



Educación XX1

ISSN: 1139-613X

ISSN: 2174-5374

educacionxx1@edu.uned.es

Universidad Nacional de Educación a Distancia

España

Arroyo Resino, Delia; Constante-Amores, Alexander; Castro, María; Navarro, Enrique

School effectiveness and high reading achievement of
Spanish students in PISA 2018: a machine learning approach

Educación XX1, vol. 27, núm. 2, 2024, Julio-Diciembre, pp. 223-251

Universidad Nacional de Educación a Distancia

Madrid, España

DOI: <https://doi.org/10.5944/educxx1.38634>

Disponible en: <https://www.redalyc.org/articulo.oa?id=70681678009>

- Cómo citar el artículo
- Número completo
- Más información del artículo
- Página de la revista en [redalyc.org](https://www.redalyc.org)

[redalyc.org](https://www.redalyc.org)

Sistema de Información Científica Redalyc

Red de revistas científicas de Acceso Abierto diamante

Infraestructura abierta no comercial propiedad de la academia

School effectiveness and high reading achievement of Spanish students in PISA 2018: a machine learning approach

Eficacia escolar y alto rendimiento del alumnado español en PISA 2018: un enfoque de machine learning

Delia Arroyo Resino ¹ 

Alexander Constante-Amores ^{2*} 

María Castro ¹ 

Enrique Navarro ¹ 

¹ Universidad Complutense de Madrid, Spain

² Universidad Camilo José Cela, Spain

* Corresponding author. E-mail: israelalexander.constante@ucjc.edu

Cómo referenciar este artículo/ How to reference this article:

Arroyo Resino, D., Constante-Amores, A., Castro, M., & Navarro, E. (2024) School effectiveness and high reading achievement of Spanish students in PISA 2018: a machine learning approach. *Educación XX1*, 27(2), 223-251. <https://doi.org/10.5944/educxx1.38634>

Date received: 14/10/2023

Date accepted: 01/02/2024

Published online: 28/06/2024

ABSTRACT

In the last few decades, the analysis of school effectiveness has gained increasing importance in the field of education. This current research focuses on studying the factors of school effectiveness associated with high performance in reading comprehension. The sample is comprised of Spanish students who participated in PISA 2018. The dependent variable is high performance in reading comprehension, and a total of 159 predictors related to school effectiveness have been considered. The data were analyzed using the Random Forest algorithm and binary multilevel logistic regression. Among the key findings, it is highlighted

that the most important variables are process variables: enjoyment of reading and metacognition: evaluating credibility. Furthermore, the relative importance of context or input factors and process factors explains 41% and 38%, respectively, of the variance of the criterion variable. The final model (comprising both groups of factors) explains approximately 54% of reading success. In this model, the predictor that has the most significant effect is metacognition: evaluating credibility, which refers to the subject's ability to assess the quality and credibility of a text (for example, whether the information is valid, accurate, and impartial), with its effect being roughly double that of context or input variables. Among the main conclusions, it is highlighted that it is possible to increase the scarce number of high-performing students in the Spanish context through the development of educational policies that promote a love for reading and metacognitive capacity.

Keywords: PISA, high achievement, machine learning, school effectiveness

RESUMEN

En las últimas décadas, el análisis de la eficacia escolar ha adquirido una creciente importancia en el ámbito educativo. La presente investigación se centra en estudiar los factores de eficacia escolar asociados al alto rendimiento en la comprensión lectora. La muestra se encuentra conformada por los estudiantes españoles que participaron en PISA 2018. La variable criterio es el alto rendimiento en comprensión lectora y se ha contado con un total de 159 predictores relacionados con la eficacia escolar. Los datos se han analizado con el algoritmo de Random Forest y regresión logística binaria multinivel. Entre los principales resultados se destaca que las variables más importantes son las variables de proceso: placer por la lectura y metacognición: evaluar la credibilidad. Además, se demuestra la importancia relativa que tienen los factores de contexto o entrada y proceso explicando un 41% y 38%, respectivamente, de la varianza de la variable criterio. El modelo final (formado por ambos grupos de factores) explica aproximadamente el 54% del éxito lector. En este modelo, el predictor que tiene un mayor efecto es la metacognición: evaluar la credibilidad, referido a la capacidad del sujeto para evaluar la calidad y credibilidad de un texto (por ejemplo, si la información es válida, precisa e imparcial), siendo su efecto aproximadamente el doble que el de las variables de contexto o entrada. Entre las principales conclusiones se destaca la posibilidad de aumentar el escaso número de estudiantes de alto rendimiento en el contexto español mediante el desarrollo de políticas educativas que fomenten el placer por la lectura y la capacidad metacognitiva.

Palabras clave: PISA, alto rendimiento, machine learning, eficacia escolar

INTRODUCTION

School effectiveness is a standard issue in educational research (Creemers et al., 2022), to the extent that it is an area of study that has its own place in scientific literature and far from being an outdated tradition, it continues to be a fruitful and necessary field for reflection and research (Scheerens & Creemers, 2022).

Its origins date back to the Coleman Report (1966), which focused on the study of inequality in academic achievements and which highlights that the socio-economic context in educational results is of more significance than school variables (López-González et al., 2021). According to de la Orden et al. (1997), this phenomenon should be expressed by means of indicators that reflect the relationship between the achievements or results of the system and the goals and objectives set by this system. Therefore, school effectiveness does not only involve the level of performance attained by students, classes and individual educational centres (quality), but also the equitable distribution of learning outcomes among students with different background characteristics (equity) (Kyriakides et al., 2019); consequently, as indicated by Hu et al. (2021) it is fundamental to study the peculiarities of the top-performing group, in reference to the students who outperform the rest. Society, including national/state organisations and schools/teachers, should assume responsibility for offering learning opportunities and additional guidance to the more disadvantaged groups of students, in order to achieve equitable, high-quality education (Creemers et al., 2022). In this way, models of school effectiveness focus on studying the processes implemented by the educational centres that make a significant contribution to students' academic performance, as these are the basic criteria for judging educational effectiveness. Thus, the identification of educational factors associated with student performance is a key aspect in educational research into school effectiveness (Creemers et al., 2022; Murillo, 2007; Scheerens et al., 2013).

The results of international educational assessments, such the Programme for International Student Assessment (PISA), which assess the performance in competencies of students from different educational backgrounds, act as national indicators of effectiveness (Kyriakides et al., 2019). The questionnaires applied to students, families, teachers and head teachers have made it possible to conduct further in-depth research into school effectiveness, by analysing the school factors with the greatest effect extensively, accurately and rigorously. This may lead to the improvement of educational processes and policies in the different regions assessed (Murillo, 2007), a subject which has not yet been fully explored using this database (Kyriakides et al., 2019). Until now, school effectiveness has been studied by PISA from a dual perspective. Firstly, through research which analyses this phenomenon from the aspect of high and low levels of school effectiveness (Gamazo et al.,

2018 and Martínez-Abad et al., 2020). Secondly, through studies that focus on the most significant factors which influence performance in science by high- and low-performing students, such as the study by Hu et al. (2021) who used Creemers & Kyriakides' (2008) dynamic model of school effectiveness in an international assessment for the first time. However, as of today, there has been no analysis of the school effectiveness factors associated with high-performing students compared to other students, a subject of great interest in enabling strategies to be established to improve the quality of educational systems by increasing the percentages of this type of student.

