Do changes in instructional time, professional development of teachers and age of students explain changes in reading comprehension at the country level? An exploration of PIRLS 2006 and 2016

Van Damme Jan[1], Bellens Kim[2], Tielemans Kelly[3], Wim Van den Noortgate[4]

[1] *KU Leuven, Belgium*
E-mail: jan.vandamme@kuleuven.be
ORCID https://orcid.org/0000-0002-2692-9120

[2] *KU Leuven, Belgium*
E-mail: kim.bellens@kuleuven.be
ORCID https://orcid.org/0000-0001-9695-5167

[3] *KU Leuven, Belgium*
E-mail: kelly.tielemans@kuleuven.be
ORCID https://orcid.org/0000-0003-4139-7262

[4] *KU Leuven, Belgium*
E-mail: wim.vandennoortgate@kuleuven.be
ORCID https://orcid.org/0000-0003-4011-219X

DOI: 10.26907/esd14.2.02

Abstract
Several international large-scale assessments in education take place every 3 to 5 years. PIRLS (the Progress in International Reading Literacy Study) takes place every 5 years. If data from only one moment are considered, it is difficult to explain differences in achievement between countries. But we can also consider data from two (or more) years and focus on changes over time within each of the participating countries. Following countries over a certain period in such a longitudinal approach facilitates causal inferences on the effects of characteristics of educational systems.
In this study, we investigate the effects of instructional time for language and reading and of amount of professional development of teachers on reading comprehension in Grade 4. We also study the effect of changes in the average age of students. By accounting for the effects of age and schooling we come to a slightly different ranking of countries in PIRLS. We use a difference-in-difference approach with correlation matrices. One of our conclusions is that professional development of teachers has an effect on the achievement level of students.
Keywords: reading comprehension, instructional time, professional development, PIRLS 2006, PIRLS 2016, Grade 4, age.

Объясняются ли изменения в понимании прочитанного на уровне стран изменениями в учебном времени, профессиональном развитии учителей и возрасте учащихся? Исследования PIRLS 2006 и 2016

Ван Дамм Ян[1], Белленс Ким[2], Тилеманс Келлиу[3], Вим Ван ден Нортгейт[4]

[1] *Лёвенский католический университет, Бельгия*
E-mail: jan.vandamme@kuleuven.be
ORCID https://orcid.org/0000-0002-2692-9120

[2] *Лёвенский католический университет, Бельгия*
E-mail: kim.bellens@kuleuven.be
ORCID https://orcid.org/0000-0001-9695-5167

[3] *Лёвенский католический университет, Бельгия*
E-mail: kelly.tielemans@kuleuven.be
ORCID https://orcid.org/0000-0003-4139-7262

[4] *Лёвенский католический университет, Бельгия*
E-mail: wim.vandennoortgate@kuleuven.be
ORCID https://orcid.org/0000-0003-4011-219X

DOI: 10.26907/esd14.2.02

Аннотация
Ряд международных сравнительных исследований в образовании проводится циклами раз в 3-5 лет. Международное исследование качества чтения и понимания текста PIRLS проводится раз в 5 лет. При рассмотрении данных только одного цикла представляется затруднительным объяснить различия в уровне достижений между странами. Но также возможно рассмотрение двух (или более) циклов и анализ изменений во времени для каждой из стран-участниц. В таком случае рассматривается динамика по странам за определенный период, и данный лонгитюдный подход позволяет делать каузальные выводы относительно эффектов характеристик образовательных систем.
В данной статье рассмотрены эффекты учебного времени (изучение языка и чтение) и уровня профессионального развития учителей на понимание прочитанного в 4-х классах. Также изучен эффект изменения среднего возраста учащихся. При учете эффектов возраста и обучения сформирован немного другой рейтинг стран по результатам исследования PIRLS. Использован подход «разность разностей» с корреляционными матрицами. Один из выводов заключается в выявлении эффекта профессионального развития учителей на уровень образовательных результатов учащихся.
Ключевые слова: понимание прочитанного, учебное время, профессиональное развитие, PIRLS-2006, PIRLS-2016, 4 класс, возраст.

Introduction

By many people, international large-scale assessments are considered as important because of their ranking of countries based on the average overall achievement of their students at a certain moment. However, for researchers the change in average achievement over time is more valuable. Are the results of the students of a specific country in a new assessment better (or worse) than in a previous assessment? And even more interest lies in answering the question as to whether we can explain why the results of a country are improving (or declining). According to Gustafsson (2007) this is possible by studying the relation between changes in average achievement over time and changes in characteristics of the educational systems, as this permits to draw causal conclusions on generic factors that influence educational outcomes. Gustafsson (2007) considered changes in factors as the average age of the participating students, the socio-economic status of the family of students (SES) and class size, i.e., the number of students in a class.

Indeed, when comparing the average achievement of educational systems at a certain moment two major problems can lead to wrong conclusions. First of all, in studies at a certain moment it is impossible to consider all relevant variables. And in such a case correlations at a certain moment are not enough to make causal inferences. By considering a group of countries and changes over time in each of these countries, many relevant variables (e.g. characteristics of the socio-economic development and of the cultural and historical background) can be assumed to be more or less stable. So, the problem of omitting relevant variables is solved to a certain extent. Secondly, in educational research the phenomenon of 'reversed causality' is not exceptional. Suppose that teachers can decide themselves whether they participate in professional development activities or not. (Or that the school principal takes that decision.) One can expect that the teachers with less well performing pupils will engage in more professional development. Then, the study

can indicate that more professional development goes together with lower achievement at the end of the school year. But it would be wrong to conclude that more professional development leads to lower achievement. The reverse is true: lower achievement has led to more professional development. Selection bias can thus lead to wrong conclusions. One solution for that problem can be to consider the effect of, e.g., professional development at a higher level: at the school or even at the country level instead of at the individual teacher level. If professional development is promoted in the whole educational system, we hope to see an increase in the average achievement level of that system.

Research aims

This study is based on the Progress in International Reading Literacy Study (PIRLS) and especially on two waves: PIRLS 2006 and 2016. (See table 1 for the participating countries which will be considered in part of our study.) PIRLS is organized by the International Association of Educational Achievement (IEA) and maps students' reading comprehension together with a large amount of context factors.

The targeted student population in PIRLS are Grade 4 students. To avoid testing very young children, PIRLS has a policy that the average age of the children in the tested grade should not be below 9.5 years (Martin et al., 2007). Unavoidably, the average age of the pupils tested can be different across countries (because of other regulations and traditions) and even between different years within one country (because of a reform of the educational system) (see Table 1).

We firstly focus on some indicators of instructional quality and teacher quality, in a broad sense. Our choice is partly influenced by the ongoing discussion in our 'country', Flanders ('Belgium Flemish' in the IEA-terminology) on the reasons for the decline in achievement in PIRLS between 2006 and 2016.

