also is the fact that in PISA 2022, when testing was in the Autumn and mathematics was the major domain, 18.5% of male students performed below Level 2, which was higher than the Spring administration of PISA 2012 (15.2%) when mathematics was also a major domain (Donohue et al., 2023a; Perkins et al., 2013). However, the lower response rates for males in the PISA 2022 study should be borne in mind for this comparison (Donohue et al., 2023b).

This study also points to a number of issues around the engagement of students in learning following the summer break, including the behavioural engagement of students. There was a statistically significantly larger percentage who reported skipping one to two days in the two weeks before the PISA test in the Autumn compared to the Spring. These behavioural issues have been noted during interviews with principals in a number of schools who administered the test in Autumn 2018 (Denner, 2022). Indeed, the opinion of some principals was that emotional and behavioural difficulties rather than learning loss may help to explain the poor performance of some male students at the beginning of the school year. Further research into both issues, as well the prevalence of absenteeism early in the school year, should be undertaken to ensure a better understanding of the dynamics of performance in PISA when the assessment takes place in the Autumn rather than the Spring.

References

Cohen, L., Mannion, L., & Morrison, K. (2011). *Research methods in education*. Routledge.

Cooper H., Nye B., Charlton K., Lindsay J., & Greathouse S. (1996). The effects of summer vacation on achievement test scores: A narrative and meta-analytic review. *Review of Educational Research, 66*, 227–268. https://doi.org/10.3102/00346543066003227

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. https://www.erc.ie/documents/vol40chp2.pdf

Cosgrove, J. & Hislop, H. (2011). *Looking back at PISA 2009 and forward to PISA 2012*. Presentation to the National Association of Principals and Deputy Principals' Conference. Killarney, October 2011.

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. https://doi.org/10.1186/2196-0739-2-2

Denner, S. (2022). *A comparison of outcomes from spring and autumn administrations of PISA Tests in Ireland* (unpublished doctoral dissertation). Dublin City University.

Denner, S. (2023). *PISA testing, spring vs autumn 2018: A feasibility study.* Educational Research Centre. https://www.erc.ie/wp-content/uploads/2023/12/37-ERC-Feasibility-Study-Report_A4_Web.pdf

Department of Education. (1993). *Transition year programme: Guidelines for schools.* Stationery Office. https://www.education.ie/en/Schools-Colleges/Information/Curriculum-and-Syllabus/Transition-Year-/ty_transition_year_school_guidelines.pdf

Department of Education and Science. (2004). *A Brief Description of the Irish Education System*. Communications Unit Department of Education and Science. https://www.gov.ie/en/policy/655184-education/#post-primary-education

DES [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*. https://assets.gov.ie/24521/9e0e6e3887454197a1da1f9736c01557.pdf

DES. (2012). *Circular No. 0025/2012. Implementation of the national literacy and numeracy strategy*. https://circulars.gov.ie/pdf/circular/education/2012/25.pdf

DES. (2017a). *National strategy: Literacy and numeracy for learning and life 2011-2020. Interim review 2011-2016. New targets: 2017-2020.* https://www.education.ie/en/Publications/Education-Reports/pub_ed_interim_review_literacy_numeracy_2011_2020.PDF

DES. (2017b). *DEIS plan 2017: Delivering Equality Of Opportunity In Schools*. https://www.education.ie/en/Publications/Policy-Reports/DEIS-Plan-2017.pdf

DES. (2017c). *STEM Education: Policy statement 2017 - 2026*. https://www.education.ie/en/The-Education-System/STEM-Education-Policy/stem-education-policy-statement-2017-2026-.pdf

DES. (2021). *DEIS identification process.* https://www.education.ie/en/Schools-Colleges/Services/DEIS-Delivering-Equality-of-Opportunity-in-Schools-/DEIS-Identification-Process.pdf

Department of Education. (2023a). *Junior Cycle Profile of Achievement (JCPA) handbook for schools.* https://www.gov.ie/en/publication/a26b9-junior-cycle-profile-of-achievement-jcpa/

Department of Education. (2023b). *Summer Programme 2023.* https://www.gov.ie/en/publication/5d15a-summer-programme/

Department of Education. (2024). *Ireland's Literacy, Numeracy and Digital Literacy Strategy 2024-2033: Every Learner from Birth to Young Adulthood.* https://assets.gov.ie/293241/65e50c84-48fa-45da-b32a-a963fb7818ca.pdf

Donohue, B., Perkins, R., Walsh, T., O'Neill, B., Ó Duibhir, C., & Duggan, A. (2023a). *Education in a Dynamic World: the performance of students in Ireland in PISA 2022*. Educational Research Centre.