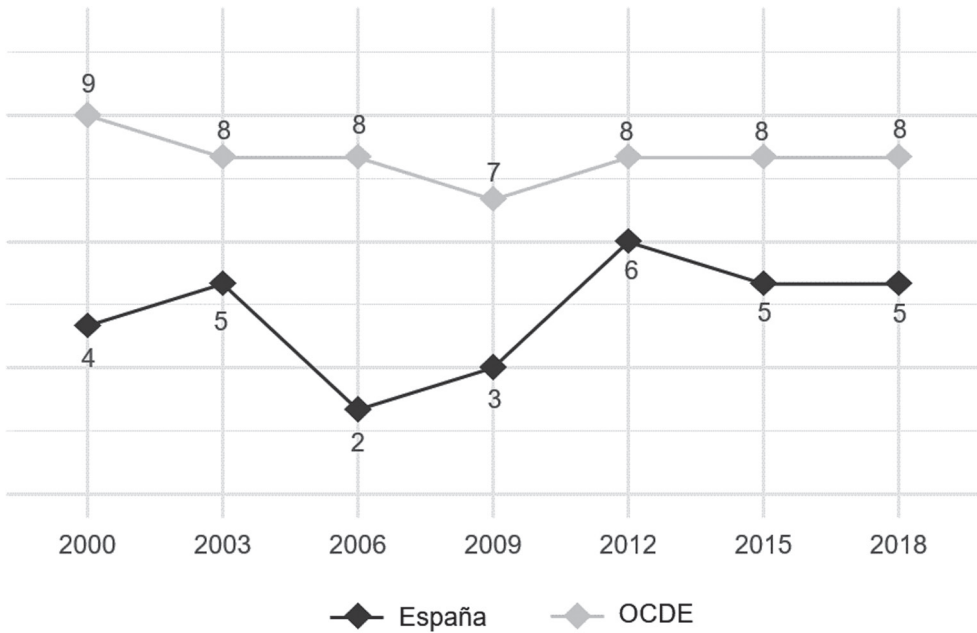
Of the three core competencies assessed triennially by PISA, reading comprehension is an essential tool in the educational field, as it allows students to access and understand reality by developing meaningful learning. Furthermore, it is an essential skill, as we constantly find ourselves producing and understanding texts; this is an everyday activity in the modern world and the basis of independent learning in the knowledge society (García et al., 2018; Molina, 2020).

In this sense, there seems to be a positive relationship between reading literacy and learning. García et al. (2018) establish a significative, positive correlation between the level of students' reading comprehension and their performance in four areas (Spanish Language, Mathematics, Social Sciences and Natural Sciences), which is not surprising, since reading skills are an essential tool for building meanings and knowledge (Gómez et al., 2014). Therefore, students with a low level of reading comprehension are usually low achievers since this competence directly affects learning (Viramontes et al., 2019).

In PISA 2018 there are six levels of reading comprehension, which consider students to be "high-performing" if they attain level 5 or 6. They are characterised as readers who can locate, organise and infer information, engage in critical reflection (level 5), make comparisons, fully understand texts, integrate information from different texts and even create abstract categories (level 6) (OECD, 2019). In Spain, the percentage of students in these "high performing" levels in 2018 was 5%, which is consistently below the average for OECD countries (8% in 2018), as shown in the following figure:

Figure 1

Evolution of the percentage of high-performing students in reading in PISA 2018



Source: Author's own compilation based on the database of the different PISA reports

Therefore, one of the main problems of the Spanish education system is that there are very few students at the higher levels in this competence. This is a genuine source of concern because it directly affects effectiveness and social equity, so consequently educational systems should be able to increase the percentage of students who attain these higher levels in order to achieve the maximum student performance (Gaviria, 2004).

Effectiveness factors associated with reading performance

The Context, Input, Process, and Output (CIPO) model by Scheerens (1991) synthesises the traditional idea of school effectiveness. There is extensive literature (Gamazo et al., 2018; Martínez-Abad, et al., 2020; Murillo, 2007) that classifies the factors of school effectiveness most commonly associated with student performance into contextual or background factors (student characteristics, teacher characteristics, infrastructures, etc.), process factors (teacher performance in the classroom and tutorials, didactic methodology, etc.) and output factors (academic

performance). In turn, these factors can be divided into two levels: students and educational centres.

Contextual or input factors

Among the student contextual factors that are most closely related to high levels of performance in reading comprehension is a favourable socio-economic and cultural environment (Campos and Arantegui, 2022). The research by Franco et al. (2016) emphasises the importance for reading comprehension of the number of books at home, the texts used to perform schoolwork, and parental supervision and support during these tasks.

The variables of student gender and immigration background are also significant. Regarding the former, the research conducted by Frutos and Santaren (2020) reveals greater reading comprehension among female students than among male students, a pattern that is observable in all the PISA and PIRLS reports, although in PISA the gender gap will have narrowed in recent reports. In relation to the latter, this variable has a major impact on reading performance. Specifically, native-born students score higher in reading than first-generation immigrant students (immigrants born in other countries) (Cordero et al., 2013).

As regards the educational centre factors that are most commonly linked to high levels of performance in reading comprehension, school ownership is significant in the area of contextual variables. Asensio-Muñoz et al. (2018) state that it is this educational centre variable that has the closest relationship to reading comprehension performance and that students who come from socio-economically disadvantaged homes have a lower level of performance in this competence. Another important variable is having changed schools two or more times: these student display lower performance in reading compared to those who have not changed schools or who have done so only once (Gamazo et al., 2018).

Process factors

Among the process factors, the student's academic expectations are of significance (Garrido et al., 2020 and Hu et al., 2021). Students who expect to complete a university degree have a greater chance of being high performers in reading literacy. The research by Franco et al. (2016), as well as that by Herrera et al. (2017) indicate the importance of a positive reading self-concept, as this has a favourable impact on performance. Another variable that is related to academic success is metacognition (Wu et al., 2020). In this sense, greater student metacognitive knowledge leads to higher achievement in reading literacy in the PISA

assessment (Artelt and Schneider, 2015). Specifically, Qi (2021) indicates that out of the different metacognitive strategies, those related to the process of summarising texts are the most important. Another extremely important factor is the enjoyment of reading, as students who spend more time on this activity display a higher level of performance in this competence (Akande & Oyedapo, 2018; Molina, 2020).

Regarding the process factors related to the centre, one of the most important variables is the school environment (Linnakyla et al., 2004). Also of significance is the use of ICT, which plays a fundamental role in high levels of proficiency in reading comprehension, since we are currently in the midst of the age of technology and students use these tools to read, so this situation should be exploited to promote reading (Rivera, 2013). Furthermore, Avendaño and Martínez (2013) emphasise the importance of ICT in establishing new scenarios where students can take an innovative approach to texts and interact with them within the framework of the digital age, which aids the development of reading literacy. Finally, another variable which is also noteworthy is teacher feedback (Hu et al., 2021), which is beneficial when it is high-quality, equitable and timely.

