Frade-Martínez, C., Gamazo, A., and Olmos-Migueláñez, S. (2024). Evolution of contextual predictors of Spanish students' proficiency level: a comparative study between PISA 2015 and 2018. *Revista de Investigación Educativa, 42*(2). DOI: https://doi.org/10.6018/rie.548151

Evolution of contextual predictors of Spanish students' proficiency levels: a comparative study between PISA 2015 and 2018

Evolución de los predictores contextuales del nivel competencial de las y los estudiantes españoles: un estudio comparativo entre PISA 2015 y 2018

Cristina Frade-Martínez*[1] , Adriana Gamazo* and Susana Olmos-Migueláñez*.

*University Institute of Education Sciences.
University of Salamanca (Spain)

Abstract

The Programme for International Student Assessment (PISA) has been assessing the competency level of 15-year-old students for more than 20 years, influencing in turn the establishment of educational policies and practices based on its results at the international level. Although its configuration does not allow for the establishment of longitudinal studies, this article proposes the design of a study of trends that makes it possible to assess the evolution of those factors of a socio-demographic nature or educational context that best predict the competency level of students. Through a multilevel regression analysis (hierarchical linear models) with the Spanish sample of the 2015 and 2018 editions of PISA, comprising 65,684 students and 1,873 schools, we observe changes in the contextual predictors of performance in reading comprehension, science and mathematics. The most notable findings are the reduction of the impact of the migratory status of first-generation immigrants, the reduction of the gender gap in STEM subjects (and its increase in reading comprehension) or the reduction of the contextual effect of the average socio-economic level of a school's student body. It concludes with the need for more in-depth analyses, both at the statistical and policy levels, in order to produce more detailed results to clarify which measures are useful for reducing the impact of socio-economic, demographic and educational contextual factors on the performance of Spanish students.

Keywords: PISA; performance; educational assessment; secondary education.

[1] **Correspondence:** Cristina Frade-Martínez, cristina.frade@usal.es, Paseo Canalejas 169, 37008, Salamanca, Spain.

Resumen

El Programa para la Evaluación Internacional de Alumnos (PISA) lleva más de 20 años evaluando el nivel competencial del alumnado de 15 años, influyendo a su vez en el establecimiento de políticas y prácticas educativas basadas en sus resultados a nivel internacional. Aunque su configuración no permite el establecimiento de estudios longitudinales, este artículo plantea el diseño de un estudio de tendencias que posibilita la valoración de la evolución de aquellos factores de carácter sociodemográfico o de contexto educativo que mejor predicen el nivel competencial de los y las estudiantes. A través de un análisis de regresión multinivel (modelos jerárquicos lineales) con la muestra española de las ediciones 2015 y 2018 de PISA, conformada por 65684 estudiantes y 1873 centros educativos, se observan los cambios en las variables predictoras de carácter contextual del rendimiento en comprensión lectora, ciencias y matemáticas. Los hallazgos más reseñables son la reducción del impacto del estatus migratorio de los y las inmigrantes de primera generación, la disminución de la brecha de género en las materias STEM (y su aumento en la comprensión lectora) o la reducción del efecto contextual del nivel socioeconómico medio del estudiantado de un centro. Se concluye con la necesidad de realizar análisis más profundos, tanto a nivel estadístico como de política educativa, para poder producir resultados más detallados que permitan esclarecer qué medidas son útiles para la reducción del impacto de los factores socioeconómicos, demográficos y de contexto educativo en el rendimiento de las y los estudiantes españoles.

Palabras clave: PISA; rendimiento; evaluación educativa; educación secundaria.

Introduction and objectives

The Programme for International Student Assessment (PISA), developed by the Organisation for Economic Co-operation and Development (OECD), is one of the large-scale international assessment tests that has had the greatest impact on societies and education systems since its introduction in 2000.

This programme was conceived as a "resource to provide rich and detailed information to enable member countries to make the necessary public decisions and policies to improve educational standards" (OECD, 2006, p. 3). (OECD, 2006, p. 3).. Certainly, since its first application, PISA results have generated diverse reactions from governments of participating countries depending on their overall performance. Some authors refer to the tendency of governments to change their education policies in reaction to PISA results as 'PISA shock', usually based on the initial descriptive results rather than on the deeper secondary analyses of the data (Wiseman, 2013). However, these results can lead to misleading inferences about the quality of education systems (Jornet, 2016).

On the other hand, although data collection is cyclical, the cross-sectional nature of the test makes it impossible to establish causal relationships and to track the effects of various factors on the evolution of subjects' performance over time (Rutkowski et al., 2017). However, this temporal perspective is vital for an analysis of the effectiveness of the various policies of an education system over time.

Because of these two issues, this article proposes a secondary analysis of the PISA data using a trend study approach with a twofold objective. On the one hand, to offer a deeper multivariate and multilevel analysis than the PISA results that usually transcend the general population and policy makers, and on the other hand, to obtain a temporal perspective that allows us to analyse the evolution of the effect of contextual factors on the performance of Spanish students across two different editions of PISA.

Factors associated with performance

Although the main purpose of large-scale assessments is to study the performance of participating students in specific competences or skills, the most commonly performed secondary analysis is the study of factors of various kinds that are related to this performance.

The 1990s saw a boom in research on school effectiveness. This led to the development of different models to develop this research, among which the dynamic model proposed by Creemers and Kyriakides stands out. . According to this model, there are different levels of factors that explain students' educational performance, ranging from the national level to the student level, the school, the teacher (or class) and the contextual characteristics of the students themselves (Figure 1). Following the above model, the most widely used to date, our research focuses on factors at the student and context level and at the school level that are shown to be interrelated and support us in defining the nature of school efficacy (Kyriakides et al., 2009)..