Instructional time has always been considered to be an important factor in explaining differences in achievement. This was already the case in the well-known model of school learning of Carroll (1963) and it is still the case in the rather recent dynamic model of educational effectiveness (Creemers & Kyriakides, 2008). For an overview of the relevant research, see Scheerens (2016). Taking into account the available data in PIRLS we distinguish between two types of instructional time: the percentage of the total instructional time spent to language instruction in general and the percentage of instructional time spent to reading instruction. As for the teacher variables we choose to focus on the quantity of their professional development in relation to reading instruction. Literature indicates that the extent of professional development during the career of a teacher can be an effective factor in raising educational outcomes (for recent studies: see Gustafsson and Nilsen (2017), and Nilsen, Gustafsson, and Yang Hansen (2017)).

In the second part we focus on the average age of the students. In PIRLS mostly it is Grade 4 students that are participating. Typically there are differences in average age between countries. And in some countries there are also differences in average age over the measurement period. The effect of age has been suggested in many studies, especially in Gustafsson (2007). But he considered other subjects, i.e., mathematics and sciences. Van Damme, Liu, Vanhee, and Pustjens (2010) have shown that differences in average age explained changes in reading comprehension in PIRLS between 2001 and 2006. Liu, Bellens, Gielen, Van Damme, and Onghena (2014) have shown this was also the case between 2006 and 2011. We wish to see whether changes in achievement between 2006 and 2016 are caused by age differences.

But also at a given time there are age differences between countries. Rindermann (2007) took notice of the fact that international student assessment and intelligence test studies suffered from different problems of sample representativeness and statistical

methods. In particular the comparability across countries was questionable. Therefore, he made suggestions for adjustments that could improve representativeness and comparability. Based on several intelligence test studies and student assessment studies, Rindermann observed that the gain of students of a specific country in one year was on average about 42 points (referring to a scale with a mean of 500 and a standard deviation of 100, as is the case in PIRLS). A Swedish study compared Grade 3 and 4 in PIRLS 2001 and showed that in Sweden about a third of this progress is attributable to an extra year of age and two thirds to an extra year of schooling (Gustafsson, 2007). Based on these findings, one can subtract 14 points from the score of a country whose students are one year older (as we have done in Van Damme et al., 2010).

The ranking of countries obtained after such an age correction is only justifiable when we have reasons to assume that having become one year older has more or less similar effects in each country. To test this assumption, we investigate the size of the age and schooling effect by means of the regression discontinuity approach in three countries who participated in PIRLS 2001 or 2006 with two adjacent grades. We develop a new ranking of the countries participating in 2006 and also of those participating in 2016.

In the last part we use these new rankings to study again whether changes in achievement between 2006 and 2016 can be caused by the changes in the earlier considered instructional and teacher variables. Because of the difference-in-difference approach we use, we expect a confirmation of the results obtained in the first part.

Research questions

RQ 1 Do changes in the percentage of instructional time for (a) language instruction, and (b) reading instruction explain changes in achievement at the country level?

RQ 2 Do changes in amount of professional development of teachers explain changes in achievement at the country level?

RQ 3 Do changes in average age at the country level explain changes in achievement at the country level?

RQ 4 What is the average achievement (and the ranking) of countries in 2006 and in 2016 after correcting for age differences across countries at each moment?

RQ 5 What is the answer to research question 1 and 2 after correcting for age differences between countries at each moment (as performed in research question 4)?

Data and methodology

Data

Data from PIRLS 2006 and 2016 were used to answer the research questions. In PIRLS 2006, 46 countries participated (Mullis, Martin, Kennedy, & Foy, 2007), whilst in PIRLS 2016 data were collected in 61 countries (Mullis, Martin, Foy, & Hooper, 2017); 35 countries participated in both data cycles, which gives researchers the opportunity to work with longitudinal data at the country level.

Prior to the analyses, we deleted six countries from our dataset as their data of 2006 and 2016 are not comparable (Mullis et al., 2017): in Poland and South Africa the target population changed between 2006 and 2016; in Israel, Morocco, Qatar, and Kuwait, some changes were made to their trend instruments. After these exclusions, our dataset consisted of 29 countries. Table 1 gives an overview of the countries included in the analyses, together with their average reading achievement score and average age. Norway was included twice because in both years data were available for Grade 4 and Grade 5,

which gives a total of 30 units to be included in the analyses. However, to guarantee an equal contribution of each educational system in the results of our analyses, the data from Norway Grade 4 and Grade 5 were weighted with 0.5, whereas all other countries were weighted with a weight of 1. In Table 2 and following tables we include results for Norway in general, based on the two different grades, which leads to a total amount of countries included in the analysis of 29.

Variables

Dependent variables

*Change in reading achievement between 2006 and 2016 (**Read_Diff**).* The change in the average reading achievement score between 2006 and 2016 at the country level was used as one of our dependent variables. Read_Diff was calculated by means of subtracting countries' average reading achievement score in 2006 (Read_06) from the score in 2016 (Read_16). Data of Read_06 and Read_16 used in our analyses can be found in the international PIRLS 2006 and 2016 reports and are shown in Table 1. These scores take into account the five plausible values, as they are considered as separate estimates of each student's underlying reading ability (Martin, Mullis, & Hooper, 2017).

*Change in reading achievement between 2006 and 2016 after correcting for age differences (**ReadCorr_Diff**).* Next to Read_Diff, we used the change in reading achievement between 2006 and 2016 after we corrected for age differences between countries as a dependent variable. In a first step, Read_06 and Read_16 (cf. supra) form the foundation for the calculation of the average reading achievement score with correction for cross country differences in age at a certain moment, i.e., in 2006 (ReadCorr_06) and 2016 (ReadCorr_16). An explanation of how the correction for age differences was performed can be found in the 'Methods and procedures' section. In a second step, ReadCorr_Diff was calculated by subtracting ReadCorr_06 from ReadCorr_16.

Independent variables

*Change in percentage of instructional time for language instruction (**LanguageTime_Diff**).* Data from the international PIRLS 2006 and 2016 reports were used to obtain data on the percentage of instructional time for language instruction in 2006 (LanguageTime_06) and 2016 (LanguageTime_16). Data on LanguageTime_Diff was obtained by subtracting LanguageTime06 from LanguageTime_16.

*Change in percentage of instructional time for reading instruction (**ReadingTime_Diff**).* Data from the international PIRLS 2006 and 2016 reports were used to obtain data on the % of instructional time for reading instruction in 2006 (ReadingTime_06) and 2016 (ReadingTime_16). Data on ReadingTime_Diff was obtained by subtracting ReadingTime06 from ReadingTime_16.

