Contribution of machine learning to the analysis of grade repetition in Spain: A study based on PISA data

Contribución del machine learning *al análisis de la repetición escolar en España: un estudio con datos PISA*

Alexander CONSTANTE-AMORES. Associate Lecturer. Universidad Camilo José Cela, Spain (*israelalexander.constante@ucjc.es*).

Delia ARROYO-RESINO, PhD. Lecturer. Universidad Complutense de Madrid, Spain (*garroy01@ucm.es*).

María SÁNCHEZ-MUNILLA, PhD. Research Assistant. Universidad Complutense de Madrid, Spain (*marsan27@ucm.es*).

Inmaculada ASENSIO-MUÑOZ, PhD. Senior Lecturer. Universidad Complutense de Madrid, Spain (*macu@edu.ucm.es*).

Abstract:

Introduction: The rate of grade repetition is excessively high in Spain despite being a controversial measure. In order to obtain evidence to contribute to reducing it in compulsory education, the present work is an in-depth study of the PISA 2018 context indices that are most closely linked to this phenomenon. Method: With the sample of Spanish students (n = 35 943), we used an automatic machine learning method to select and order the predictors, and multilevel logistic regression (students and centres) to quantify the contribution of each one. Results: For each educational stage we obtained the 30 most significant contextual variables, which explain 65.5% of the grade repetition variance in primary education and almost 55.7% in secondary education. Conclusions: The main indicators are principally at student level, which suggests the suitability of psychoeducational interventions based more on individual support than general policies. This gives rise to potentially more efficient and equitable measures than grade repetition, aimed at, for example, the management of learning time or academic/professional guidance, and predictors with specific differential significance at each stage. Methodologically, the study contributes to improving the specification of predictive models.

Date of receipt of the original: 2024-03-28.

Date of approval: 2024-07-05.

Please, cite this article as follows: Constante-Amores, A., Arroyo-Resino, D., Sánchez-Munilla, M., & Asensio-Muñoz, I. (2024). Contribution of machine learning to the analysis of grade repetition in Spain: A study based on PISA data [Contribución del *machine learning* al análisis de la repetición escolar en España: un estudio con datos PISA]. *Revista Española de Pedagogía, 82* (289), 539-562. https://doi.org/10.22550/2174-0909.4014

Keywords: PISA, grade repetition, machine learning, contextual variables, multilevel logistic regression, compulsory education.

Resumen:

Introducción: La repetición de curso tiene una excesiva incidencia en España a pesar de ser una medida controvertida. A fin de obtener evidencias que contribuyan a su reducción en la educación obligatoria, el presente trabajo profundiza en el estudio de los índices de contexto de PISA 2018 más vinculados con dicho fenómeno. Método: Con la muestra de estudiantes españoles (n = 35 943), se utiliza un método de aprendizaje automático para seleccionar y ordenar los predictores, y una regresión logística multinivel (estudiantes y centros) para cuantificar la contribución de cada uno. Resultados: Se obtienen las 30 variables de contexto más relevantes en cada etapa educativa, que explican el 65.5% de la varianza de la repetición en primaria y casi el 55.7% en secundaria. Conclusiones: Los principales indicadores son sobre todo del nivel de estudiantes, lo que sugiere la idoneidad de intervenciones psicoeducativas basadas en el apoyo individualizado más que en políticas generalizadas. De ahí emergen medidas potencialmente más eficientes y equitativas que la repetición, centradas, por ejemplo, en la gestión del tiempo de aprendizaje o en la orientación académico-profesional, así como predictores con importancia específica diferencial en cada etapa. En el ámbito metodológico, el estudio hace una aportación a la mejora de la especificación de los modelos predictivos.

Palabras clave: PISA, repetición escolar, aprendizaje automático, variables de contexto, regresión logística multinivel, educación obligatoria.