Based on the foregoing, the general objective of this research consists of studying the factors (context or background and process) associated with a high level of performance in reading comprehension within the framework of PISA 2018 using the Spanish sample. To achieve this, we formulated the following specific objectives:

1. Identify the factors with the greatest influence on high levels of reading performance.
2. Determine the relative contribution of the factors with the most impact on the student's chances of attaining a high level of achievement in reading literacy, bearing the hierarchical data structure in mind (level 1: student and level 2; school).

METHOD

A secondary analysis was conducted of the PISA 2018 international assessment data. Therefore, this is a quantitative study with a non-experimental design, based on cross-sectional data and limited to ex post facto research. It is worth mentioning that this research used a dual methodological strategy. Firstly, the Random Forest machine learning algorithm was used to conduct descriptive and exploratory analysis. Secondly, a predictive logistic model was run, bearing the hierarchical structure of the educational data in mind (level 1: student and level 2: school) (Lee, 2000).

Muestra

This research used the PISA 2018 database provided by the OECD (<https://www.oecd.org/pisa/data/2018database/>). A total of 79 countries and approximately 600000 students aged between 15 and 16 years participated in the most recent survey. The study sample is composed of the Spanish students (35943) who participated in the most recent survey of the PISA international assessment, with an average age of 15.836 years ($SD = 0.288$). In this group of students, 49.957% are female and 50.043% are male. It should be borne in mind that 22265 students (64.7%) attend public educational centres/public centres, 9722 (28.3%) are at privately-owned, publicly funded schools and 2410 (7%) are at private centres, in 17 Autonomous Communities. In the preliminary phase of machine learning, the entire dataset was utilized. Subsequently, during the multilevel logistic regression stage, schools with 20 or fewer students were omitted, resulting in a refined sample of 34,411 students across 976 schools.

Variables

The response variable in the study is Reading Comprehension in PISA 2018, which has been dichotomised (0 = average and low level of performance) and 1 = high level of performance) (see Table 1). A student is considered to be high-performing when they achieve level 5 or 6 in reading literacy in PISA 2018 (OECD, 2019).

Table 1

Number of high-performing Spanish students in reading literacy in PISA 2018

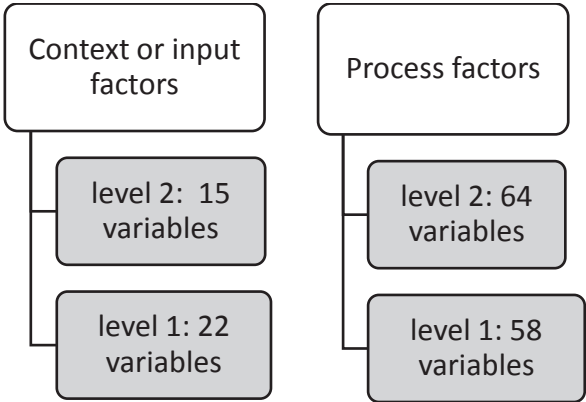
High level of reading performance	N	%
Yes	1797	5
No	34146	95

Source: Author's own compilation based on the OECD (2019).

With reference to the predictor variables, this research is based on a total of 159 predictors (see ANNEX 1), which have been grouped into context or input factors and process factors. Furthermore, these have in turn been grouped into two levels: student (level 1) and school (level 2) (see Figure 2). It is important to note that the predictors are composed of complex indices but that some direct variables from the questionnaire have also been included regarding the student, centre and

educational path. The selection of the context or input, and process factors was based on previous studies by Gamazo et al. (2018), Hu et al. (2021) and Martínez-Abad et al. (2020) which address school effectiveness in the PISA international assessment.

Figura 2
Diagram of the independent variables in the study



Procedure and data analysis

To meet the first objective, which consisted in identifying the factors most closely related to a high level of reading performance, we used the Random Forest classification algorithm, after conducting data pre-processing (Kassambara, 2018). This supervised machine learning algorithm was chosen for its high accuracy in identifying and ranking key predictor variables, as noted by Sterne (2018). Its implementation necessitates thorough and meticulous data pre-processing before execution, a step underscored by Kassambara (2018) for optimal results. To this purpose, firstly, the missing values in the database were imputed for the student and educational centre. We selected the method of multiple imputation by chained equations, as this is the most suitable method for obtaining accurate estimates (Sterne et al., 2009). The initial dataset of 159 predictor variables was narrowed down to 137 by removing those with over 20% missing data, following the criteria set by Medina and Galván (2007). Exceptions were made for key variables related to learning time: LMMINS (reading), MMINS (mathematics), and SMINS (science), despite their missing data percentages of 24%, 24%, and 25%, respectively. This decision is grounded in the critical importance of learning time as a process variable,

a relationship well-documented in the literature (Martinez-Abad et al., 2020; Hu et al., 2021). Secondly, the data were split randomly into two sets; the first set relates to training (60% of the data) and was used to fit the model and, subsequently, the validation sample (40%) was used to test the model performance (Raschka, 2015). Thirdly, the continuous variables were standardised with the purpose of avoiding predictors with a greater magnitude having a major impact on the model (Kassambara, 2018). Finally, we analysed the variance of the predictors, as it was necessary to eliminate the predictors variables with zero or near-zero variance. None were eliminated, as they all contain information, they have variability.

In order to obtain the best model, the hyperparameters were optimised by means of 10-fold cross-validation. This consisted in dividing the training set into subsets of the same size called folds. In the first iteration the model was fitted with all the observations except the first fold, which was used to predict. In the second the model was trained with all the observations except the second fold, which was used to predict and so on until the tenth iteration (Sarkar et al., 2018).

We used the Random Forest technique on each of the 10 plausible values (PV) and selected the one with the largest area under the curve (AUC), as this metric is the most appropriate when the levels of the criterion variable are unbalanced (Bonaccorso, 2017). The most accurate plausible value was nine (0.827), so consequently this PV was used to report the most significant variables. The great limitation of this technique is that there is no cut-off point for selecting the predictors with the greatest influence on the response variable (Sarkar et al., 2018); for this reason, in line with the study by Gorostiaga and Rojo-Álvarez (2016), who recommend that to select the optimum number of variables it is necessary to assess several sets of variables (20 and 30) and select the set with the best performance, we decided to examine the following sets of variables from the 137 predictors: 15, 20, 25, 30 and 35. In doing so, the aim was greater accuracy in indicating the optimal set of variables. The method of relative importance was selected to sort the variables.