Table 1

Countries participating in PIRLS 2006 and 2016, average reading achievement and average age (N=30)

Country	Read_06	Read_16	Age_06	Age_16
Austria (AUS)	538	541	10.3	10.3
Belgium, Flemish (BFL)	547	525	10.0	10.1
Belgium, French (BFR)	500	497	9.9	10.0
Bulgaria (BGR)	547	552	10.9	10.8
Canada, Ontario (COT)	555	544	9.8	9.8
Canada, Quebec (CQU)	533	547	10.1	10.1

Country	Read_06	Read_16	Age_06	Age_16
Chinese Taipei (TWN)	535	559	10.1	10.1
Denmark (DNK)	546	547	10.9	10.8
England (ENG)	539	559	10.3	10.3
France (FRA)	522	511	10.0	9.8
Georgia (GEO)	471	488	10.1	9.7
Germany (DEU)	548	537	10.5	10.3
Hong Kong, SAR (HKG)	564	569	10.0	9.9
Hungary (HUN)	551	554	10.7	10.6
Iran, Islamic Republic of (IRN)	421	428	10.2	10.2
Italy (ITA)	551	548	9.7	9.7
Latvia (LVA)	541	558	11.0	10.9
Lithuania (LTU)	537	548	10.7	10.8
Netherlands (NLD)	547	545	10.3	10.1
New Zealand (NZL)	532	523	10.0	10.1
Norway, Grade 4 (NOR4)	498	517	9.8	9.8
Norway, Grade 5 (NOR5)	541	559	10.8	10.8
Russian Federation (RUS)	565	581	10.8	10.8
Singapore (SGP)	558	576	10.4	10.4
Slovak Republic (SVK)	531	535	10.4	10.4
Slovenia (SVN)	522	542	9.9	9.9
Spain (ESP)	513	528	9.9	9.9
Sweden (SWE)	549	555	10.9	10.7
Trinidad And Tobago (TTO)	436	479	10.1	10.2
United States (USA)	540	549	10.1	10.1

Change in professional development of teachers (***Develop_Diff***). The time spent in the past two years on formal professional development activities that dealt directly with reading or teaching reading in 2006 (Develop_06) and 2016 (Develop_16), was calculated using the variable ATBGSEMI and ATBG06 in the PIRLS 2006 and 2016 dataset, respectively. Teachers could indicate the amount of formal professional development activities during the last two years by means of a 5-point Likert scale ranging from (1) none to (5) 35 hours or more. These variables were considered as continuous. TCHWGT was used in computing countries' average professional development of teachers. Develop_Diff was calculated by subtracting Develop_06 from Develop_16.

*Change in average students' age (**Age_Diff**).* Change in average students' age between 2006 and 2016 was calculated by subtracting average students' age in 2006 (Age_06) from average students' age in 2016 (Age_16). Data on Age_06 and Age_16 can be found in the international PIRLS 2006 and 2016 reports and in Table 1.

Descriptive statistics of all variables included can be found in Table 2.

Table 2

Descriptive statistics

	N	Mean	SD	MIN	MAX
Read_06	29	529.56	34.45	420.93	565.00
Read_16	29	536.70	31.74	428.00	580.99

	N	Mean	SD	MIN	MAX
Read_Diff	29	7.14	13.52	-21.98	43.81
ReadCorr_06	29	527.48	32.90	420.93	568.91
ReadCorr_16	29	535.55	29.59	428.00	576.37
ReadCorr_Diff	29	8.07	13.42	-24.46	40.54
LanguageTime_06	29	31.36	5.25	22.00	39.00
LanguageTime_16	29	29.79	4.94	22.00	41.00
LanguageTime_Diff	29	-1.57	3.39	-12.00	4.00
ReadingTime_06	29	20.72	6.43	9.00	32.00
ReadingTime_16	29	19.53	7.42	9.00	39.00
ReadingTime_Diff	29	-1.19	4.86	-11.00	10.00
Develop_06	29	2.67	0.57	1.74	4.32
Develop_16	29	2.90	0.53	1.95	3.85
Develop_Diff	29	0.23	0.45	-0.85	1.01
Age_06	29	10.28	0.37	9.70	10.97
Age_16	29	10.24	0.37	9.74	10.90
Age_Diff	29	-0.04	0.10	-0.34	0.09

Methods and procedures

Data were analyzed by means of correlation matrices, in combination with a difference-in-difference approach, a method often used by Gustafsson (2007; 2013). First, correlations between all variables under study were calculated, taking into account data at a certain moment (2006 and 2016) as well as difference scores (2016-2006). The results are represented in tables and in scatterplots.

PIRLS and other IEA-studies are grade-based. In PIRLS most participants are Grade 4 students. Typically there are differences in average age between countries. Several studies indicated that the variations in students' age between countries have a troublesome influence on the comparability of the achievement scores of these countries (Rindermann, 2007; Gustafsson, 2007). To take into account the differences in age we applied a correction to the average achievement scores at each moment in research question 4. In an earlier publication we have treated this problem, comparing PIRLS 2001 and 2006 (Van Damme et al., 2010). We summarize and actualize our earlier publication, partly by citing from it.

Iceland and Norway, as an additional effort for their own purposes, administered PIRLS 2006 to small samples of their fifth-grade students, while Sweden participated in PIRLS 2001 with grades three and four. These sampling designs enabled us to estimate age and schooling effects by means of a regression-discontinuity approach described by Luyten (2006). We explain the most important part of the procedure while illustrating it with Swedish data.

For a *non-technical explanation*, we refer the reader to Figure 1, which shows the relation between the achievement and the month of birth within each grade. Within each grade there are older and younger students (referring to the month of birth) but all students have had the same number of years of schooling. The different outcomes between older and younger students within a grade allow us to estimate how big the age effect is. Figure 1 also shows a gap in achievement, a regression discontinuity (between the two regression lines), between the oldest students in grade 3 and the youngest students in grade 4. This gap allows us to estimate the size of the effect of one more year of schooling.

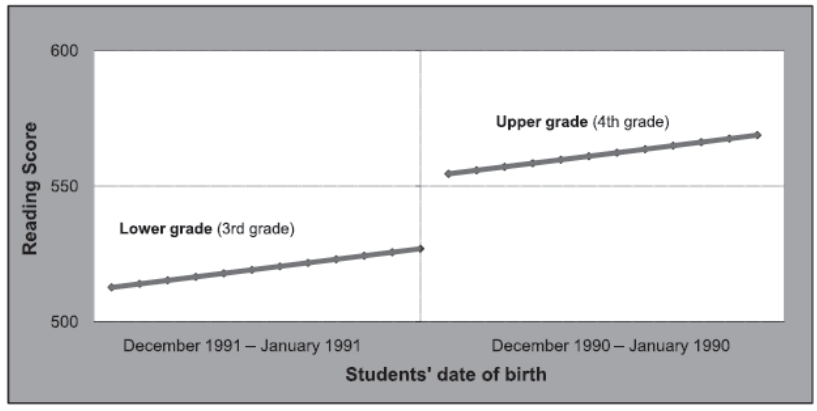

Figure 1. Relationship between date of birth and reading achievement in Sweden, estimated by means of the regression model in Equation 1

And now the *technical explanation* of the procedures used. First we merged the files of the Swedish data from Grade 3 and 4 in PIRLS 2001. Grade-level was re-coded to assign scores of zero to students in the lower grade and scores of one to students in the higher grade. The variable that denotes a student's date of birth is based on the year and month of birth. Each date was transformed into a single number. For example, a student born in March 1990 received a score of 90.25, and a student born in April 1990 received a score of 90.33. Students born in January 1991 or later were in the lower grade, and the ones born earlier were in the higher grade. The cut-off value (91.08) was then subtracted from these scores, giving each of the oldest students in the lower grade (the comparison group) a positive score, students of the upper grade a negative score. Table 3 illustrates the transformation of the original birth dates to the scores used in the analyses.