1. Introduction

As is generally known, grade repetition is a strategy that consists of retaining the student at a level of a specific educational stage when they have not proved that they have acquired the minimum levels of knowledge (Jimerson & Ferguson, 2007; López & García, 2020). Nevertheless, it is a controversial measure, that some authors consider to be positive because it helps lower-performing students to improve their skills and gain maturity (Battisttin & Schizzerotto, 2019; Ikeda & García, 2014; Valbuena et al., 2021) and that others see as a merely disciplinary measure (Schwerdt et al., 2017). Furthermore, although this issue has been the subject of synthesis studies, systematic reviews and meta-analysis covering practically the entire twentieth century (Allen et al., 2009; Jimerson 2001), there are doubts as to its effectiveness. Moreover, it even seems to be detrimental, as it is associated with a decrease in student performance (Asensio et al., 2018; Hattie, 2017; López et al., 2018, Nieto-Isidro & Martínez-Abad, 2023) and is linked to negative aspects, such as frustration (López-Rupérez et al., 2021), feelings of discrimination

(Van Canegem et al., 2021) or a lowering of self-confidence, academic self-concept or motivation, thus affecting academic expectations (González-Nuevo et al., 2023; Mathys et al., 2019; Peixoto et al., 2016).

The corresponding regulations in Spain take this situation into account and the Organic Law 1/1990 of 3 October, on the General Organisation of the Educational System (LOGSE) established that repetition is a measure that focuses on exceptional diversity. Although, in fact, its use is not so exceptional. As reported by the Ministerio de Educación y Formación Profesional [Ministry of Education and Vocational Training] (2022), in the first stage of secondary education the repetition rate is 8.7% and in primary education 2.4% of students have repeated a year, a percentage that is only surpassed in the European Union by Portugal (3.6%), the Slovak Republic (2.9%) and Austria (2.9%). Furthermore, based on data from the Programme for International Student Assessment (PISA) for 2009, 2015 and 2018, López and García (2020), who compare the percentage of 15-year-old students who have repeated at least once in the European Union, conclude that Spain, despite having considerably reduced the rate between 2009 and 2018, in the latter year still has a rate of 28.71%. That puts the country in third place, only behind Luxembourg (32.02%) and Belgium (29.06%), when the average for the Organisation for Economic Co-operation and Development (OECD) stands at 11.4% and for Finland it is 3.2%.

In view of the above, further research into this phenomenon is necessary, with the ultimate aim of providing information that makes it possible to act based on the evidence, not only from a political perspective, but also through educational practice, thereby contributing to data optimisation so that Spain can get closer, at least, to the OECD average.

To this effect, the present study takes the line of a search for predictors by means of secondary exploitation of the PISA data, for which there is already research based on previous PISA publications, showing that grade repetition has many causes (Arroyo et al., 2019; Carabaña, 2013; Cordero et al., 2014).

In line with the classification of background questionnaires used by PISA (OECD, 2019), shown in Figure 1 below, there is a brief summary of the state of affairs regarding the most significant variables of the three constructs.

Regarding precedents, there is evidence that, above all, the students who repeat grades usually come from underprivileged socio-economic backgrounds (López & García, 2020). However, immigration background shows inconclusive results, as Cordero (2014), García-Pérez et al. (2014) o Warren et al. (2014) indicate that first-generation immigrants (born abroad) have a greater probability of repeating a grade than natives, while, for example, in the study by Choi et al. (2018) this variable does not prove to be significant.

Revista Española de Pedagogía
year 82, n. 289, September-December 2024, 539-562

FIGURE 1. Modular structure of contextual variables in PISA 2018.

	Student background constructs		**Schooling constructs**			**Non-cognitive and metacognitive constructs**
Reading literacy		5. Out-of scool scientific experience	TEACHING AND LEARNING			4. Reading related outcomes: attitudes, motivation and strategies
			1. Teacher qualifications and knowledge	2. Teaching practices for science	11. Learning time and curriculum	
			SCHOOL POLICIES			
				3. School-level learning environment for reading		
General categories	6. Socio-economic status of student and family 7. Migration and culture	8. Educational pathways in early childhood	13. Implicación de los padres	12. School climate: interpersonal relationships, trust and expectations	14. School context and resources	9. Dispositional and school-focused variables
			GOVERNANCE			
				16. Student assessment, institutional assessment and accountability	15. Allocation, selection and choice	10. Dispositions for global competence

Source: OECD (2019).