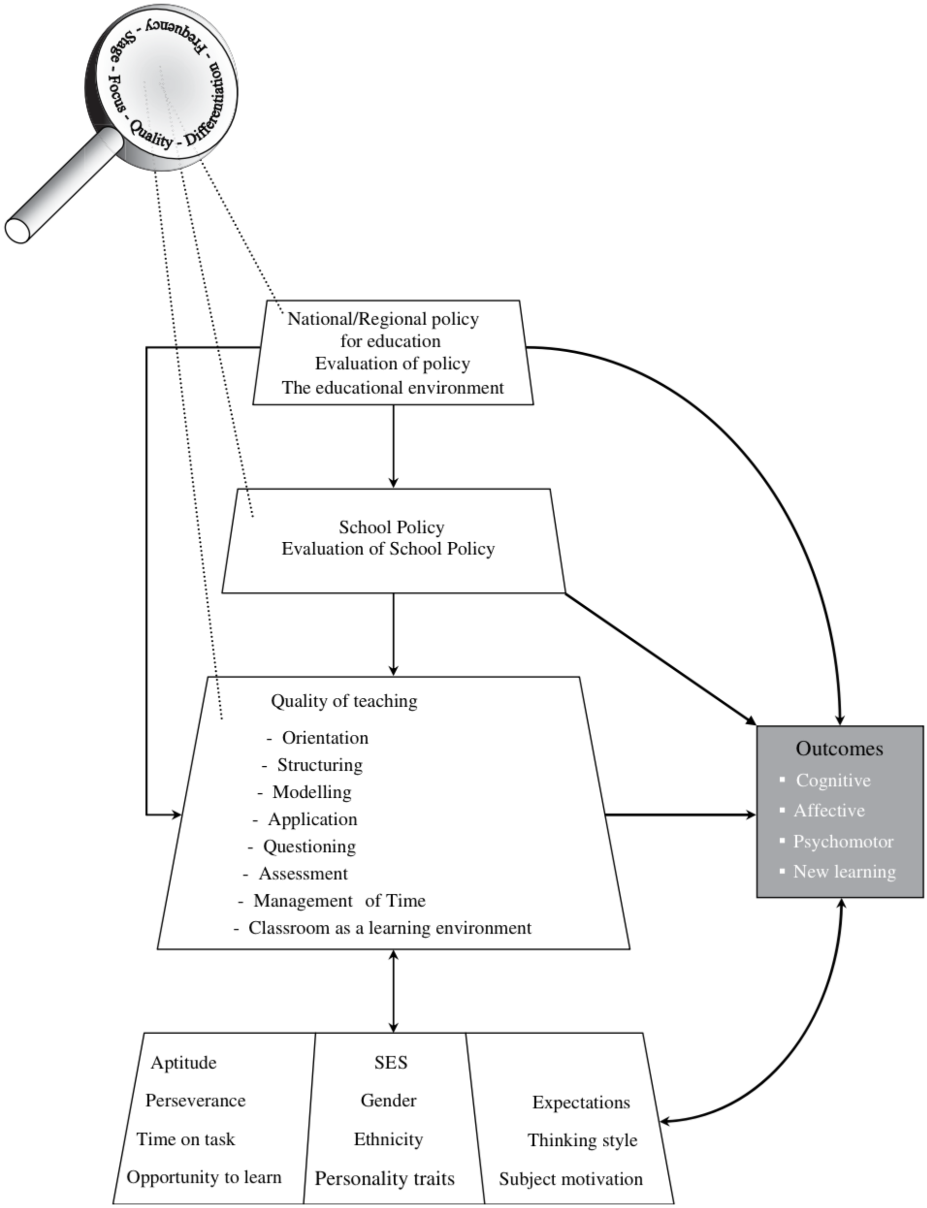

Figure 1. A dynamic model of school effectiveness (Kyriakides et al., 2009, p. 13).

The set of factors that are significantly related to student achievement varies widely depending on the proficiency being studied or the research methodology employed, as well as on the country or cultural background from which students come. However, there are certain variables, mainly of a socio-demographic and educational background nature, which consistently demonstrate a significant influence on the achievement of school-age students.

Given the nature of large-scale assessments, which collect data at two levels (student and school), we can find two main types of variables studied based on the previous literature review.

Student-level variables tend to show the most influence on student performance, the most frequent being socio-economic status (Lenkeit, 2012; Pomianowicz, 2021; Yetişir and Bati, 2021)gender (Laukaityte and Rolfsman, 2020; Martínez-Abad et al., 2020; H. Wu et al., 2020), migration status (migration status (Doncel Abad and Cabrera Álvarez, 2020; Gómez-Fernández and Mediavilla, 2021; Pomianowicz, 2021), grade repetition (having repeated a school year (Doncel Abad and Cabrera Álvarez, 2020; Autora et al., 2018; Martínez-Abad, 2019) or speaking a language other than the vehicular language of the educational system at home (Doncel Abad and Cabrera Álvarez, 2020; Martínez-Abad et al., 2020; Pomianowicz, 2021)..

With respect to school factors, the only variable that consistently shows an influence on students' achievement is the contextual effect of the socio-economic status of the school's student body (Ding and Homer, 2020; Kameshwara et al. (Ding and Homer, 2020; Kameshwara et al., 2020; H. Wu et al., 2020).although some studies also find other types of variables such as school size or school ownership to be relevant (Hu et al., 2018; Martínez-Abad, 2019)..

Longitudinality and trend studies

One of the most frequent criticisms of PISA's configuration is its cross-sectional nature, which does not allow for an analysis of the evolution of students and their circumstances or for establishing causal relationships between the study variables. (Carabaña, 2015; Author, 2020; Han, 2018).or, alternatively, data on the students' previous performance (Dumay and Dupriez, 2014; Willms, 2010)This makes it impossible to establish value-added models that provide more accurate information on school effects.

The inherent characteristics of this large-scale assessment test do not allow us to carry out longitudinal studies, characterised by referring to different assessments on the same subjects over time. (de Miguel, 1985). Thus, taking into account that the aim of this study is to draw conclusions about the evolution of the results and characteristics of a specific population over time, there are two other methodological options: a cohort study and a trend study. In the case of the cohort study, although the subjects from whom information is collected are not necessarily the same in each measurement, the population from which these subjects are drawn is stable (e.g. subjects born in a specific year or people from the same university graduating class) (Bisquerra, 2004; Cohen et al., 2017). However, given that the target population of PISA is students who are 15 years old at the time of the assessment, this methodology is not feasible either.

As an alternative to the impossibility of carrying out a longitudinal or cohort study, this research proposes a trend study. Trend studies are a variant of the cohort study, in which the aim is to describe a change by measuring populations that are not stable over time, and therefore the subjects that make up the population under study are different in each data collection (Bisquerra, 2004; Cohen, 2017; Sullivan and Calderwood, 2017).