Table 3

Range of birth dates (cut-off = 91.08) for Swedish students in PIRLS 2001

	Month and year	In decimals	Minus cut-off
Upper grade	January 1990	90.08	-1.00
	February 1990	90.17	-0.92
	March 1990	90.25	-0.83
	April 1990	90.33	-0.75
	May 1990	90.42	-0.67
	June 1990	90.50	-0.58
	July 1990	90.58	-0.50
	August 1990	90.67	-0.42
	September 1990	90.75	-0.33
	October 1990	90.83	-0.25
	November 1990	90.92	-0.17
	December 1990	91.00	-0.08
Lower grade	January 1991	91.08	0.00
	February 1991	91.17	0.08
	March 1991	91.25	0.17
	April 1991	91.33	0.25

Month and year	In decimals	Minus cut-off
May 1991	91.42	0.33
June 1991	91.50	0.42
July 1991	91.58	0.50
August 1991	91.67	0.58
September 1991	91.75	0.67
October 1991	91.83	0.75
November 1991	91.92	0.83
December 1991	92.00	0.92

Next, we explain the regression-discontinuity model expressed in equation (1). The coefficient β_1 expresses the effect of age and β_2 the effect of one extra year of schooling. In this equation, the effect of age is assumed linear and identical in both grades (i.e., no interaction between age and grade).

$$Y_i = \beta_0 + \beta_1(x_i - x_0) + \beta_2 z_i + R_i \quad (1)$$

Where:

Y_i = *reading score*
x_i = *date of birth, pupil i*
x_0 = *cut-off value (here 91.08, referring to January, 1991)*
z_i = *grade, student i (0 if lower grade; 1 if upper grade)*
β_0 = *parameter for comparison group intercept at cut-off*
β_1 = *age effect*
β_2 = *effect of being in the upper grade (i.e., having received an extra year of schooling)*
R_i = *random residual*

To see whether the results of the Swedish study are generalizable over countries, we compare them with the results from Iceland and Norway. These countries included Grade 4 and 5 in PIRLS 2006. The overall progress that students from Sweden, Iceland and Norway made in one year was respectively 41.71, 38.65 and 42.28. On average, the expected progress in one year can be set to 40.88. This is close to the findings of Rindermann (2007). The results on the grade effect and the age effect are included in Table 4, while Figure 1 and 2 give a visual perception of the results for Sweden and Iceland.

Table 4

Grade and age effects for Sweden, Iceland and Norway

	Grade effect	Age effect
Sweden	26.34***	15.54**
Iceland	10.29 (*one-tailed)	28.00***
Norway	12.09	30.18***

*** $p < .001$; ** $p < .01$; * $p < .05$

From Table 4 it is obvious that the age effect is larger than the grade effect in Iceland and Norway. In contrast with our publication of 2010 in which our correction of the age differences in PIRLS 2006 was only based on the Swedish data, we now make a correction for PIRLS 2006 and 2016, based on the average of the age effect of the three countries, i.e.,

$$(15.54 + 28.00 + 30.18)/3 = 24.57$$

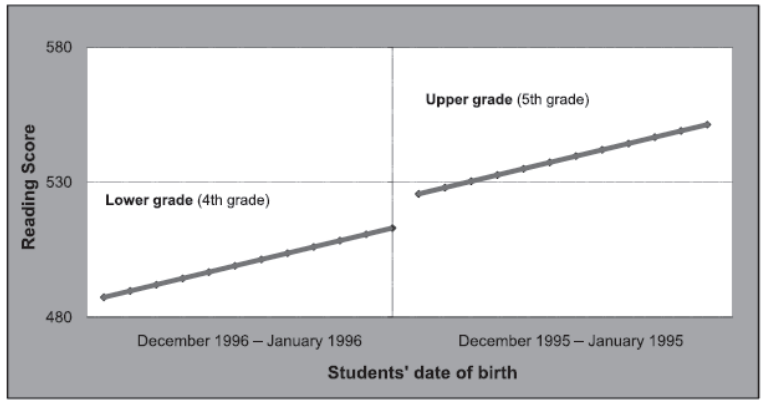

Figure 2. Relationship between date of birth and reading achievement in Iceland, estimated by means of the regression model in Equation 1

The level of significance (α) chosen in this study is .05. For reasons of prudence, 2-tailed tests of significance were used in analyzing the results of the correlational data.

Results

RQ 1: The effect of the percentage of total instructional time spent on (a) language instruction, and (b) reading instruction on achievement

A. Effect of percentage of total instructional time spent on language instruction on achievement

The correlation between the variables is given in Table 5 and presented in Figure 3. In Table 5 one can see that the correlation of the levels of achievement of countries at the two moments is high (r = .92). Also the correlation between the average instructional time at the two moments is high (r = .78). But we are more interested in the relation between instructional time and achievement. When considering the correlation at one measurement moment, the correlation between % time for language instruction and achievement is not significant (r = -.14 in 2006 and r = .04 in 2016). When considering Figure 3 we can observe that the majority of the countries have decreased the time for language instruction. The overall correlation between changes in time spent and changes in language instructional time is positive (r = .16), but not significant (p = .39).

Table 5

Correlation between (changes in) the percentage of the total instructional time spent on language instruction and (changes in) reading achievement in PIRLS 2006 and 2016 (N=29)

	Read_06	Read_16	Read_Diff	LanguageTime_06	LanguageTime_16
Read_06	1.00				
Read_16	.92***	1.00			
Read_Diff	-.39*	.01	1.00		
LanguageTime_06	-.14	-.23	-.17	1.00	
LanguageTime_16	.07	.04	-.07	.78***	1.00
LanguageTime_Diff	.32	.42*	.16	-.41*	.25

* $p < .05$; *** $p < .001$ (2-tailed)

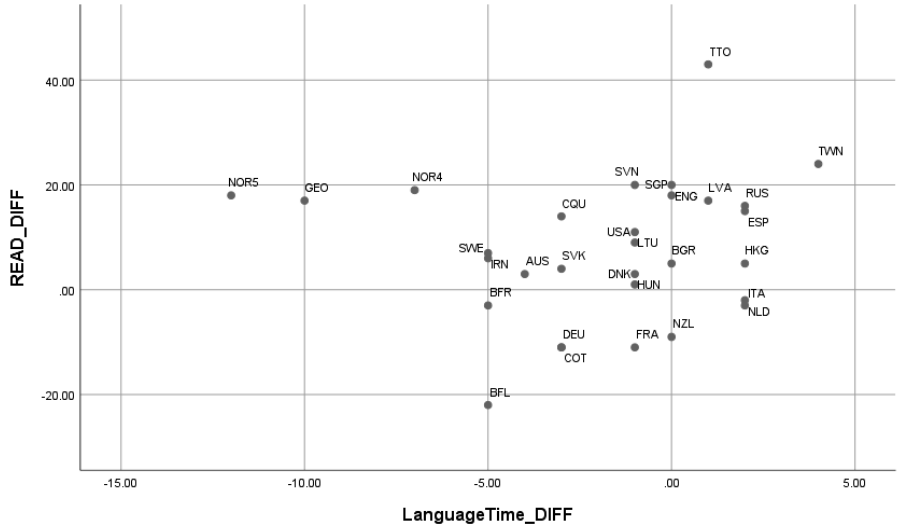

Figure 3. Relationship between changes in the percentage of the total instructional time spent on language instruction and changes in reading achievement in PIRLS 2006 and 2016

B. Effect of percentage of total instructional time spent on reading instruction on achievement

When studying Table 6 and Figure 4, we do not find a significant relation between instructional time and achievement: no correlation at each moment (r = -.23 in 2006 and r = -.14 in 2016) and no correlation between the change variables (r = .05; p = .78).