Donohue, B., Perkins, R., Millar, D., Walsh, T., & Delaney, M., with Karakolidis, A., & Taylor, K. (2023b). *PISA 2022: Non-response bias analysis for Ireland.* Educational Research Centre.

Ferraro, D., Kali, J., & Williams, T. (2009). *Program for International Student Assessment (PISA) 2003: U.S. Nonresponse bias analysis (NCES 2009-088).* National Centre for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC. https://nces.ed.gov/pubs2009/2009088.pdf

Hislop, H. (2011). *Teacher education and Ireland's national strategy to improve literacy and numeracy*. SCOTENS Annual Conference, Cavan. September 2011. https://scotens.org/docs/2011-Hislop-speech.pdf

IEA. (2016). *Help manual for the IDB Analyzer*. www.iea.nl/data

Jerrim, J. (2011). *England's "plummeting" PISA test scores between 2000 and 2009: Is the performance of our secondary school pupils really in relative decline?* DoQSS Working Paper No. 11-09. Department of Quantitative Social Science, Institute of Education, London. http://repec.ioe.ac.uk/REPEc/pdf/qsswp1109.pdf

Kuhfeld, M. (2019). Surprising new evidence on summer learning loss. *Phi Delta Kappan, 101(1)*, 25–29. https://doi.org/10.1177/0031721719871560

McKeown, C., Denner, S., McAteer, S., Shiel, G., with O'Keeffe, L. (2019). *Learning for the Future: The performance of 15-Year-olds in Ireland on reading literacy, science and mathematics in PISA 2018*. Educational Research Centre. https://www.erc.ie/wp-content/uploads/2020/07/B23321-PISA-2018-National-Report-for-Ireland-Full-Report-Web-4.pdf

Muthén, L. K., & Muthén, B. O. (2017). *Mplus user's guide. Eighth Edition*. Muthén & Muthén.

National Council for Curriculum and Assessment. (2024). *Junior Cycle Profile of Achievement.* https://www.curriculumonline.ie/junior-cycle/level-1-learning-programmes-(l1lps)/assessment-and-reporting/junior%20cycle%20profile%20of%20achievement/

OECD [Organization for Economic Co-operation and Development]. (1999). *Measuring student knowledge and skills: A new framework for assessment.* https://www.oecd-ilibrary.org/education/pisa_19963777

OECD. (2019a). *PISA 2018 assessment and analytical framework*. https://www.oecd-ilibrary.org/education/pisa_19963777

OECD. (2019b). *PISA 2018 Results (Volume I): What students know and can do.* https://www.oecd-ilibrary.org/education/pisa_19963777

OECD. (2019c). *PISA 2018 Results (Volume II): Where All Students Can Succeed.* Paris: https://www.oecd-ilibrary.org/education/pisa_19963777

OECD. (2019d). *PISA 2018 Results (Volume III): What School Life Means for Students' Lives*. https://doi.org/10.1787/acd78851-en.

OECD. (2020). *PISA 2018 technical report.* https://www.oecd-ilibrary.org/education/pisa_19963777

Opdenakker, M-C., Maulana, R., & den Brok, P. (2012). Teacher-student interpersonal relationships and academic motivation within one school year: developmental changes and linkage. *School Effectiveness and School Improvement*, *23,* 95–119. https://doi.org/10.1080/09243453.2011.619198

Ouweneel, E., Schaufeli, W. B., & Le Blanc, P. M. (2013). Believe, and you will achieve: Changes over time in self-efficacy, engagement, and performance. *Applied psychology: Health and well-being*, *5*, 225–247. https://doi.org/doi:10.1111/aphw.12008

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*. Educational Research Centre.

Perkins, R. (2015). The Role of Engagement and Test-Taking Behaviour in PISA 2012 in Ireland. *Irish Journal of Education*, *40,* 45-67. https://www.erc.ie/documents/vol40chp3.pdf

Schmidt, W. H., Zoido, P., & Cogan, C. (2013). *Schooling Matters: Opportunity to learn in PISA 2012*. OECD Education Working Papers No. 95. https://dx.doi.org/10.1787/5k3v0hldmchl-en

What Works Clearinghouse. (2014). *Procedures and standards handbook.* Version 3.0. Washington: US Department of Education, Institution of Educational Sciences. https://ies.ed.gov/ncee/wwc/docs/referenceresources/wwc_procedures_v3_0_standards_handbook.pdf

Wise, S. L., Ma, L., Kingsbury, G. G., & Hauser, C. (2014). *An investigation of the relationships between time of testing and test-taking effort.* Northwest Evaluation Association (NWEA).