Although varying the sample size between editions of PISA introduces elements that may distort the results, trend studies are a reasonably sound strategy capable of identifying trends in behaviour by analysing information collected at different points in time (Bisquerra, 2004). (Bisquerra, 2004).

Taking all this into account, the general objective of this study is to compare the impact of the main socio-demographic variables on the competency performance of Spanish students between the PISA 2015 and 2018 assessments in order to assess their evolution.

Method

This study presents a secondary data analysis of two different PISA cycles based on an ex post facto design, as it is conducted on previously collected data and does not involve any manipulation of variables. Although each of the data collections is cross-sectional in nature, by pooling the analysis of both editions, a trend study is proposed, as justified in the previous section.

Population and Sample

The study population consists of Spanish students who were 15 years old in 2015 and 2018. The selection of these PISA cycles is motivated by the variables available in the background questionnaires, which serve as predictor variables in the present study. The 2015 and 2018 editions contain the same variables; however, the pre-2015 editions are missing some variables that have been shown to be relevant in predicting performance, such as the number of school changes a student has made during his/her school history or the length of schooling in the period before compulsory education (pre-primary education). (Author et al., 2018). For this reason, and to ensure the highest possible degree of comparability, it was decided to select only data from the 2015 and 2018 assessments.

The sample is obtained through a two-stage sampling process, which initially selects schools to participate (representative in terms of region/country and school ownership) and then determines which students should take the tests within each school. (OECD, 2019).

In the case of Spain, the sample collected is representative at Autonomous Community level, which allows a disaggregation of the data by region that is not available in all participating countries and facilitates inter-regional comparison.

Since one of the aims of the study is to analyse the effect of school-level factors, schools with fewer than 20 students surveyed have been excluded, following the recommendation of some authors (Joaristi et al., 2014; Martínez-Abad, 2019; Martínez-Abad et al. (Joaristi et al., 2014; Martínez-Abad, 2019; Martínez-Abad et al., 2017; Meunier, 2011).. Thus, the study sample consists of a total of 65684 students and 1873 schools, distributed by year as shown in Table 1.

Table 1

Study sample, by PISA cycle and level

	2015	2018
Alumni	31273	34411
Educational establishments	897	976

Instruments

The present study uses information collected through two types of instruments, both of which are part of the PISA assessment[2] . First, use is made of the results of the competency assessment questionnaires (mathematics, science and reading). These questionnaires are made up of different types of items, depending on the specific content to which they refer, the level of difficulty or the type of response required (multiple, closed, open-ended).

The context questionnaires, applied to students and schools, are aimed at gathering information on non-cognitive outcomes (self-efficacy, motivation, attitudes towards school and learning), individual conditions (educational and socio-economic environment), and the procedural and organisational characteristics of the institutional environment (structure, resources and processes of the school). These questionnaires are self-reported, and are answered by the students and the management team of the participating schools.

Variables

The criterion variables used in this study are the scores of the participating students in each of the main competences assessed: mathematical competence, scientific competence and reading comprehension competence[3] .

On the other hand, the predictor variables included in the study are some of the contextual variables at student and school level reflected in the PISA context questionnaires (Table 2) and based on the literature review above.

Table 2

Contextual variables of multilevel models

Variable	Label		Range
	PISA 2015	PISA 2018	

[2] The PISA databases are publicly available and can be found at the following link: https://www.oecd.org/pisa/data/

[3] For an in-depth definition of these variables, see the Spanish PISA Report 2018 (Ministry of Education and Vocational Training, 2019).

Level 1 - Students	Gender	ST004D01T	ST004D01T	0: Male 1: Female
	Month of birth (2015)	ST003D02T		1-12
	Age (2018)		AGE	Continua
	Course	GRADE	GRADE	1st ESO - 1st Bachillerato
	Socio-economic and cultural index (ESCS)	ESCS	ESCS	Continua
	Migration status	IMMIG	IMMIG	0: Native, 1: 2nd generation immigrant 2: 1st generation immigrant
	Repeater status	REPEAT	REPEAT	0: No 1: Yes
	Number of school changes (2015)	SCCHANGE		0: No change 1: A change 2: Two or more changes
	Number of school changes (2018)		SCCHANGE	Continua
	Number of years spent in pre-school education	DURECEC	DURECEC	Continua
	Language spoken at home	ST022Q01TA	ST022Q01TA	0: Test language 1: Other language
Level 2 - Centres	Size of the centre	SCHSIZE	SCHSIZE	Continua
	Class sizes	CLSIZE	CLSIZE	Continua
	Shortage of resources	EDUSHORT	EDUSHORT	Continua
	Teacher shortage	STAFFSHORT	STAFFSHORT	Continua
	Ownership of the centre	SCHLTYPE	SCHTYPE	1: Private 2: Concerted 3: Public
	Teacher-student ratio	STRATIO	STRATIO	Continua
	Location of the centre	SC001Q01TA	SC001Q01TA	1: Rural area (less than 3000 inhabitants) 2: Small town (3 000-15 000 inhabitants) 3: Town (15 000-100 000

			inhabitants) 4: City (100 000 to 1 000 000 inhabitants) 5: Large city (More than 1 000 000 inhabitants)
Proportion of teachers with a Master's degree	PROAT5AM	PROAT5AM	Continua
Proportion of teachers with PhD	PROAT6	PROAT6	Continua
ESCS medium	ST_ESCS	ST_ESCS	Continua
Proportion of repeaters	ST_REPEAT	ST_REPEAT	Continua
Proportion of immigrant pupils*.	ST_INM	ST_INM	Continua
Proportion of female students	ST_GEN	ST_GEN	Continua

Note: * Variable aggregated at school level from the student database selected for its relevance demonstrated in previous research.

Categorical variables with more than two response options (migration status, changes of school in 2015, ownership of the school, etc.) were converted into *dummy* variables to improve the clarity of the analysis.