Secondly, among the school-related constructs, studies have mainly been conducted on learning time, school climate or the curriculum, which are associated with grade repetition (OECD, 2019). In this way, Asensio et al. (2018) and López et al. (2018) demonstrate the importance of the number of classes per week or the duration of classes for grade repetition, while Seabra and Ferrão (2016) find a significant relationship between grade repetition and disciplinary problems in schools, with the disciplinary climate having a greater impact than the socio-economic level. On the other hand, Arroyo et al. (2019) link repetition with curriculum-related variables, such as having studied science the previous year or having chosen an optional subject in this field.

In the group of non-cognitive and metacognitive constructs (hereinafter NC+MCC), motivation, self-confidence, future academic expectations and a sense of belonging to the group are highly significant variables (Fernández-Lasarte et al., 2019; Hornstra et al., 2017; Van Canegem et al., 2021), so that positive development of these aspects minimises the effect of grade repetition (Marsh et al., 2018). Likewise, academic self-concept and focusing on goals provide a high level of predictive capacity regarding the phenomenon under study (Ferla et al., 2009; Rodríguez-Rodríguez, 2022).

Despite the amount of existing research, one limitation that makes it difficult to define a more accurate picture lies in the fact that predictors are usually introduced into models in a random way, perhaps due to the lack of theoretical references that, in education, would produce modelling that aims to minimise specification error. Although the main assumption of multilevel models (which are the most widely used in secondary exploitation of PISA data, due to their isomorphism with the reality they aim to model) is that they are well defined, that is to say, that they meet the condition of "not excluding major predictors from the model" (Gaviria & Castro, 2005, p. 86).

In view of this, it is interesting to explore predictive models in which statistical error reduction is matched by the attempt to minimise specification error as well. The present research takes this line, as do others that use data mining, based on machine learning (Urbina & De la Calleja, 2017), both in international studies (Gamazo & Martínez-Abad, 2020; Kılıç et al., 2017; Kiray et al., 2015; Liu & Ruiz, 2008; Martínez-Abad & Chaparro, 2017; Martínez-Abad et al., 2020), and those focusing on Spanish samples (Arroyo et al., 2024a; Arroyo et al., 2024b; Asensio et al., 2018).

Another limitation is that the previous research does not tend to distinguish between students from the two compulsory stages, with most studies concentrating on secondary school students, which prevents the appearance of differentially significant variables in primary education.

In order to contribute to overcoming the aforementioned limitations, the overall objective pursued by this research is to obtain pertinent information for modelling that is a closer fit for grade repetition in the two compulsory stages in Spain, based on the PISA 2018 databases (Ministerio de Educación y Formación Profesional, 2019), which provide a vast amount of information about contextual variables (OECD, 2019).

The specific goals are:

- To identify the most accurate machine learning algorithm [individual decision trees (CART and C5.0), random forest and stochastic gradient boosting (GBM)] and, with it, select the variables that are most closely linked to grade repetition in the two stages and sort them according to their differential significance in each stage.

- To determine the relative contribution of each factor (student background, school-related and NC+MCC constructs) to the probability of a student repeating a grade in primary and secondary education taking into account the two-level structure of the data (students and schools).

2. Method

A secondary analysis was conducted of the PISA 2018 international assessment data, using a non-experimental, quantitative, cross-disciplinary, correlational design, in accordance with the PISA manual on technical guidelines (OECD, 2019).

2.1. Participants

Using two-stage sampling, the sample of Spanish students who participated in PISA 2018 is 35943, which represents coverage of 92% of the national population of 15-year-olds (Ministerio de Educación y Formación Profesional, 2019), composed of 49.96% women and with an average age of 15.84 years old (*SD* = 0.29). By type of school, 65% are students who are enrolled in public centres, 28% in state-contracted, privately-owned centres and 7% in private schools. The full information can be found in the second annex of the document published by the OECD (2020).

2.2. Instruments

All the variables have been measured with the PISA 2018 background questionnaires for students and principals (OECD, 2019). The dependent variables are grade repetition in primary and secondary education, with almost double the rate in the latter stage as reported by the participants themselves (Table1).

In terms of the independent variables, 85 were introduced in the initial stage (see Tables S1 and S2 at https://acortar.link/hfDnvO). Of this set of predictors, 83 are composite indices (WLE and SUM), of which 67 indices were constructed based on student responses (level 1) and 17 based on the school principals' responses (level 2). The two remaining variables are simple items that have been included in the model due to the large amount of previously existing evidence: *student gender* (level 1) and *percentage of 15-year-old students from socio-economically disadvantaged homes* (level 2).