Table 6

Correlation between (changes in) the percentage of the total instructional time spent on reading instruction and (changes in) reading achievement in PIRLS 2006 and 2016 (N=29)

	Read_06	Read_16	Read_Diff	ReadingTime_06	ReadingTime_16
Read_Diff	-.39*	.01	1.00		
ReadingTime_06	-.23	-.19	.15	1.00	
ReadingTime_16	-.19	-.14	.17	.74***	1.00
ReadingTime_Diff	.01	.03	.05	-.22	.50**

** p < .01; *** p < .001 (2-tailed)

The correlations across all countries are presented in Table 7 and shown in Figure 5. At both measurement moments, there is no significant correlation between the time spent on professional development and the achievement level of the students (r = -.22 in 2006 and r = -.04 in 2016). But the changes in achievement at the country level are significantly linked to changes in the amount of professional development (r = .41; p < .05), as illustrated in Figure 5.

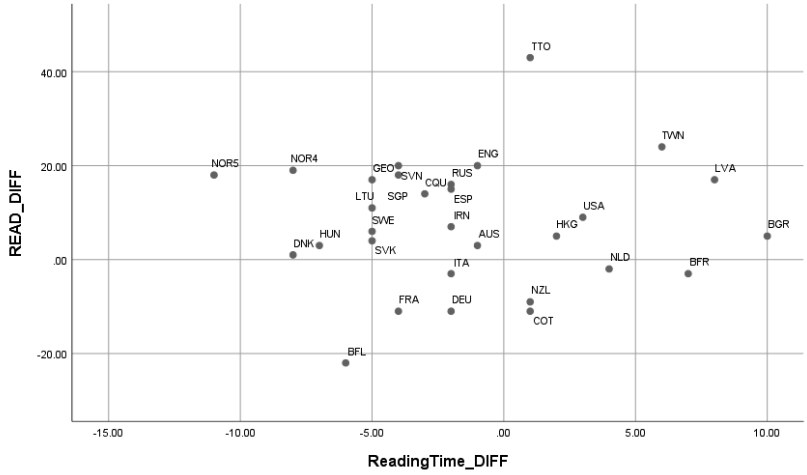

Figure 4. Relationship between changes in the percentage of the total instructional time spent on reading instruction and changes in reading achievement in PIRLS 2006 and 2016

RQ 2: The effect of amount of professional development in reading instruction on achievement

Table 7

Correlation between (changes in) teachers' professional development and (changes in) reading achievement in PIRLS 2006 and 2016 (N=29)

	Read_06	Read_16	Read_Diff	Develop_06	Develop_16
Read_Diff	-.39*	.01	1.00		
Develop_06	-.22	-.15	.20	1.00	
Develop_16	-.26	-.04	.56**	.66***	1.00
Develop_Diff	-.03	.14	.41*	-.48**	.35

* $p < .05$; ** $p < .01$; *** $p < .001$ (2-tailed)

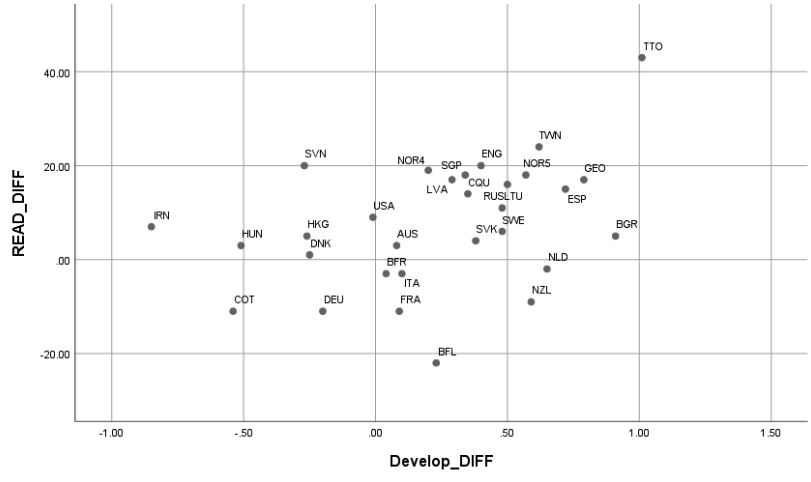

Figure 5. Relationship between changes in teachers' professional development and changes in reading achievement in PIRLS 2006 and 2016

RQ 3: The effect of age on achievement

In Figure 6, we present a scatter diagram of the age and achievement differences between 2006 and 2016 at the country level. Whereas in 2016, countries with older students have a higher level of reading comprehension (r = .38, see Table 8), no significant correlation exists between average age and average achievement in 2006 (r = .27). Furthermore, the change in age is not significantly linked to change in achievement (r = .10).

Table 8

	Read_06	Read_16	Read_Diff	Age_06	Age_16
Read_Diff	-.39*	.01	1.00		
Age_06	.27	.35	.13	1.00	
Age_16	.28	.38*	.16	.97***	1.00
Age_Diff	.04	.09	.10	-.16	.11

Correlation between (changes in) students' age and (changes in) reading achievement in PIRLS 2006 and 2016 (N=29)

* $p < .05$; *** $p < .001$ (2-tailed)

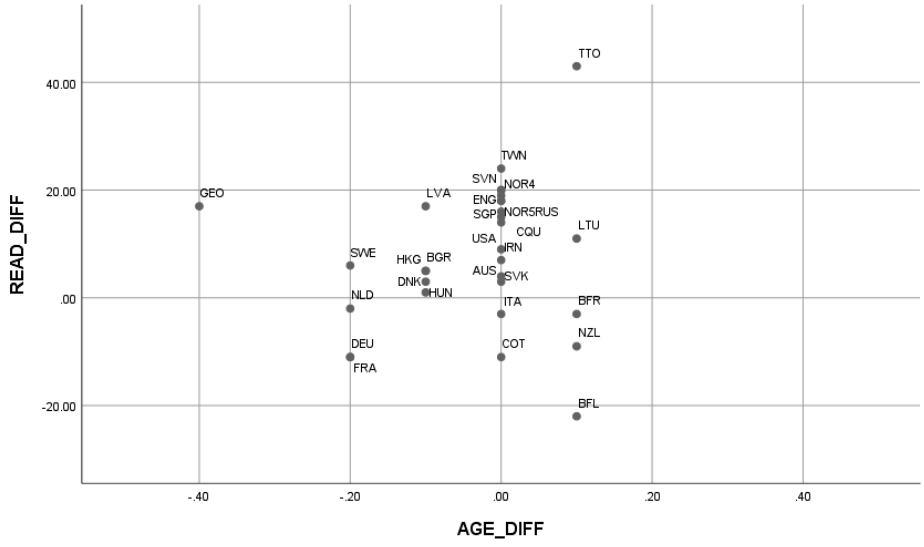

Figure 6. Relationship between changes in students' age and changes in reading achievement in PIRLS 2006 and 2016