Data analysis

The data analysis technique used is multilevel regression, also known as Linear Hierarchical Modelling. The choice of this type of analysis is primarily motivated by the nature of the data from large-scale assessments, which present a nested structure at two levels (students within schools), thus assuming the existence of variability at both levels. The use of multilevel regression allows the simultaneous consideration of the effect of variables at different levels, thus avoiding the biases that could come from using aggregation or disaggregation techniques to carry out simple regression analyses. (Gaviria Soto and Castro Morera, 2005; Raudenbush and Bryk, 2002; Snijders and Bosker, 2011)..

The model is a mixed effects design, with a random intercept and fixed slopes. The random intercept allows the higher level units in the hierarchy of the model (the schools in this case) to have different means in the dependent variable or criterion (Hayes, 2006). The effects of the covariates in this case remain fixed across schools. A random slopes model, in which the covariates are allowed to have a different slope in each school, would be a more accurate reflection of the educational reality. However, it also introduces a level of complexity to the analysis, and given that it is particularly advisable when the aim of the study is to carry out differential effectiveness analyses of schools (Clarke et al., 2010), it is not considered necessary in the present case. Given that the regression coefficient obtained for the models of both editions is not standardised, the results will be interpreted using the t-statistic, which is standardised in terms of the standard error (De Veaux et al., 2021).

In order to check the fulfilment of the previous assumptions of multilevel regression and to assess the appropriateness of this type of analysis, the Intraclass Correlation Coefficient (ICC) of the null model (without covariates) of the criterion variables is calculated, which indicates the proportion of variance in students' performance that is attributable to the second level of analysis (schools). (Author et al., 2018). The value of this coefficient must be equal to or greater than 10% to consider the application of this methodology appropriate (Lee, 2000). (Lee, 2000)This is the case for all the criterion variables used.

Table 3

Intraclass Correlation Coefficient of the null model, by cycle and competence

	2015			2018		
ICC	Mathematics	Science	Reading	Mathematics	Science	Reading
	12.26%	12.41%	12.04%	13.14%	11.39%	14.21%

The analysis was carried out with the statistical programme specialised in multilevel regression HLM. This programme allows the processing of data from large-scale evaluations while complying with all the statistical safeguards indicated for this type of data. (OECD, 2019).. First, it allows the treatment of criterion variables with plausible values, as is the case for PISA performance data. Plausible values can be defined as the range of skills that a student could reasonably have, given his/her responses to the items (Wu, 2005). (Wu, 2005)They are particularly useful in the analysis of data with complex designs such as PISA because they facilitate addressing problems related to biases in the estimation of population parameters and also facilitate the computation of standard estimation errors in complex sample designs.

On the other hand, the HLM programme also allows the simultaneous use of sampling weights at student and school level. These weights reflect the unequal probability of students and schools to be selected in the sampling process, and their application ensures the correct representativeness of the sample.

Results

The results of the multilevel modelling process for each of the competences are presented below.

Mathematics

After selecting only those variables with a significant relationship with performance in mathematical competence, the resulting models respond to the following equations (1 and

2).

$$Matemáticas\ 2015_{ij} = \gamma_{00} + \gamma_{01} * ESCS\ nivel\ 2_j + \gamma_{10} * Género_{ij} + \gamma_{20} * Inmigrante\ 1^{a}\ generación_{ij} + \gamma_{30} * Curso_{ij} + \gamma_{40} * Repetición_{ij} + \gamma_{50} * ESCS_{ij} + \gamma_{60} * Cambio\ centro\ 1_{ij} + \gamma_{70} * Cambio\ centro\ 2_{ij} + u_{0j} + r_{ij}$$

Equation 1

$$Matemáticas\ 2018_{ij} = \gamma_{00} + \gamma_{01} * Escasez\ personal_j + \gamma_{02} * ESCS\ nivel\ 2_j + \gamma_{10} * ESCS_{ij} + \gamma_{20} * Género_{ij} + \gamma_{30} * Curso_{ij} + \gamma_{40} * Edad_{ij} + \gamma_{50} * Lengua_{ij} + \gamma_{60} * Repetición_{ij} + \gamma_{70} * Cambio\ centro_{ij} + u_{0j} + r_{ij}$$

Equation 2

Where:
$Matemáticas_{ij}$ is the school's average score in mathematics;
γ_{00} is the average of all schools in mathematics;
$\gamma_{01} - \gamma_{02}$ are the level 2 covariates;
$\gamma_{10} - \gamma_{70}$ are the level 1 covariates;
u_{0j} is the difference between the school's mathematics score and the overall average of all schools; and
r_{ij} is the residue of level 1.

The coefficients for each of the covariates, together with their t-ratio value and significance level can be found in table 4.

Table 4

Multilevel regression results for mathematics proficiency, 2015 and 2018

		2018			2015	
Fixed effect	**Coef.**	**t-ratio**	**Sig.**	**Coef.**	**t-ratio**	**Sig.**
INTERCEPT	375.84	8.831	<0.001	543.069	313.438	<0.001
N2 Staff shortages	-2.635	-2.264	0.024			
N2 ESCS medium	14.864	6.860	<0.001	13.000	6.820	<0.001
N1 ESCS	11.624	12.776	<0.001	9.787	8.729	<0.001
N1 Gender	-18.223	-9.225	<0.001	-22.578	-11.474	<0.001
N1 Course	26.800	5.844	<0.001	34.859	9.048	<0.001
N1 Age	10.050	3.709	<0.001			
N1 Language	-6.396	-2.863	0.006			
N1 Repetition	-55.031	-8.158	<0.001	-35.803	-7.243	<0.001

N1 Changes of centre	-9.499	-9.643	<0.001			
N1 Change of centre (1)				-10.649	-5.724	<0.001
N1 Change of centre (2+)				-15.602	-4.724	<0.001
N1 1st gen immigrant				-11.806	-2.433	0.026

Most of the significant variables for the mathematics performance of Spanish students remain the same between the two editions of the assessment. However, there are some changes, both in terms of the variables included and the magnitude of their influence on performance.

Firstly, it is worth noting that the variables related to socio-economic status, both at individual and school level, maintain similar coefficients in both assessments, so it seems that their influence on performance is stable.

In the section of variables whose effect has reduced in magnitude between 2015 and 2018 are gender, whose negative influence for females has decreased slightly, and grade, whose impact for those students who are in lower grades than what would correspond to their age has decreased considerably. The variable relating to the effect of being a first-generation immigrant deserves a separate mention, as it has disappeared as a relevant variable for the model between the two editions.