The basis of the second objective was to determine the relative contribution of context or input and process factors which most influenced the chances of a student attaining a high level of achievement in reading literacy, considering the hierarchical data structure (level 1: student and level 2; school). As the criterion variable is of a qualitative, nominal and dichotomous nature, with a significant random variance in level 2 (educational centre) and the intraclass correlation coefficient being higher than 10% (Lee, 2000), we used the multilevel binary logistic regression technique. Prior to running the models, the assumption of multicollinearity was checked in two stages. Firstly, we correlated the variables using Spearman's correlation coefficient, as none of the variables met the assumption of normality. The work by Kassamabara (2018) states that the relationship is considered to be strong when values are larger

than 0.7. Table 2 shows that there are two associations with a high magnitude of correlation. Specifically, we eliminated the variables of the index's *parents' highest occupational status* and *household possessions*, as these have a lower relative importance than the *index of economic, social and cultural status*.

Table 2
Spearman's correlation

	Index of economic, social and cultural status
Index of parents' highest occupational status	0.815***
Household possessions	0.704***

Note. Difference is (*) significant with an $\alpha = .05$ (**) significant with an $\alpha = .01$; (***) significant with an $\alpha = .001$

Secondly, we checked that the Variance Inflation Factor was lower than 10 and as this was true in all cases, we concluded that the assumption of multicollinearity had been met.

Following this, we eliminated the educational centres with fewer than 20 subjects in order to conduct a multilevel analysis correctly (Gamazo et al., 2018). In particular, a total of 1532 subjects belonging to 133 centres were eliminated, so the final sample consisted of 34411 students. Four models were created, each based on a combination of key contextual (or input) factors and process variables as selected by the machine learning algorithm: model 0 or zero-inflated model, containing no predictors; model 1, composed of the context or input variables; model 2, composed of the process variables; and finally, model 3, composed of all the variables, in order to analyse the contribution of the process variables once the context or input factors had been checked. The odds ratio (Lee, 2000) was used to interpret the coefficients, with reference to the probability of a student having a high level of performance in reading comprehension and the probability of this not occurring. The Percentage of Explained Variance (PEV) was also calculated, which indicates the amount of variance that the model explains (equation 1).

$$PEV = \left(\frac{\sigma_{Zero-inflated\ model}^2 - \sigma_{Final\ model}^2}{\sigma_{Zero-inflated\ model}^2} \right) * 100 \quad \text{Equation 1}$$

Finally, to check the fit of the models, we used the AIC and BIC indices, as well as Deviance, in order to compare nested models; the significance of this reduced statistical model was calculated and the percentage of variance reduction was estimated (R^2) (Cameron & Windmeijer, 1997).

The variables were introduced into the model in the order obtained from the machine learning algorithm, as undertaken in the work by Arroyo et al. (2019), Fernández-Mellizo and Constante-Amores (2020) and Constante-Amores et al. (2021).

All the analyses were conducted using the statistical software R version 4.2.0 (R Core Team, 2022). The mice and lme4 packages were used to implement the method of multiple imputation by chained equations and multilevel binary logistic regression respectively. The Random Forest model was created using the machine learning library H2O, which is written in Java but can be used with the programming language R.

RESULTS

The results are delineated below, organized in alignment with the established objectives.

Specific objective 1: The most significant contextual or background factors

The following table shows how the first set of predictors composed of 15 variables is the set with the highest level of performance (AUC = 0.839), that is to say, the one that best represents the characteristics of high-performing students in reading literacy.

Table 3

Model performance in the validation sample of the five sets of candidate characteristics using random forests.

	AUC	Accuracy
15 variables	0.839	0.952
20 variables	0.837	0.947
25 variables	0.830	0.950
30 variables	0.830	0.945
35 variables	0.831	0.949

Table 4 shows the 15 predictors that are most closely linked to high levels of performance in reading comprehension, selected from a total of 137 contextual or input and process variables. Overall, it can be seen that all these variables are related to student characteristics. In relation to the context or input factors, the most significant variables are: *I like or enjoy reading, index of economic, social and cultural status and father's occupation*. Regarding the process variables, the most significant are *metacognition: credibility assessment, reading self-concept: perception of proficiency and reading self-concept: perception of difficulty*.

Table 4

The most significant variables associated with high levels of performance in reading comprehension.

			Relative importance (%)
Context or input	6 th	Index of economic, social and cultural status	27.432
	8 th	Father's occupation	20.300
	10 th	Index of parents' highest occupational status	13.569
	11 th	Household possessions	10.736
	12 th	Mother's occupation	10.506
	13 th	Cultural household possessions	10.497
Process	1 st	I like or enjoy reading	100.000
	2 nd	Metacognition: credibility assessment	86.520
	3 rd	Reading self-concept: perception of proficiency	61.827
	4 th	Reading self-concept: perception of difficulty	39.826
	5 th	Learning time in science (minutes per week)	32.459
	7 th	Metacognition: summarising	23.747
	9 th	Student's expected occupational status	19.379
	14 th	Use of ICT outside of school (for school activities)	7.167
	15 th	ICT as a subject in social interaction	4.127

Specific objective 2: Probability of attaining a high level of reading performance

Once the most important variables had been identified, we conducted the multilevel binary logistic analysis. Table 5 shows the results of the predictive models. In model 1, which was composed of the context or input variables, it is noticeable that

all of them are statistically significant with a positive directionality, and they explain approximately 41% of the variance of the dependent variable. The most significant effect occurs in the predictor *I like or enjoy reading*. Specifically, for each additional point in this predictor, there is a 100% increase in the probability of attaining a high level of reading performance. In model 2, composed solely of the process variables, as in the previous model, all of them are statistically significant, explaining 38.258% of the high level of performance in reading literacy. The greatest effect appears in the variable *metacognition: credibility assessment*. In particular, for each additional point in this predictor, there is a 110% increase in the probability of attaining a high level of performance in reading literacy. In model 3 all of the variables are included, with all of them being statistically significant. Furthermore, they move in a positive direction, except in *reading self-concept: perception of proficiency, use of ICT outside of school (for school activities)* and *TIC as a subject in social interaction*. The variable with the greatest effect is *metacognition: credibility assessment*. Once the context or input variables and the process factors have been controlled, the percentage of explained variance increases by around 13% compared to model 1, thereby producing a total explained variance of 54.676%.

Lastly, as regards the model fits, the model with the lowest AIC and BIC score is the model including all the predictors (model 3), where there is significant reduction in the variance, equivalent to R^2 11% compared to model 1. The reduction in the variance of model 1 is also significant compared to model 0 (12%) and to model 2 compared to the zero-inflated model (20%).