RQ 4: Average achievement and ranking of countries in 2006 and in 2016 after correcting for age differences

The results are included in Tables A2 and A3 in the Appendix. For each year, we included all the participating countries and regions, even those which did not participate in the other year. Overall, many countries keep more or less the same place in the ranking after correcting for age differences, especially in 2016. The differences in age across

countries were smaller in 2016 than in 2006. And of course, countries with older students (in comparison to other countries) go down in the ranking, while countries with younger students go up. In 2006, Russia loses some places in the ranking. In the official ranking Russia was the best performing country. In our new ranking Russia is only on the 8th place. So one of the reasons why Russia was 1st, was the older age of its students.

Let us go a bit in more detail in relation to the results of 2016. In the new ranking for this year, see Table A3, Russia moved from the 2nd place in the official ranking to the 4th place in the new ranking. Moscow, which participated in PIRLS 2016 as a separate region, takes the 1st rank in both rankings. Most other countries with relatively older students, lost several places in the new ranking, e.g. Latvia (-12 places), Finland (-8 places), Bulgaria (-13 places), Lithuania (-15 à 16 places) and Denmark (-14 à 15 places)[1]. Some countries with younger participating students sometimes gained many places in the ranking, e.g. Italy (+ 12 à 13 places), but this was not the case for e.g. the less well performing country Georgia.

We have also considered a way of correcting for differences in schooling. Indeed some countries participated with Grade 5 (see table A2 and A3). Their students have one year more schooling. As the difference between two adjacent grades is about 40 points (see supra), we can estimate the effect of one year more schooling on about 40 – 24.57 (the age effect) = 15.43 points. But because each country participating with Grade 5, has its own reasons to do so, we did not correct for different numbers of years schooling in this study.

RQ 5: The effects of changes in instructional and teacher factors after correction for age differences

When we re-analyzed the data of the same countries as in RQ 1 and RQ 2 with the average achievement after correcting for age differences as reported in Tables A2 and A3, the main results on the correlations between the changes over time were more or less the same as in the first part (see Tables A4, A5 and A6). This means: no significant correlation between the changes in instructional time (r = .13 for % instructional time for language in Table A4 and r = .05 for % instruction time for reading in Table A5) and a significant positive correlation between the changes in the amount of professional development (r = .42 in Table A6) respectively, on the one hand, with the changes in reading achievement on the other hand.

Discussion and conclusion

We start with some considerations on separate research questions and finish with some more general comments.

As for our instructional time variables we were not successful in showing they have an effect on changes in reading comprehension at the country level. Some possible explanations for this result can be given. Firstly, the operationalization of our time variables was rather specific and complex: it was the percentages of total instruction time. But this total instruction time is different across countries. Therefore, future research has to consider some indicators of 'absolute' instruction time differences. Secondly, the exact questions to ask information from the teachers were somewhat different between 2006 and 2016: in hours and minutes in 2006 and in minutes in 2016. Perhaps this difference was partly responsible for different results. Thirdly, we think it is appropriate to consider the instruction time together with information on what is being done during this time.

[1] When two countries were at the same rank, we have worked with half ranks in the table, while we name two ranks in the text.

In particular, a combination of time and indicators of instructional quality seems to be a good variable for future research.

As for time spent on professional development in reading (instruction), we were able to demonstrate its 'causal' effect on changes in reading comprehension. This is in line with the research literature and with some recent results on TIMSS (Trends in Mathematics and Science Study) data (see Gustafsson et al., 2017 and Nilsen et al., 2017). This gives clear indications for schools and policy makers. Nevertheless, the operationalization of this variable in IEA-studies is rather weak. Our suggestion to the IEA is to collect in future waves better and more data on professional development of teachers: considering a period of 4 to 5 years, with attention for different forms of professional development, etc.

Our study does not really support the importance of differences in age in international grade-based studies. Probably this unexpected result is understandable because the age differences between measurement moments within countries were rather small and also the age differences between countries in 2016 were rather small. One important limitation is that we do not really know how big the age effects are in different countries, for different subject domains and at different age levels. Is the Swedish result based on a comparison of grade 3 and 4 and showing big schooling effects and small age effects generalizable over countries? And/or is the result from Iceland and Norway based on a comparison between grade 4 and 5 and showing small schooling and big age effects generalizable? It would be great to include in a future PIRLS study (as has been done in a TIMSS study in the past) two adjacent grades. This would allow the IEA to publish also a valid 'ranking of countries corrected for age differences.'

Overall, our impression on the used method is positive. Nevertheless, we have observed that in most analyses the data of some individual countries go against the general pattern. The scatterplots showed these outliers.

For one variable at least – professional development – we could observe a 'causal' effect on achievement.

After correction for age differences between countries at a certain moment, we have found the same results for the three considered instructional or teacher variables. This is an indication of the robustness of the procedure. It can help to convince skeptical readers that a difference-in difference approach considering changes in only one independent variable in a group of countries allows to reach valid conclusions.

Perhaps using this approach with only two measurement moments is not optimal. In future research we have to include all available waves of a study as PIRLS, TIMSS or PISA (Programme for the International Student Assessment). (For an example of a study including three waves of PIRLS, see Liu et al., 2014.)

Last but not least, our study is a rather early study using the PIRLS 2016 database, which has been made publicly available only recently. We hope the broader research community will continue its efforts to explain these new data and to be able to use these data to do suggestions for improvements to our schools and educational policy makers.

Acknowledgements

We thank the Flemish ministry of Education for the support for participating in PIRLS 2016 and for analysing its database.

Statements

1. The authors declare that they are unaware of any potential conflict of interest.

2. All data were collected in line with the Belgian law and the European regulations on the privacy of the subjects and schools involved. The data are – in an anonymous form – processed by the IEA and made publicly available at the IEA-websites.

References

Creemers, B., & Kyriakides, L. (2008). *The dynamics of educational effectiveness. A contribution to policy, practice and theory in contemporary schools.* New York, United States: Routledge.

Gustafsson, J.-E. (2007). Understanding causal influences on educational achievement through analysis of differences over time within countries. In T. Loveless (Ed.), *Lessons learned: What international assessments tell us about math achievement*, (pp. 37–63). Washington, DC: Brookings Institution Press.