On the other hand, the variable related to grade repetition has increased its negative effect on repeating students.

Finally, the 2018 edition includes some new variables that were not significant in 2015 (staff shortages, student age and language spoken at home), although none of them has a very significant effect either in terms of coefficient or relative importance for the model (t-ratio).

Science

After the data analysis process and the selection of significant variables, the explanatory models of Spanish pupils' scientific competence are configured according to the following equations (3 and 4).

$$\begin{aligned} Ciencia\ 2015_{ij} = {} & \gamma_{00} + \gamma_{01} * Tamaño\ centro_j + \gamma_{02} * Escasez\ personal_j + \gamma_{03} \\ & * ESCS\ nivel\ 2_j + \gamma_{04} * Porcentaje\ de\ repetidores_j + \gamma_{05} \\ & * Porcentaje\ de\ alumnas_j + \gamma_{10} * Mes\ nacimiento_{ij} + \gamma_{20} * Género_{ij} \\ & + \gamma_{30} * Inmigrante\ 1^{a}\ generación_{ij} + \gamma_{40} * Curso_{ij} + \gamma_{50} * Repetición_{ij} \\ & + \gamma_{60} * ESCS_{ij} + \gamma_{70} * Cambio\ centro\ 1_{ij} + \gamma_{80} * Cambio\ centro\ 2_{ij} \\ & + u_{0j} + r_{ij} \end{aligned}$$

Equation 3

$$\begin{aligned} Ciencia\ 2018_{ij} = {} & \gamma_{00} + \gamma_{01} * ESCS_j + \gamma_{10} * ESCS_{ij} + \gamma_{20} * Género_{ij} + \gamma_{30} * Curso_{ij} + \gamma_{40} \\ & * Edad_{ij} + \gamma_{50} * Lengua_{ij} + \gamma_{60} * Repetición_{ij} + \gamma_{70} * Cambio\ centro_{ij} \\ & + u_{0j} + r_{ij} \end{aligned}$$

Equation 4

Where:
$Ciencia_{ij}$ is the school's average score in science;
γ_{00} is the average of all centres in science;
γ_{01}-γ_{05} are the level 2 covariates;
γ_{10} - γ_{70} are the level 1 covariates;
u_{0j} is the difference between the school's science score and the overall average of all schools; and
r_{ij} is the residue of level 1.

The coefficients and significance levels of these variables can be found in table 5.

Table 5

Multilevel regression results for science proficiency, 2015 and 2018

		2018			2015	
Fixed Effect	**Coef.**	**t-ratio**	**Sig.**	**Coef.**	**t-ratio**	**Sig.**
INTERCEPT	440.24	11.360	<0.001	537.68	73.71	<0.001
N2 ESCS medium	12.378	4.960	<0.001	20.75	7.96	<0.001
N2 Size of centre				-0.01	-2.24	0.025
N2 Staff shortages				2.49	2.14	0.032
N2 Percentage of repeaters				31.95	3.35	<0.001
N2 Percentage female students				31.84	2.66	0.008
N1 ESCS	10.630	10.978	<0.001	9.78	13.99	<0.001
N1 Gender	-13.174	-7.083	<0.001	-19.76	-11.70	<0.001
N1 Course	23.026	6.060	<0.001	38.29	13.25	<0.001
N1 Age/Month of birth	6.040	2.537	0.013	-0.59	-2.56	0.011
N1 Language	-10.410	-3.751	<0.001			
N1 Repetition	-53.281	-10.756	<0.001	-34.29	-8.86	<0.001
N1 Changes of centre	-8.751	-7.377	<0.001			
N1 Change of centre (1)				-10.36	-5.34	<0.001
N1 Change of centre (2+)				-16.96	-6.43	<0.001
N1 1st gen immigrant				-9.63	-3.34	<0.001

As in the case of mathematical competence, some variables remain the same between the two editions, while many others vary.

Again, the variables focusing on socio-economic status at both student and school level remain the same. Although the socio-economic status of students shows a similar impact in 2015 and 2018, it seems that the impact of the contextual effect of the ESCS (at school level) has decreased considerably between the two editions.

Of the remaining variables that are repeated, some have also reduced their impact, such as, for example, gender. This variable shows a reduction in the negative impact for female students, thus reducing the gender gap in scientific competence. The impact of grade has also decreased, bringing students who are in different grades at the time of the assessment closer together in terms of scores.

On the other hand, the increased impact of grade repetition on student performance is again noteworthy.

As can be seen, there are many variables that disappear between the two editions, such as school size, staff shortages, the percentage of repeaters or the percentage of female students at school level, or first-generation immigrant status at student level.

In the analysis of the 2018 edition, some new variables appear, such as the language spoken at home, with a relatively low and negative impact on the performance of those students who do not speak the vehicular language of the education system at home.

Reading

After the modelling process, and once all variables not significantly related to student performance in reading comprehension have been eliminated, the models are configured according to the following equations (5 and 6).

$$\begin{aligned} Lectura\ 2015_{ij} = {} & \gamma_{00} + \gamma_{01} * ESCS\ medio_j + \gamma_{02} * Porcentaje\ de\ repetidores_j + \gamma_{03} \\ & * Porcentaje\ de\ alumnas_j + \gamma_{10} * Género_{ij} + \gamma_{20} * Curso_{ij} + \gamma_{30} \\ & * Repetición_{ij} + \gamma_{40} * ESCS_{ij} + \gamma_{50} * Lengua_{ij} + \gamma_{60} \\ & * Cambio\ centro\ 1_{ij} + \gamma_{70} * Cambio\ centro\ 2_{ij} + u_{0j} + r_{ij} \end{aligned}$$

Equation 5

$$\begin{aligned} Lectura\ 2018_{ij} = {} & \gamma_{00} + \gamma_{01} * ESCS\ medio_j + \gamma_{10} * Género_{ij} + \gamma_{20} * Curso_{ij} + \gamma_{30} \\ & * Repetición_{ij} + \gamma_{40} * ESCS_{ij} + \gamma_{50} * Lengua_{ij} + \gamma_{60} \\ & * Inmigrante\ de\ 1^{a}\ generación_{ij} + \gamma_{70} * Cambio\ centro_{ij} + \gamma_{80} \\ & * Edad_{ij} + u_{0j} + r_{ij} \end{aligned}$$

Equation 6

Where*:*

$Lectura_{ij}$ is the school's average reading score*;*

γ_{00} is the average of all schools in reading*;*

γ_{01}-γ_{03} are the level 2 covariates*;*

γ_{10} -γ_{80} are the level 1 covariates*;*

u_{0j} is the difference between the school's reading score and the overall average of all schools; *and*

r_{ij} is the residue of level 1.