Table 5

Estimation of fixed effects with robust standard errors using multilevel binary logistic regression (odds ratio)

	Model 0 (zero-inflated)	Model 1 (context or input)	Model 2 (process)	Model 3 (all)
Intercept	0.045 ***	0.015 ***	0.002***	0.001***
Index of economic, social and cultural status		1.245 ***		1.152***
Father's occupation		1.008 ***		1.007***
Mother's occupation		1.007 ***		1.005***
Cultural household possessions		1.195 ***		1.145***
I like or enjoy reading			2.000 ***	1.369***
Metacognition: credibility assessment			2.100***	1.989***

	Model 0 (zero-inflated)	Model 1 (context or input)	Model 2 (process)	Model 3 (all)
Reading self-concept: perception of proficiency			1.875***	1.436***
Reading self-concept: perception of difficulty			0.782***	0.790***
Learning time in science (minutes per week)			1.002***	1.001***
Metacognition: summarising			1.777***	1.358***
Student's expected occupational status			1.027***	1.022***
Use of ICT outside of school (for school activities)			0.852***	0.809***
ICT as a subject in social interaction			0.788***	0.819***
Variance	0.528	0.312	0.326	0.241
PEV (%)		40.909	38.258	54.356
AIC	14881	13045	11884	11509
BIC	14895	13104	11969	11636
Loglik	-7563	-6515	-5932	-5739
Deviance	14850	13031	11864	11479
X ²		1818	2985	1552
Pr(>Chisq)		0.000	0.000	0.000

Note. Difference is (*) significant with an $\alpha = .05$ (**) significant with an $\alpha = .01$; (***) significant with an $\alpha = .001$

DISCUSSION AND CONCLUSIONS

In this study, there has been a deepening in the analysis of the factors (context or input and process) of school effectiveness associated with high performance in reading comprehension within the framework of PISA 2018, which is one of the major problems of the educational system.

Our findings identify 15 variables with a greater influence on high reading performance, among which the most important is the process variable referred to as the enjoyment of reading. These results are also observed in the works carried out by Akande and Oyedapo (2018), Franco et al. (2016), and Molina (2020), indicating

the importance of this predictor in academic reading success. In this sense, there is extensive literature (Gil, 2011) indicating the relevance of the family in fostering reading enjoyment. Additionally, the research of Butlen (2005) shows that the school also plays an important role. Therefore, educational measures must be implemented from the school to increase students' enjoyment of reading (Dezcallar et al., 2014), highlighting the work on students' reading habits.

It is also observed that the five most important predictors belong to process factors (enjoyment of reading, metacognition: evaluating credibility, reader self-concept understood as perception of competence and perception of difficulty and learning time in science - minutes per week). Therefore, considering these results, it seems that these factors play a very relevant role in high reading comprehension performance. This is contrary to extensive literature that states that high reading performance is primarily explained by context or input factors (Cordero et al., 2013; Franco et al., 2016). Although it coincides with the work of Martínez-Abad et al. (2020), in which the most important variables are related to students.

Likewise, it should be noted that in our study, student gender and immigrant status are not important variables in high reading performance, unlike other studies (de Frutos & Santaren, 2020, and Cordero et al., 2013), even Hu et al. (2021) demonstrate that these predictors are important in discriminating and differentiating between students with high and low scientific performance in PISA 2015.

Regarding the second objective based on determining the relative contribution of the factors that most affect the probability of a student achieving high performance in reading competence, the importance of both context or input factors and process factors in reading performance can be appreciated. Each set of predictors explains a very similar percentage variance of 39% and 36%, respectively. The final model explains approximately 52% of the variability of the criterion variable. These results complement the conclusions of Coleman (1966) who showed the preponderance of context or input factors such as socioeconomic status and ethnicity.

In the final model, the predictor with the largest effect is metacognition (odds ratio = 1.989), referring to the subject's ability to evaluate the quality and credibility of a text (for example, if the information is valid, accurate, and impartial, etc.) incorporated for the first time in PISA 2018 (OECD, 2019). This predictor has approximately twice the effect of certain context or input variables that have been of great importance in the scientific literature, such as the index of economic, social, and cultural status, father's occupation, mother's occupation, and cultural possessions of the home (Barrera et al., 2019).

In fact, the other variable related to metacognition (summarizing) also has a greater effect than this set of predictors.

In this sense, the great importance of metacognition in high reading performance stands out compared to other more socio-economic variables, following the line of Qi (2021) and Wu et al. (2020), which point out the important role played by metacognitive variables in high reading performance. Therefore, it is essential that metacognitive capacity be worked on in schools.

Likewise, as in the research of Franco et al. (2016) and Herrera et al. (2017), it should be emphasized, the relevance of reader self-concept (perception of competences and perception of difficulty) in high reading scores in PISA 2018. Also, the academic and professional educational expectations of students about reading competence play a very important role, as was the case in the study by Hu et al. (2021).

From a methodological point of view, this research not only indicates the most important variables but also carries out a predictive model (multilevel binary logistic regression) exhaustively addressing this educational phenomenon for a specific group, which are the high-achieving students compared to the rest. Also, this study represents a methodological advancement in the phenomenon of school effectiveness factors, as it employs for the first time a machine learning approach in the Spanish context. This is a complementary approach to works that use data mining techniques (Martínez-Abad et al., 2020) and binary logistic regression (Gamazo et al., 2018).

With the desirable goal for the Spanish educational system of increasing the low number of high-performing students, educational policies could be developed that promote reading enjoyment and the learning of metacognitive tools in the field of reading, well-known and necessary tools such as evaluating the credibility of texts and summarizing, accredited by the results of this research. Additionally, it is necessary for schools to work on improving self-concept, as it affects students' academic expectations and, therefore, the performance of different subjects (Carrillo et al., 2022).

As prospective research, it would be necessary to determine the factors (context or input and process) associated with high reading performance in other countries and contexts, such as some in Southern Europe (Portugal, Italy, France, Cyprus, Greece, and Malta). Regarding the limitations of the study, it should be noted that the results obtained cannot be interpreted in terms of causality; for this, structural equation models or experimental studies would be necessary. Additionally, the PISA evaluation has been methodologically reviewed and criticized, as shown by the work of Fernández-Cano (2016) alluding, among other issues, to problems of validity of measurement instruments and the way of estimating scores on the performance scale. Another problem is the questionnaires used to collect context information (Jornet, 2016), which lack a clear theory to develop the different constructs measured and the lack of information about their psychometric characteristics.

REFERENCIAS BIBLIOGRÁFICAS

- Akande, S., & Oyedapo, R. (2018). Developing the reading habits of secondary school students in Nigeria: the way forward. *International Journal of Library Science*, 7(1), 15-20.
- Arroyo, D., Constante-Amores, I. A., & Asensio, I. (2019). La repetición de curso a debate: un estudio empírico a partir de PISA 2015. *Educación XX1*, 22(2), 69-92. <https://doi.org/10.5944/educxx1.22479>
- Artelt, C., & Schneider, W. (2015). Cross-country generalizability of the role of metacognitive knowledge in students' strategy use and reading competence. *Teachers College Record*, 117(1), 1–32. <https://bit.ly/4akHyKl>
- Asensio-Muñoz, I., Carpintero, E., Expósito, E., & López-Martín, E. (2018). ¿Cuánto oro hay entre la arena? Minería de datos con los resultados de España en PISA 2015. *Revista de Educación*, 270, 225-245
- Avendaño, I., & Martínez, D. (2013). Competencia lectora y el uso de las nuevas tecnologías de la información y comunicación. *Revista Escenarios*, 11(1), 7-22.
- Barrera, J. E., Polanco, J. G. C., & Acosta, J. D. (2019). Comprensión lectora de estudiantes universitarios. Factores asociados y mecanismos de acción. *Revista Venezolana de Gerencia*, 24(87), 874-889.
- Bonaccorso, G. (2017). *Machine learning algorithms*. Packt Publishing Ltd.
- Butlen, M. (2005). Paradoja de la lectura escolar. *Revista de Educación*, núm. Extraordinario, 139-151.
- Cameron, A. C., & Windmeijer, F. A. (1997). An R-squared measure of goodness of fit for some common nonlinear regression models. *Journal of Econometrics*, 77(2), 329-342.
- Campos, I. O., & Arantegui, M. (2022). Exploración de la mediación parental en el uso de las TIC y su correlación con la comprensión lectora del alumnado preadolescente. *Lenguaje y Textos*, 55, 43-54. <https://doi.org/10.4995/lyt.2022.15948>
- Carrillo-López, P. J., Constante-Amores, A., Arroyo-Resino, D., & Sánchez-Munilla, M. (2022). Self-concept and academic achievement in primary school: a predictive study. *International Journal of Education in Mathematics, Science and Technology*, 10(4), 1057-107
- Constante-Amores, A., Florenciano Martínez, E., Navarro, E., & Fernández-Mellizo, M. (2021). Factores asociados al abandono universitario. *Educación XX1*, 24(1), 17-44.
- Coleman, J. (1966). *Equality of educational opportunity*. US Government Printing Office.