Gustafsson, J.-E. (2013). Causal inference in educational effectiveness research: a comparison of three methods to investigate effects of homework on student achievement. *School Effectiveness and School Improvement*, 24(3), 275-295. doi:10.1080/09243453.2013.806334

Gustafsson, J.-E., & Nilsen, T. (2017). *Changes in mathematics performance from 2011 to 2015 in Norway and Sweden as a function of teachers' professional development.* Paper presented at the European Conference on Educational Research, Copenhagen, Denmark.

Liu, H., Bellens, K., Gielen, S., Van Damme, J., & Onghena, P. (2014). A country level longitudinal study on the effect of student age, class size and socio-economic status - based on PIRLS 2001, 2006 & 2011. In R. Strietholt, W. Bos, J. E. Gustafsson, & M. Rósen (Reds.), *Educational Policy Evaluation through International Comparative Assessments* (pp. 225-242). Münster, Germany: Waxmann.

Luyten, H. (2006). An empirical assessment of the absolute effect of schooling: Regression-discontinuity applied to TIMSS-95. *Oxford Review of Education*, 32, 397-429. doi:10.1080/03054980600776589

Martin, M. O., Mullis, I. V. S., & Hooper, M. (Eds.). (2017). *Methods and Procedures in PIRLS 2016*. Retrieved from https://timssandpirls.bc.edu/publications/pirls/2016-methods.html

Martin, M. O., Mullis, I. V. S., & Kennedy, A. M. (2007). *PIRLS 2006 Technical Report.* Retrieved from https://timssandpirls.bc.edu/pirls2006/tech_rpt.html

Mullis, I. V. S., & Martin, M. O. (Eds.). (2015). *PIRLS 2016 Assessment Framework* (2nd ed.). Retrieved from http://timssandpirls.bc.edu/pirls2016/framework.html

Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, M. (Eds.) (2017). *PIRLS 2016 International Results in Reading.* Retrieved from http://timssandpirls.bc.edu/pirls2016/international-results/pirls/summary/

Mullis, I. V. S., Martin, M. O., Kennedy, A. M., & Foy, P. (2007). *PIRLS 2006 international report: IEA's Progress in International Reading Literacy Study in primary schools in 40 countries.* Chesnut Hill, MA: Boston College.

Nilsen, T., Gustafsson, J.-E., & Yang Hansen, K. (2017). *Changes in teacher quality and school climate influencing changes in achievement from 2011 to 2015.* Paper presented at the European Conference on Educational Research, Copenhagen, Denmark.

Rindermann, H. (2007). The g-factor of international cognitive ability comparisons: the homogeneity of results in PISA, TIMSS, PIRLS and IQ-tests across nations. *European Journal of Personality*, 21(5), 667-706. doi:10.1002/per.634

Scheerens, J. (2016). *Educational effectiveness and ineffectiveness: A critical review of the knowledge base.* Dordrecht, the Netherlands: Springer.

Van Damme, J., Liu, H., Vanhee, L., & Pustjens, H. (2010). Longitudinal studies at the country level as a new approach to educational effectiveness: explaining change in reading achievement (PIRLS) by change in age, socio-economic status and class size. *Effective Education, 2*(1), 53-84. doi:10.1080/19415531003616888

Appendix

Table A1

The percentage of the total instructional time spent on (a) language instruction, and (b) reading instruction, and (c) the time spent in the past two years on formal professional development activities that dealt directly with reading or teaching reading in PIRLS 2006 and 2016

Country	Language Time_06	Language Time_16	Reading Time_06	Reading Time_16	Develop_06	Develop_16
AUS	38	34	14	13	2.96	3.04
BFL	31	26	15	9	1.74	1.97
BFR	39	34	20	27	2.02	2.06
BGR	33	33	29	39	2.14	3.05
COT	34	31	23	24	3.53	2.99
CQU	37	34	19	16	2.30	2.65
TWN	22	26	9	15	2.60	3.22
DNK	24	23	21	13	2.65	2.40
ENG	28	28	13	12	2.17	2.57
FRA	38	37	23	19	1.86	1.95
GEO	36	26	24	19	3.06	3.85
DEU	32	29	13	11	2.30	2.10
HKG	22	24	11	13	3.03	2.77
HUN	36	35	32	25	2.92	2.41
IRN	28	23	22	20	4.32	3.47
ITA	26	28	16	14	2.52	2.62
LVA	29	30	20	28	3.12	3.41
LTU	30	29	28	23	2.56	3.04
NLD	32	34	15	19	2.22	2.87
NZL	37	37	23	24	2.71	3.30
NOR4	36	29	30	22	2.69	2.89
NOR5	35	23	28	17	2.43	3.00
RUS	39	41	29	27	3.21	3.71
SGP	27	27	16	12	3.06	3.40
SVK	33	30	28	23	2.02	2.40
SVN	27	26	16	12	2.81	2.54
ESP	23	25	18	16	2.58	3.30
SWE	27	22	17	12	2.56	3.04
TTO	35	36	29	30	2.40	3.41
USA	31	30	29	32	3.61	3.60

Table A2

Countries in PIRLS 2006 ranked according to reading achievement versus ranked according to reading achievement corrected for differences in age (24.57 points per year) (in order of the new ranking)

Country	Age	Reading score	Original ranking 2006	Corrected reading score	Corrected ranking	Ranking difference
Hong Kong	10.0	564	2.0	569	1	1.0
Canada, British Columbia	9.8	558	4.5	568	2	2.5
Canada, Alberta	9.9	560	3.0	567	3	0.0
Canada, Ontario	9.8	555	7.0	565	4	3.0
Italy	9.7	551	8.5	563	5	3.5
Singapore	10.4	558	4.5	553	6	-1.5
Belgium (Flemish)	10.0	547	13.0	552	7	6.0
Russian Federation	10.8	565	1.0	550	8	-7.0
Canada, Nova Scotia	10.0	542	16.0	547	9	7.0
Netherlands	10.3	547	13.0	545	10	3.0
United States	10.1	540	19.0	542	11	8.0
Germany	10.5	548	11.0	541	12	-1.0
Hungary	10.7	551	8.5	539	13	-4.5
Chinese Taipei	10.1	535	23.0	537	14	9.0
New Zealand	10.0	532	25.0	537	15	10.0
England	10.3	539	20.0	537	16	4.0
Austria	10.3	538	21.0	536	17	4.0
Canada, Quebec	10.1	533	24.0	535	18	6.0
Scotland	9.9	527	27.0	534	19	8.0
Sweden	10.9	549	10.0	532	20	-10.0
Bulgaria	10.9	547	13.0	530	21	-8.0
Slovenia	9.9	522	28.5	529	22	6.5
Denmark	10.9	546	15.0	529	23	-8.0
Luxembourg (5)	11.4	557	6.0	528	24	-18.0
France	10.0	522	28.5	527	25	3.5
Poland	9.9	519	30.0	526	26	4.0
Norway (5)	10.8	541	17.5	526	27	-9.5
Slovak Republic	10.4	531	26.0	526	28	-2.0
Lithuania	10.7	537	22.0	525	29	-7.0
Latvia	11.0	541	17.5	521	30	-12.5
Iceland	9.8	511	33.0	521	31	2.0
Spain	9.9	513	31.0	520	32	-1.0
Israel	10.1	512	32.0	514	33	-1.0
Norway (4)	9.8	498	36.0	508	34	2.0
Belgium (French)	9.9	500	34.5	507	35	-0.5
Moldova	10.9	500	34.5	483	36	-1.5
Georgia	10.1	471	38.0	473	37	1.0
Romania	10.9	489	37.0	472	38	-1.0
Trinidad and Tobago	10.1	436	40.0	438	39	1.0
Macedonia	10.6	442	39.0	432	40	-1.0
Iran	10.2	421	41.0	421	41	0.0
Indonesia	10.4	405	42.0	400	42	0.0
Qatar	9.8	353	43.0	363	43	0.0
Kuwait	9.8	330	44.0	340	44	0.0
Morocco	10.8	323	45.0	308	45	0.0
South Africa (5)	11.9	302	46.0	260	46	0.0