The coefficients and significance levels of these variables can be found in table 6.

Table 6

Multilevel regression results for reading literacy proficiency, 2015 and 2018

	2018			2015		
Fixed Effect	**Coef.**	**t-ratio**	**Sig.**	**Coef.**	**t-ratio**	**Sig.**
INTERCEPT	396.46	11.310	<0.001	522.475	84.912	<0.001
N2 ESCS medium	14.353	6.114	<0.001	20.293	7.633	<0.001
N2 Percentage of repeaters				36.502	3.371	<0.001
N2 Percentage female students				24.719	2.338	0.020
N1 ESCS	10.167	12.998	<0.001	8.299	8.748	<0.001
N1 Gender	15.442	10.362	<0.001	6.776	3.665	<0.001
N1 Course	23.543	7.360	<0.001	41.092	11.757	<0.001
N1 Age	7.542	3.416	<0.001			
N1 Language	-10.475	-4.387	<0.001	-6.718	-2.220	0.033
N1 Repetition	-52.633	-12.657	<0.001	-30.72	-6.09	<0.001
N1 Changes of centre	-8.953	-12.230	<0.001			
N1 Change of centre (1)				-8.857	-4.164	<0.001
N1 Changes centre (2+)				-20.642	-6.403	<0.001

With respect to the socio-economic variables, the result of the two previous models is repeated: while the effect of the variable at the individual level is maintained (and in this case even increases), its contextual effect at the school level decreases.

Another variable whose effect decreases between the two editions is the course.

On the other hand, the phenomenon related to the repetition variable is repeated, the effect of which increases to almost double as a function of t-ratio. In the case of reading comprehension, in contrast to the previous two, the effect of gender increases considerably, thus widening the gender gap, which in this skill works in favour of female students.

In the case of reading comprehension competence, only two variables disappear between the 2015 and 2018 editions: percentage of repeaters and percentage of female students. On the other hand, the age of the student appears as a new variable.

Descriptive study of the gender gap

Given the results obtained in the three models, which show changes in the gender gap between male and female students (a reduction in the case of mathematical and scientific competences and an increase in the case of reading comprehension), a descriptive analysis should be carried out to check whether the gap has narrowed (or widened, in the case of reading) because female students have improved or because male students have worsened. In order to answer this question, an analysis of means according to the gender of the students in both competences is proposed in order to check the evolution of the competence levels. The results of this analysis can be seen in table 7.

Table 7

Mean scores for the three competencies according to gender

Competition	Alumnae			Students		
	2018	2015	Diff. 2018-2015	2018	2015	Diff. 2018-2015
Mate	487,96	487,47	0,48	495,76	499,22	-3,46
Science	489,63	495,84	-6,22	493,12	503,57	-10,45
Reading	495,76	511,06	-15,29	471,25	493,06	-21,81

As can be seen, students' overall performance has worsened between 2015 and 2018 in all skills. The key, therefore, is the magnitude of the decline, which in all cases has been smaller for female students, and especially in the area of reading comprehension.

Intraclass correlation coefficients of the final models

To complete the multilevel analysis, the intraclass correlation index can be explored again, this time including the significant covariates in each model, in order to check the proportion of variability at the centre level that each model has managed to explain (table 8).

Table 8

Intraclass correlation coefficient of the null model and the final model, by cycle and proficiency.

	2015			2018		
	Mate	Science	Reading	Mate	Science	Reading
ICC nil	12.26%	12.41%	12.04%	13.14%	11.39%	14.21%
Final ICC	4.55%	5.6%	5.07%	6.15%	6.86%	10.47%
Proportion explained	62.89%	54.88%	57.89%	52.05%	39.77%	26.32%

These data illustrate the great relevance of socio-demographic and educational context variables in explaining the variability of the criterion variables presented at school level, explaining up to 60% in some models.

Discussion

Although each model has its particularities, some characteristics can be found that are consistently observed in all three.

Firstly, the contextual effect of socio-economic status, i.e. the mean of this variable in each of the schools studied, has decreased considerably, which is a positive aspect of the evolution of the education system, in line with what has been observed in other studies (Lenkeit, 2012; Sirin, 2005)..

Secondly, a striking issue is the increase in the effect of repetition on performance in the three competences studied, as in the study by Martínez-Abad (2019). However, another feature common to all three models may help to explain this issue, at least in part. Just as the effect of repetition has increased consistently and considerably, the course variable has seen its impact reduced to a notable level in all three competences. Given that whether a subject has repeated or not and what year they are in at the time of the assessment are highly correlated variables, it could be hypothesised that, for some reason, part of the effect assigned to the year in 2015 has been transferred to year repetition in 2018, although further studies would be necessary to explore this issue in more depth.

Another issue to analyse is the gender gap in the performance of the participants. In the two science skills (science and mathematics) this gap, which favours male students, has narrowed significantly (between 30% and 50% both in terms of relative relevance to the model), which is in line with the results of other recent studies (Molina Portillo et al., 2022). (Molina Portillo et al., 2022)which also indicate a narrowing of the gender gap in STEM subjects. However, this is not the case for reading comprehension, where the gender gap, which in this case favours female students, has more than doubled its effect between the 2015 and 2018 editions.

Another notable variable is first-generation immigrant status. In 2015, this variable was part of the science and mathematics models. In 2018, this impact has disappeared from these two models, making first-generation immigrant students equal to the rest of their peers in terms of performance in these skills. However, while in the reading comprehension model in 2015 this variable was not among the significant variables, in 2018 it has appeared in this model and with a medium-high relative importance.

Conclusions

Although there are numerous studies dedicated to analysing the impact of socio-demographic and educational variables on student performance through large-scale assessments, the cross-sectional nature of these assessments makes it difficult to assess the evolution of education systems in terms of their ability to control for the effects of the socio-economic context. For this reason, a comparison between the last two editions of PISA is proposed in this study with the aim of gaining some temporal perspective to help assess the effectiveness of the education system over time.