- Cordero, J. M., Crespo, E., & Pedraja, F. (2013). Rendimiento educativo y determinantes según PISA: una revisión de la literatura en España. *Revista de Educación*, 362, 273-297
- Creemers, B. P., Peters, T., & Reynolds, D. (2022). *School effectiveness and school improvement*. Routledge
- de Frutos, S. F., & Santaren, V. R. (2020). El papel del sexo en comprensión lectora. Evidencias desde PISA y PIRLS. *Revista de Investigación en Educación*, 18(2), 99-117.
- de la Orden, A., Asensio, I., Carballo, R., Fernández Díaz, J., Fuentes, A., García Ramos, J. M., & Guardia, S. (1997). Desarrollo y validación de un modelo de calidad universitaria como base para su evaluación. *RELIEVE - Revista Electrónica De Investigación Y Evaluación Educativa*, 3(1).
- Dezcallar, T., Clariana, M., Cladellas, R., Badia, M., & Gotzens, C. (2014). La lectura por placer: su incidencia en el rendimiento académico, las horas de televisión y las horas de videojuegos. *Ocnos: Revista de Estudios sobre Lectura*, (12), 107-116.
- Franco, M., Cárdenas, R., & Santrich, E. (2016). Factores asociados a la comprensión lectora en estudiantes de noveno grado de Barranquilla. *Psicogente*, 19(36), 296-310. <https://doi.org/10.17081/psico.19.36.1299>
- Gamazo, A., Martínez-Abad, F., Olmos-Miguelañez S., & Rodríguez-Conde M. J. (2018). Evaluación de los factores relacionados con la eficacia escolar en PISA. Un análisis multinivel. *Revista de Educación*, 379, 56-84.
- García M. A., Arévalo M. A., & Hernández C. A. (2018). La comprensión lectora y el rendimiento escolar. *Cuadernos de Lingüística Hispánica*, (32), 155- 174.
- Garrido, R., Gallo-Rivera, M. T., & Martínez-Gautier, D. (2020). ¿Cuáles son y cómo operan los determinantes del fracaso escolar? Replanteando las políticas públicas para el caso de España y sus regiones. *Revista Internacional de Ciencias del Estado y de Gobierno*, 1(4), 509-540.
- Gaviria, J. L. (2004). La situación española: el rendimiento de los estudiantes. In G. Haug, J. L. Gaviria, C. Lomas, M.D. de Prada, & D. Gil (Eds.), *El rendimiento de los estudiantes al final de la educación obligatoria: objetivos europeos y situación española* (pp.18- 83). Santillana.
- Gil, J. (2011). Hábitos lectores y competencias básicas en el alumnado de educación secundaria obligatoria. *Educación XX1*, 14(1), 117-134.
- Gómez, I. M., García, J. A., Vila, J. O., Elosúa, M. R., & Rodríguez, R. (2014). The dual processes hypothesis in mathematics performance: beliefs, cognitive reflection, working memory and reasoning. *Learning and Individual Differences*, (29), 67-73.

- Gorostiaga, A., & Rojo-Álvarez, J. L. (2016). On the use of conventional and statistical-learning techniques for the analysis of PISA results in Spain. *Neurocomputing*, 171, 625-637.
- Herrera, J. C., Treviño, A., & Navarrete, G. (2017). *Factores de riesgo asociados a falta de competencia para la comprensión lectora en niños de primaria* [Comunicación]. X Congreso Nacional de Investigación Educativa.
- Hu, J., Peng, Y., & Ma, H. (2021). Examining the contextual factors of science effectiveness: a machine learning-based approach. *School Effectiveness and School Improvement*, 33, 21-50. <https://doi.org/10.1080/09243453.2021.1929346>
- Kassambara, A. (2018). *Machine learning essentials: Practical guide in R*. Sthda.
- Kyriakides, L., Charalambous, E., Creemers, B. P. M., Antoniou, P., Devine, D., Papastyliaou, D., & Fahie, D. (2019). Using the dynamic approach to school improvement to promote quality and equity in education: a European study. *Educational Assessment, Evaluation and Accountability*, 31(1), 121-149. <https://doi.org/10.1007/s11092-018-9289-1>
- Lee, V. E. (2000). Using hierarchical linear modeling to study social contexts: the case of school effects. *Educational Psychologist*, 35(2), 125-141.
- Linnakyla, P., Malin, A., & Taube, K. (2004). Factors behind low reading literacy achievement. *Scandinavian Journal of Educational Research*, 48(3), 231-249.
- López-González, E., Navarro, E., García San Pedro, M. J., Lizasoain, L., & Tourón, J. (2021). Estudio de la eficacia escolar en centros educativos de primaria mediante el uso de modelos jerárquicos lineales. *Bordón*, 73(1), 59-80. <https://doi.org/10.13042/Bordon.2021.80530>
- Martínez-Abad, F., Gamazo, A., & Rodríguez-Conde, M. J. (2020). Educational data mining: identification of factors associated with school effectiveness in PISA assessment. *Studies in Educational Evaluation*, 66, Artículo 100875.
- Medina, F., & Galván, M. (2007). *Imputación de datos: teoría y práctica*. Cepal.
- Molina, I. (2020). Comprensión lectora y rendimiento escolar. *Revista Boletín Redipe*, 9(1), 121-131.
- Murillo, F. J. (Coord.) (2007). *Investigación iberoamericana sobre eficacia escolar*. Convenio Andrés Bello.
- Organización para la Cooperación y el Desarrollo Económicos. (2019). *PISA 2018 assessment and analytical framework*. OECD publishing.
- Qi, X. (2021). Effects of self-regulated learning on student's reading literacy: evidence from Shanghai. *Frontiers in Psychology*, 11, Artículo 555849.
- Rivera, J. (2013). Punto de encuentro entre los jóvenes y la lectura: estrategia constructivista para fortalecer su comportamiento lector. *Simbiosis Estudiantil* 7(1).