Table A3

Countries in PIRLS 2016 ranked according to reading achievement versus ranked according to reading achievement corrected for differences in age (24.57 points per year) (in order of the new ranking)

Country	Age	Reading score	Original ranking 2016	Corrected reading score	Corrected ranking	Ranking difference
Russia, Moscow	10.8	612	1.0	597	1	0.0
Hong Kong	9.9	569	4.0	576	2	2.0
Singapore	10.4	576	3.0	571	3	0.0
Russian Federation	10.8	581	2.0	566	4	-2.0
Chinese Taipei	10.1	559	10.0	561	5	5.0
Italy	9.7	548	18.5	560	6	12.5
Northern Ireland	10.4	565	7.5	560	7	0.5
Ireland	10.5	567	5.0	560	8	-3.0
England	10.3	559	10.0	557	9	1.0
Spain, Madrid	9.9	549	16.5	556	10	6.5
Canada, Ontario	9.8	544	24.5	554	11	13.5
Poland	10.7	565	7.5	553	12	-4.5
United States	10.1	549	16.5	551	13	3.5
Finland	10.8	566	6.0	551	14	-8.0
Macao	10.0	546	22.0	551	15	7.0
Canada	9.9	543	26.5	550	16	10.5
Canada, Quebec	10.1	547	20.5	549	17	3.5
Slovenia	9.9	542	28.0	549	18	10.0
Australia	10.0	544	24.5	549	19	5.5
Netherlands	10.1	545	23.0	547	20	3.0
Norway (5)	10.8	559	10.0	544	21	-11.0
Hungary	10.6	554	14.0	544	22	-8.0
Sweden	10.7	555	13.0	543	23	-10.0
Latvia	10.9	558	12.0	541	24	-12.0
Czech Republic	10.3	543	26.5	541	25	1.5
Austria	10.3	541	29.0	539	26	3.0
Portugal	9.8	528	34.5	538	27	7.5
Bulgaria	10.8	552	15.0	537	28	-13.0
Spain	9.9	528	34.5	535	29	5.5
Israel	10.0	530	33.0	535	30	3.0
Spain, Andalusia	9.8	525	36.5	535	31	5.5
Germany	10.3	537	30.0	535	32	-2.0
Kazakhstan	10.3	536	31.0	534	33	-2.0
Lithuania	10.8	548	18.5	533	34	-15.5
Denmark (4)	10.8	547	20.5	532	35	-14.5
Slovak Republic	10.4	535	32.0	530	36	-4.0
Belgium (Flemish)	10.1	525	36.5	527	37	-0.5
Norway (4)	9.8	517	39.0	527	38	1.0
New Zealand	10.1	523	38.0	525	39	-1.0
United Arab Emirates, Dubai	9.9	515	40.0	522	40	0.0

Country	Age	Reading score	Original ranking 2016	Corrected reading score	Corrected ranking	Ranking difference
France	9.8	511	41.0	521	41	0.0
Denmark (3)	9.8	501	42.0	511	42	0.0
Belgium (French)	10.0	497	43.0	502	43	0.0
Georgia	9.7	488	45.0	500	44	1.0
Chile	10.1	494	44.0	496	45	-1.0
Argentina, Buenos Aires	10.0	480	46.0	485	46	0.0
Trinidad and Tobago	10.2	479	47.0	479	47	0.0
Azerbaijan	10.1	472	48.0	474	48	0.0
Malta	9.7	452	49.0	464	49	0.0
United Arab Emirates	9.8	450	50.0	460	50	0.0
Bahrein	9.9	446	51.0	453	51	0.0
Qatar	10.0	442	52.0	447	52	0.0
Saudi Arabia	9.9	430	53.0	437	53	0.0
Oman	9.7	418	55.0	430	54	1.0
Iran	10.2	428	54.0	428	55	-1.0
United Arab Emirates, Abu Dhabi	9.7	414	56.0	426	56	0.0
Kuwait	9.6	393	58.0	408	57	1.0
RSA, Eng/Afr/Zulu (5)	11.6	406	57.0	372	58	-1.0
Morocco	10.2	358	59.0	358	59	0.0
Egypte	10.0	330	60.0	335	60	0.0
South Africa	10.6	320	61.0	310	61	0.0

Table A4

Correlation between (changes in) the percentage of the total instructional time spent on language instruction and (changes in) reading achievement in PIRLS 2006 and 2016 after correction for differences in age (N=29)

	ReadCorr_06	ReadCorr_16	ReadCorr_Diff	LanguageTime_06	LanguageTime_16
ReadCorr_16	.91***	1.00			
ReadCorr_Diff	-.44*	-.03	1.00		
LanguageTime_06	-.15	-.24	-.18	1.00	
LanguageTime_16	.07	.03	-.10	.78***	1.00
LanguageTime_Diff	.33	.43*	.13	-.41*	.25

* p < .05; *** p < .001 (2-tailed)

Table A5

Correlation between (changes in) the percentage of the total instructional time spent on reading instruction and (changes in) reading achievement in PIRLS 2006 and 2016 after correction for differences in age (N=29)

	ReadCorr_06	ReadCorr_16	ReadCorr_Diff	ReadingTime_06	ReadingTime_16
ReadCorr_Diff	-.44*	-.03	1.00		
ReadingTime_06	-.32	-.30	.13	1.00	
ReadingTime_16	-.26	-.23	.14	.76***	1.00
ReadingTime_Diff	.02	.05	**.05**	-.16	.52**

* p < .05; ** p < .01; *** p < .001 (2-tailed)

Table A.6

Correlation between (changes in) teachers' professional development and (changes in) reading achievement in PIRLS 2006 and 2016 after correction for differences in age (N=29)

	ReadCorr_06	ReadCorr_16	ReadCorr_Diff	Develop_06	Develop_16
ReadCorr_Diff	-.44*	-.03	1.00		
Develop_06	-.23	-.16	.21	1.00	
Develop_16	-.31	-.08	.58**	.66***	1.00
Develop_Diff	-.07	.11	**.42***	-.48**	.35

* p < .05; ** p < .01; *** p < .001 (2-tailed)