In general, the conclusions regarding the evolution of the education system in its capacity to reduce educational inequalities are positive. Through the comparative study, the reduction of the impact of relevant variables such as the average socio-economic level of students in a school or the condition of being a first generation immigrant has been detected, as well as the reduction of the gender gap in skills related to the STEM field. On this point, taking into account the information contained in Table 7, it can be concluded that the narrowing of the gender gap in STEM subjects and the widening of the gap in reading comprehension is due to a smaller decrease in the scores of female students between the two editions compared to the decrease in the case of male students.

It is worth commenting on some of the limitations of this study, which should be taken into account when assessing the results and conclusions. Firstly, when taking data from the PISA assessment as the basis for the analysis, the limitations of this source must necessarily be assumed, such as the lack of data at the classroom level (Lafontaine et al., 2015; Scheerens et al., 2015). (Lafontaine et al., 2015; Scheerens et al., 2015).which prevents the study of relevant variables such as teaching methodology or the grouping of students, the cross-sectional nature of the data (Author, 2020; Yetişir and Bati, 2021)or existing criticisms of the configuration of the context questionnaires (Autora et al., 2018). (Autora et al., 2018; Li, 2016)which would fundamentally affect the reliability of socio-demographic and educational variables. On the other hand, the study design also has some limitations, such as the small sample of editions studied (only two) due to the lack of comparability with editions prior to 2015.

The results and conclusions drawn indicate some lines of research that it would be interesting to pursue in the future in order to clarify some questions that are beyond the scope of this study. It would be advisable to carry out a more in-depth statistical study of the behaviour and evolution of the course and repetition variables in order to verify the transfer of effects suggested by this study. On the other hand, the reverse behaviour of the gender gap in STEM subjects and in reading comprehension deserves to be studied in more detail, in order to examine possible variables that may explain this behaviour and to plan educational policies and practices aimed at boosting student performance in all skills regardless of gender. Finally, another question that would be interesting to investigate in order to establish educational policies and practices that promote educational equality is the reasons behind the decline in the effect exerted by variables such as the average socio-economic level of students or the status of being a first-generation immigrant.

Funding

This research is part of Project PID2021-125775NB-I00 funded by MCIN/AEI/10.13039/501100011033/ and by ERDF A way to do Europe. Similarly, the research is carried out thanks to a grant for pre-doctoral contracts for the training of doctors with reference PRE2019-087412 funded by the Ministry of Science and Innovation of the Government of Spain.

References

Bisquerra, R. (2004). *Metodología de la investigación educativa*. La Muralla.

Carabaña, J. (2015). *La inutilidad de PISA para las escuelas*. Los Libros de la Catarata, D.L.

Creemers, B. P. M. y Kyriakides, L. (2008). *The Dynamics of Educational Effectiveness: A contribution to policy, practice and theory in contemporary schools*. Routledge. https://doi.org/10.4324/9780203939185

Cohen, L., Manion, L., y Morrison, K. (2017). Surveys, longitudinal, cross-sectional and trend studies. In *Research methods in education* (pp. 334-360). Routledge. https://doi.org/10.4324/9781315456539

de Miguel, M. (1985). Estrategias metodológicas en los estudios longitudinales. *Revista de Investigación Educativa*, *3*(6), 252-270. Pearson.

Ding, H. y Homer, M. (2020). Interpreting mathematics performance in PISA: Taking account of reading performance. *International Journal of Educational Research, 102,* 101566. https://doi.org/10.1016/j.ijer.2020.101566

Doncel Abad, D. y Cabrera Álvarez, P. (2020). Comunidades Autónomas bilingües, identidades y desempeño educativo según PISA 2015. *Revista de educación, 387,* 163-188. https://doi.org/10.4438/1988-592X-RE-2020-387-443

Dumay, X. y Dupriez, V. (2014). Educational quasi-markets, school effectiveness and social inequalities. *Journal of Education Policy*, *29*(4), 510-531. https://doi.org/10.1080/02680939.2013.850536

Gamazo, A. y Martínez-Abad, F. (2020). An Exploration of Factors Linked to Academic Performance in PISA 2018 Through Data Mining Techniques. *Frontiers in Psychology*, *11.* https://doi.org/10.3389/fpsyg.2020.575167

Gamazo, A., Martínez-Abad, F., Olmos-Migueláñez, S., y Rodríguez-Conde, M. J. (2018). Evaluación de factores relacionados con la eficacia escolar en PISA 2015. Un análisis multinivel. *Revista de educación*, *379,* 56-84.

Gaviria Soto, J. L. y Castro Morera, M. (2005). *Modelos jerárquicos lineales*. La Muralla.

Gómez-Fernández, N. y Mediavilla, M. (2021). Exploring the relationship between Information and Communication Technologies (ICT) and academic performance: A multilevel analysis for Spain. *Socio-Economic Planning Sciences*, *77,* 101009. https://doi.org/10.1016/j.seps.2021.101009

Han, S. W. (2018). School-based teacher hiring and achievement inequality: A comparative perspective. *International Journal of Educational Development*, *61,* 82-91. https://doi.org/10.1016/j.ijedudev.2017.12.004

Hayes, A. F. (2006). A Primer on Multilevel Modeling. *Human Communication Research*, *32*(4), 385-410. https://doi.org/10.1111/j.1468-2958.2006.00281.x

Hu, X., Gong, Y., Lai, C., y Leung, F. K. S. (2018). The relationship between ICT and student literacy in mathematics, reading, and science across 44 countries: A multilevel analysis. *Computers & Education*, *125,* 1-13. https://doi.org/10.1016/j.compedu.2018.05.021

Joaristi, L., Lizasoain Hernández, L., y Azpillaga, V. (2014). Detección y caracterización de los centros escolares de alta eficacia de la Comunidad Autónoma del País Vasco mediante Modelos Transversales Contextualizados y Modelos Jerárquicos Lineales. *Estudios sobre educación*, *27,* 37-61.