- Sarkar, D., Bali, R., & Sharma, T. (2018). Practical machine learning with Python. A problem-solvers guide to building real-world intelligent systems. Apress.
- Scheerens, J. (1991). Process indicators of school functioning: a selection based on the research literature on school effectiveness. *Studies in Educational Evaluation*, 17(2), 371–403. [https://doi.org/10.1016/S0191-491X\(05\)80091-4](https://doi.org/10.1016/S0191-491X(05)80091-4).
- Scheerens, J., & Creemers, B. P. (2022). *School effectiveness and school improvement*. Routledge
- Scheerens, J., Witziers, B., & Steen, R. (2013). A meta-analysis of school effectiveness studies. *Revista de Educación*, 361, 619-645. <https://doi.org/10.4438/1988-592X-RE-2013-361-235>
- Sterne, J. A., White, I. R., Carlin, J. B., Spratt, M., Royston, P., Kenward, M. G., & Carpenter, J. R. (2009). Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ*, 338, 23-93
- Viramontes, E., Amparán, A., & Núñez, L. D. (2019). Comprensión lectora y el rendimiento académico en educación primaria. *Investigaciones sobre Lectura*, 12, 65-82.
- Wu, Y. J., Carstensen, C. H., & Lee, J. (2020). A new perspective on memorization practices among East Asian students based on PISA 2012. *Educational Psychology*, 40(5), 643-662.

ANNEX 1

Table 1*Context or input variables*

Level 2: centre	SC001Q01TA	Which of the following best describes the community in which your school is located?
	SCHLTYPE	School Ownership
	STRATIO	Student-Teacher ratio
	SCHSIZE	School Size (Sum)
	CLSIZE	Class Size
	EDUSHORT	Shortage of educational material
	STAFFSHORT	Shortage of educational staff
	RATCMP1	Number of available computers per student at modal grade
	RATCMP2	Proportion of available computers that are connected to the Internet
	TOTAT	Total number of all teachers at school
	PROATCE	Index proportion of all teachers fully certified
	PROAT5AB	Index proportion of all teachers ISCED LEVEL 5A Bachelor
	PROAT5AM	Index proportion of all teachers ISCED LEVEL 5A Master
	PROAT6	Index proportion of all teachers ISCED LEVEL 6
	SC048Q03NA	Percentage <national modal grade for 15-year-olds>: Students from socioeconomically disadvantaged homes
	SC048Q01NA	Percentage <national modal grade for 15-year-olds>: Students whose <heritage language> is different from <test language>
	ST003D02T	Student (Standardized) Birth - Month
	ST004D01T	Student (Standardized) Gender
Level 1: student	LANGN	Language at home (3-digit code)
	MISCED	Mother's Education
	FISCED	Father's Education
	HISCED	Highest Education of parents
	PARED	Index highest parental education in years of schooling
	PAREDINT	Index highest parental education (international years of schooling scale)
	BMMJ1	ISEI of mother

	BFMJ2	ISEI of father
	HISEI	Index highest parental occupational status
	IMMIG	Index Immigration status
	DURECEC	Duration in early childhood education and care
	SCCHANGE	Number of school changes
	CHANGE	Number of changes in educational biography
	ESCS	Index of economic, social and cultural status
Level 1: student	WEALTH	Family wealth
	CULTPOSS	Cultural possessions at home
	ICTRES	ICT resources
	ICTHOME	ICT available at home
	HOMEPOS	Home possessions
	HEDRES	Home educational resources
	ST004D01T	Student (Standardized) Gender
	LANGN	Language at home (3-digit code)
	MISCED	Mother's Education (ISCED)

Table 2
Level 2 process variables

	Learning difficulties	STUBEHA TEACHBEHA	Student behaviour hindering learning Teacher behaviour hindering learning
Level 2: school	Family participation	SC064Q01TA	Proportion of parents: Discussed their child's progress with a teacher on their own initiative
		SC064Q02TA	Proportion of parents: Discussed their child's progress on the initiative of one of their child's teachers
		SC064Q03TA	Proportion of parents: Participated in local school government (e.g. parent council or school management committee)
		SC064Q04TA	Proportion of parents: Volunteered in physical or extra-curricular activities [...]
	School assessment	SC037Q01TA	Quality assurance at school: Internal evaluation/Self-evaluation
		SC037Q02TA	Quality assurance at school: External evaluation
		SC037Q03TA	Quality assurance at school: Written specification of the school's curricular profile and educational goals
		SC037Q04TA	Quality assurance at school: Written specification of student performance standards
		SC037Q05NA	Quality assurance at school: Systematic recording of data such as [...] attendance and professional development
		SC037Q06NA	Quality assurance at school: Systematic recording of student test results and graduation rates
		SC037Q07TA	Quality assurance at school: Seeking written feedback from students (e.g. regarding lessons, teachers or resources)
		SC037Q08TA	Quality assurance at school: Teacher mentoring
		SC037Q09TA	Quality assurance at school: Regular consultation aimed at school improvement [...] over a period of at least six months
		SC037Q010NA	Quality assurance at school: Implementation of a standardised policy for reading subjects [...]

Level 2: school	Student assessment	SC154Q01HA	School's use of assessments of students: To guide students' learning
		SC154Q02WA	School's use of assessments of students: To inform parents about their child's progress
		SC154Q03WA	School's use of assessments of students: To make decisions about students' retention or promotion
		SC154Q04WA	School's use of assessments of students: To group students for instructional purposes
		SC154Q05WA	School's use of assessments of students: To compare the school to <district or national> performance
		SC154Q06WA	School's use of assessments of students: To monitor the school's progress from year to year
		SC154Q07WA	School's use of assessments of students: To make judgements about teachers' effectiveness
		SC154Q08WA	School's use of assessments of students: To identify aspects of instruction or the curriculum that could be improved
		SC154Q09HA	School's use of assessments of students: To adapt teaching to the students' needs
	Autonomy of the centre	SC154Q010WA	School's use of assessments of students: To compare the school with other schools
		SC154Q011HA	School's use of assessments of students: To award certificates to students
		SC165Q01HA	Teachers' practices: [...] students learn about the histories of diverse cultural groups that live in <country of test>
		SC165Q02HA	Teachers' practices: [...] students learn about the histories of diverse cultural groups that live in other countries.
		SC165Q03HA	Teachers' practices: [...] students learn about the cultures [...] of [...] groups that live in <country of test>
		SC165Q04HA	Teachers' practices: [...] students learn about different cultural perspectives on historical and social events.
		SC165Q05HA	Teachers' practices: Our school supports activities that encourage students' expression of diverse identities [...]
		SC165Q06HA	Teachers' practices: Our school offers an exchange programme with schools in other countries.