Wu, M. (2005). The role of plausible values in large-scale surveys. *Studies in Educational Evaluation*, *31*, 114–128. https://doi.org/10.1016/j.stueduc.2005.05.005

APPENDIX TABLE 1

Coefficients for the interaction between 'session' and other level 1 learning outcome variables on final models presented in Table 8

	Reading			Mathematics			Science		
Session[by]...	**Coefficient**	**SE**	**p-value**	**Coefficient**	**SE**	**p-value**	**Coefficient**	**SE**	**p-value**
Text 101	5.24	4.52	0.246	a	a	a	a	a	a
Text 100	7.67	5.83	0.188	a	a	a	a	a	a
Write a summary of the book/ chapter (Ref=No)	-3.40	4.95	0.492	a	a	a	a	a	a
Compare the content of the book or the chapter with your own experience (Ref=No)	2.89	4.32	0.503	a	a	a	a	a	a
Select a passage you liked or disliked and explain why (Ref=No)	3.53	4.13	0.393	a	a	a	a	a	a
How to use keywords when using a search engine (Ref=No)	2.19	4.72	0.644	a	a	a	a	a	a
To understand the consequences of making information publicly available online (Ref=No)	4.30	6.03	0.476	a	a	a	a	a	a
How to use the short description below the links in the list of search results (Ref=No)	7.30	4.49	0.104	a	a	a	a	a	a
How to detect whether the information is subjective or biased (Ref=No)	8.09	4.28	0.059	a	a	a	a	a	a
Teachers' stimulation of reading engagement	-1.20	2.23	0.591	a	a	a	a	a	a
Teacher-directed instruction	0.59	1.98	0.767	a	a	a	a	a	a
*Skipped a whole day in 2 weeks before PISA testing (Ref=No whole days skipped)^*									
1-2 whole days skipped	a	a	a	-0.12	5.66	0.983	3.76	6.64	0.572
3 or more whole days skipped	a	a	a	-17.25	10.62	0.104	**-21.54***	10.77	0.046

	Reading			Mathematics			Science		
Session[by]...	Coefficient	SE	p-value	Coefficient	SE	p-value	Coefficient	SE	p-value
*Arrived late to school in 2 weeks before PISA testing (Ref=No days late)^*									
1-2 late days	1.93	5.34	0.718	-8.06	5.31	0.129	-7.64	5.84	0.190
3 or more late days	-9.67	8.18	0.237	4.21	9.03	0.641	0.23	10.82	0.983
Understanding and remembering	-4.27	2.23	0.056	a	a	a	a	a	a
Summarising strategies	-1.58	2.15	0.460	a	a	a	a	a	a
Assessing credibility	0.29	1.94	0.880	a	a	a	a	a	a
Perceived competence in reading	1.43	1.98	0.471	a	a	a	a	a	a
Perceived difficulty in reading	**-4.54***	2.27	0.046	a	a	a	a	a	a
Enjoyment of reading	-0.39	1.99	0.845	a	a	a	a	a	a
General self-efficacy (resilience)	a	a	a	0.06	2.93	0.985	-4.11	3.29	0.213
Attitudes towards competition	a	a	a	3.16	2.44	0.196	-0.41	2.74	0.880
Motivation to master tasks	a	a	a	1.13	2.76	0.682	-3.22	2.92	0.270
Life satisfaction	-1.37	0.93	0.141	a	a	a	-0.65	1.20	0.589
Eudemonia: meaning in life	-1.45	2.34	0.536	-0.82	2.53	0.748	-4.16	2.69	0.122
Subjective wellbeing: positive affect	-0.63	2.21	0.774	1.46	2.82	0.606	-0.26	2.81	0.927
Attitude towards school/learning activities	a	a	a	1.01	2.56	0.694	-3.38	2.87	0.240

Note: [a] not included in the final model. ^Missing indicator included to retain data.

Bold indicates statistically significant difference with *indicating .05 level and **indicating .01 level.