Jornet Meliá, J. M. (2016). Análisis metodológico del proyecto PISA como evaluación internacional. *RELIEVE. Revista Electrónica De Investigación y Evaluación Educativa, 22*(1), 1-26. https://doi.org/10.7203/relieve22.1.8293

Kameshwara, K. K., Sandoval-Hernandez, A., Shields, R., y Dhanda, K. R. (2020). A false promise? Decentralization in education systems across the globe. *International Journal of Educational Research*, *104,* 101669. https://doi.org/10.1016/j.ijer.2020.101669

Kyriakides, L., Creemers, B. P. M., y Antoniou, P. (2009). Teacher behaviour and student outcomes: Suggestions for research on teacher training and professional development. *Teaching and Teacher Education*, *25*(1), 12-23. https://doi.org/10.1016/j.tate.2008.06.001

Lafontaine, D., Baye, A., Vieluf, S., y Monseur, C. (2015). Equity in opportunity-to-learn and achievement in reading: A secondary analysis of PISA 2009 data. *Studies in Educational Evaluation*, *47,* 1-11. https://doi.org/10.1016/j.stueduc.2015.05.001

Laukaityte, I. y Rolfsman, E. (2020). Low, medium, and high-performing schools in the Nordic countries. Student performance at PISA Mathematics 2003-2012. *Education Inquiry*, *11*(3), 276-295. https://doi.org/10.1080/20004508.2020.1721256

Lee, V. E. (2000). Using Hierarchical Linear Modeling to Study Social Contexts: The Case of School Effects. *Educational Psychologist*, *35*(2), 125-141. https://doi.org/10.1207/S15326985EP3502_6

Lenkeit, J. (2012). How effective are educational systems? A value-added approach to study trends in PIRLS. *Journal of Educational Research Online*, *4*(2),143-173. https://hdl.handle.net/11245/1.405026

Li, H. (2016). How is formative assessment related to students' reading achievement? Findings from PISA 2009. *Assessment in Education: Principles, Policy & Practice*, *23*(4), 473-494. https://doi.org/10.1080/0969594X.2016.1139543

Martínez-Abad, F. (2019). Identification of Factors Associated With School Effectiveness With Data Mining Techniques: Testing a New Approach. *Frontiers in Psychology*, *10*, 2583. https://doi.org/10.3389/fpsyg.2019.02583

Martínez-Abad, F., Gamazo, A., y Rodriguez-Conde, M.-J. (2020). Educational Data Mining: Identification of factors associated with school effectiveness in PISA assessment. *Studies in Educational Evaluation*, *66,* 100875. https://doi.org/10.1016/j.stueduc.2020.100875

Martínez-Abad, F., Lizasoain, L., Castro, M., y Joaristi, L. M. (2017). Selección de escuelas de alta y baja eficacia en Baja California (México). *REDIE: Revista Electrónica de Investigación Educativa*, *19*(2), 38-53. https://doi.org/10.24320/redie.2017.19.2.960

Meunier, M. (2011). Immigration and student achievement: Evidence from Switzerland. *Economics of Education Review*, *30*(1), 16-38.

Molina Portillo, E., Contreras García, J. M., Molina Muñoz, D., y Sánchez Pelegrín, J. A. (2022). Estudio por género del impacto de factores contextuales en el rendimiento matemático del alumnado español en PISA 2018. *Revista Complutense de Educación*, *33*(4), 645-656. https://doi.org/10.5209/rced.76428

OECD. (2006). *El programa PISA de la OCDE: Qué es y para qué sirve*. http://iice.institutos.filo.uba.ar/el-programa-pisa-de-la-ocde-qu%C3%A9-es-y-para-qu%C3%A9-sirve

OECD. (2019). *PISA 2018 Technical Report* (p. 468).

Pomianowicz, K. (2021). Educational achievement disparities between second-generation and non-immigrant students: Do school characteristics account for tracking effects? *European Educational Research Journal*, *22*(3), 297-324. https://doi.org/10.1177/14749041211039929

Raudenbush, S. W. y Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods*. SAGE.

Rutkowski, D., Rutkowski, L., Wild, J., y Burroughs, N. (2018). Poverty and educational achievement in the US: A less-biased estimate using PISA 2012 data. *Journal of Children and Poverty, 24*(1), 47-67. https://doi.org/10.1080/10796126.2017.1401898

Scheerens, J., Luyten, H., van den Berg, S. M., y Glas, C. A. W. (2015). Exploration of direct and indirect associations of system-level policy-amenable variables with reading literacy performance. *Educational Research and Evaluation*, *21*(1), 15-39. https://doi.org/10.1080/13803611.2015.1008520

Sirin, S. (2005). Socioeconomic Status and Academic Achievement: A Meta-Analytic Review of Research. *Review of Educational Research*, 75. https://doi.org/10.3102/00346543075003417

Snijders, T. A. B. y Bosker, R. J. (2011). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. SAGE.

Sullivan, A. y Calderwood, L. (2017). Surveys: longitudinal, cross-sectional, and trend studies. *The BERA/SAGE Handbook of Educational Research*: *Two Volume Set* (Vol.2, pp.395-415). SAGE Publications Ltd. https://doi.org/10.4135/9781473983953

Willms, J. D. (2010). School Composition and Contextual Effects on Student Outcomes. *Teachers College Record*, *112*(4), 1008-1037.

Wiseman, A. W. (2013). Policy responses to PISA in comparative perspective. En H. D. Meyer y A. Benavot (Eds.), *PISA, power, and policy: The emergence of global educational governance* (pp. 303-322). Symposium Books.

Wu, H., Shen, J., Zhang, Y., y Zheng, Y. (2020). Examining the effect of principal leadership on student science achievement. *International Journal of Science Education*, *42*(6), 1017-1039. https://doi.org/10.1080/09500693.2020.1747664

Wu, M. (2005). The role of plausible values in large-scale surveys. *Studies in Educational Evaluation*, *31*(2), 114-128. https://doi.org/10.1016/j.stueduc.2005.05.005

Yetişir, M. y Bati, K. (2021). The Effect of School and Student-Related Factors on PISA 2015 Science Performances in Turkey. *International Journal of Psychology and Educational Studies*, *8,* 170-186. https://doi.org/10.52380/ijpes.2021.8.2.433

Date of receipt: 15 November 2022.
Review date: 13 December 2022.
Date of acceptance: 29 November 2023.