Level 2: school	Autonomy of the centre	SC165Q07HA	Teachers' practices: Our school organises multicultural events (e.g. cultural diversity day).
		SC165Q08HA	Teachers' practices: In our school, we celebrate festivities from other cultures.
		SC165Q09HA	Teachers' practices: In our school, students are encouraged to communicate with people from other cultures via [...]
		SC165Q09HA	Teachers' practices: Our school adopts different approaches to educate students about cultural differences [...]
		SC165Q10HA	Teachers' practices: [...] students learn about the histories of diverse cultural groups that live in <country of test>.
		SC156Q01HA	At school: Its own written statement about the use of digital devices
		SC156Q02HA	At school: Its own written statement specifically about the use of digital devices for pedagogical purposes
		SC156Q03HA	At school: A programme to use digital devices for teaching and learning in specific subjects
		SC156Q04HA	At school: Regular discussions with teaching staff about the use of digital devices for pedagogical purposes
		SC156Q05HA	At school: A specific programme to prepare students for responsible Internet behaviour
		SC156Q06HA	At school: A specific policy about using Social Networks (<Facebook>, etc.) in teaching and learning
		SC156Q07HA	At school: A specific programme to promote collaboration on the use of digital devices among teachers
		SC156Q08HA	At school: Scheduled time for teachers to meet to share, evaluate or develop instructional materials and [...]
		SC155Q01HA	Capacidad de la escuela mediante dispositivos digitales: El número de dispositivos digitales conectados a Internet es suficiente
		SC155Q02HA	Capacidad del centro escolar mediante dispositivos digitales: El ancho de banda o la velocidad de Internet del centro es suficiente.
		SC155Q03HA	School's capacity using digital devices: The number of digital devices connected to the Internet is sufficient

Level 2: school	Autonomy of the centre	SC155Q04HA	School's capacity using digital devices: The school's Internet bandwidth or speed is sufficient
		SC155Q05HA	School's capacity using digital devices: The number of digital devices for instruction is sufficient
		SC155Q06HA	School's capacity using digital devices: Digital devices [...] are sufficiently powerful in terms of computing capacity
		SC155Q07HA	School's capacity using digital devices: The availability of adequate software is sufficient
		SC155Q08HA	School's capacity using digital devices: Teachers have the [...] skills to integrate digital devices in instruction
		SC155Q09HA	School's capacity using digital devices: Teachers have sufficient time to prepare lessons integrating digital devices
		SC155Q10HA	School's capacity using digital devices: Effective professional resources for teachers to learn how to use digital [...]
		SC155Q11HA	School's capacity using digital devices: An effective online learning support platform is available
		SC042Q01TA	School's equity-oriented policies: These students attend regular classes and receive additional periods of [...]
		SC042Q02TA	School's equity-oriented policies: Before transferring to regular classes, [...] preparatory programme aimed at [...]
		SC150Q01IA	School's equity-oriented policies: Before transferring to regular classes, [...] instruction in school subjects [...]
		SC150Q02IA	School's equity-oriented policies: These students receive [...] amounts of instruction in their <heritage language>
		SC150Q03IA	School's equity-oriented policies: Class size is reduced to cater to the special needs of these students.
		SC150Q04IA	School's equity-oriented policies: These students attend regular classes and receive additional periods of [...]
		SC150Q05IA	School's equity-oriented policies: Before transferring to regular classes, [...] preparatory programme aimed at [...]

Level 1: Student	EMOSUPS	Parents' emotional support perceived by student (WLE)
	ST225Q001HA	Do you expect to complete? <ISCED level 2>
	ST225Q002HA	Do you expect to complete? <ISCED level 3B or C>
	ST225Q003HA	Do you expect to complete? <ISCED level 3A>
	ST225Q005HA	Do you expect to complete? <ISCED level 5B>
	ST225Q006HA	Do you expect to complete? <ISCED level 5A or 6>
	EC154Q001A	Do you currently attend additional instruction? [Enrichment lessons] in [Test language]
	EC154Q002A	Do you currently attend additional instruction? [Enrichment lessons] in mathematics
	EC154Q003A	Do you currently attend additional instruction? [Enrichment lessons] in [science]
	EC154Q004HA	Do you currently attend additional instruction? [Enrichment lessons] in foreign language
	EC154Q005A	Do you currently attend additional instruction? [Remedial lessons] in [Test language]
	EC154Q006A	Do you currently attend additional instruction? [Remedial lessons] in mathematics
	EC154Q007A	Do you currently attend additional instruction? [Remedial lessons] in [science]
	EC154Q008HA	Do you currently attend additional instruction? [Remedial lessons] in foreign language
	EC154Q009A	Do you currently attend additional instruction? Lessons to improve your [study skills]
	EC158Q001HA	On the most recent day you attended school, how long did you study in the morning before going to school? Hours
	EC158Q002HA	On the most recent day you attended school, how long did you study in the morning before going to school? Minutes
	EC159Q001HA	On the most recent day you attended school, how long did you study after leaving school? Hours
	EC159Q002HA	On the most recent day you attended school, how long did you study after leaving school? Minutes
	WB176Q001HA	When was the very last time you did your homework/studied for school?
	BSMJ	Student's expected occupational status (SEI)
	MMINS	Learning time (minutes per week) - <Mathematics>
	LMINS	Learning time (minutes per week) - <test language>
	SMINS	Learning time (minutes per week) - <science>
	TMINS	Learning time (minutes per week) - in total

Level 1: Student	UNDREM	Meta-cognition: understanding and remembering
	METASUM	Meta-cognition: summarising
	METASPAM	Meta-cognition: assess credibility
	ICTSCH	ICT available at school
	DISCLIMA	Disciplinary climate in test language lessons (WLE)
	TEACHSUP	Teacher support in test language lessons (WLE)
	DIRINS	Teacher-directed instruction (WLE)
	PERFEED	Perceived feedback (WLE)
	ADAPTIVITY	Adaptation of instruction (WLE)
	TEACHINT	Perceived teacher's interest (WLE)
	JOYREAD	Joy/Like reading (WLE)
	STIMREAD	Teacher's stimulation of reading engagement perceived by student (WLE)
	SCREADCOMP	Self-concept of reading: Perception of competence (WLE)
	SCREADDIFF	Self-concept of reading: Perception of difficulty (WLE)
	PISADIFF	Perception of difficulty of the PISA test (WLE)
	PERCOMP	Perception of competitiveness at school (WLE)
	PERCOOP	Perception of cooperation at school (WLE)
	ATTINACT	Attitude towards school: learning activities (WLE)
	COMPETE	Competitiveness (WLE)
	RESILIENCE	Resilience (WLE)
	MASTGOAL	Mastery goal orientation (WLE)
	BELONG	Subjective well-being: Sense of belonging to school (WLE)
	ENTUSE	ICT use outside of school (leisure) (WLE)
	HOMESCH	Use of ICT outside of school (for school work activities) (WLE)
	USESCH	Use of ICT at school in general (WLE)
	INTICT	Interest in ICT (WLE)
	COMPICIT	Perceived ICT competence (WLE)
	AUTICT	Perceived autonomy related to ICT use (WLE)
	SOIAICT	ICT as a topic in social interaction (WLE)
	ICTCLASS	Subject-related ICT use during lessons (WLE)
	ICTOUTSIDE	Subject-related ICT use outside of lessons (WLE)
	GFOFAIL	General fear of failure (WLE)
	DISCRIM	Students' perception of discrimination at school