Table 1. Grade repetition in compulsory education in Spain.

Have you ever repeated a grade in primary education?				
	n	%	*n* (weighted)	% (weighted)
No	30 078	89.927	340 287	88.606
Yes, once or more	3369	10.073	43 756	11.394
Total valid responses	33 447			
Have you ever repeated a grade in secondary education?				
	n	%	*n* (weighted)	% (weighted)
No	27 902	80.354	310 497	77.313
Yes, once or more	6822	19.646	91 114	22.687
Total valid responses	34 724			

Source: author's own compilation based on the PISA (2018) data for Spain.

Subsequently, the multilevel models included only those that, using the more accurate machine learning algorithm, show a greater connection with grade repetition in each of the stages studied.

2.3. Procedure

A detailed summary of the procedure for applying the questionnaires can be found in the Spanish report (Ministerio de Educación y Formación Profesional, 2019).

2.4. Data analysis

There were two separate stages, so that first an exploratory, descriptive study was conducted (Rosenthal & Rosnow, 2008) using algorithms supervised by machine learning and, once the most significant variables had been selected and sorted, several predictive models were tested taking into account the hierarchical data structure (level 1: student and level 2: school).

Prior to the machine learning algorithm, the data were pre-processed, as recommended by Raschka (2015), by processing the missing values with the technique of multiple imputation by chained equations (Rivero, 2011) and proceeding to random division of the cases into two sets: one for training (70%), with which to adjust the model, and the other for validation (30%), with which to evaluate the model by means of the area under the curve (AUC), as these are two dependent variables with unbalanced levels. Centring and scaling were also performed on the continuous predictors in order to avoid overrepresentation of the variables with a greater magnitude, and no predictor with zero or near-zero variance was obtained.

To meet the first objective, we compared the performance of four algorithms [individual decision trees (CART and C5.0), random forest and stochastic gradient boosting] using the AUC, with the grid-search method based on 10-fold cross-validation, jointly optimising the most significant hyperparameters. Using the sorting method of relative importance, Stochastic Gradient Boosting proved to be the most accurate algorithm, in accordance with the Friedman test applied to the training sample. For the sake of parsimony, given that one of the limitations of this algorithm is that it does not suggest a fixed cut-off point regarding the selection of an optimal number of predictors (Sarkar et al., 2018), in line with Gorostiaga and Rojo-Álvarez (2016), we decided to examine the following sets of variables: 15, 20, 25, 30 and 35. In this way, greater precision was provided for the selection of the optimal set, which proved to be, in both primary and secondary education, the one containing 30 of all the indices initially included.

To achieve the second objective, we used the technique of multilevel binary logistic regression, as the random variance of the null model proved to be statistically significant and the intraclass correlation coefficient (ICC) was greater than 10% (Lee, 2000). In considering the assumption of multicollinearity, we eliminated *home possessions* (HOMEPOS), *mother's education* (MISCED) and *parents' highest occupational status* (HISEI), as they are implicit in the *index of economic, social and cultural status* (ESCS); and the predictor *family wealth*, due to its strong correlation with this index.

Following this, as required by multilevel analysis, we removed the schools with fewer than 20 subjects, thereby excluding 1532 subjects belonging to 133 centres, so the final sample consisted of 34 411 students.

This generated a total of five models: the null model, without predictors; model 1, composed of the student background variables; model 2, comprising the school-related variables; model 3, with the NC+MCC variables; and model 4, containing all the predictors. Thus, it was possible to analyse the contribution of the school-related and NC+MCC factors once the student background had been checked. The variables were introduced into the model in the order obtained from the machine learning algorithm (Arroyo et al., 2019; Constante-Amores et al., 2021). Finally, to check the fit of the models, we used the AIC and BIC indices and Deviance, a statistical reduction that allows nested models to be compared by estimating the percentage of reduced variance (R2), and we calculated their significance (Cameron & Windmejier, 1997). The interpretation of the results is based on the percentage of explained variance (PEV), which quantifies the explanatory capacity of the final model, and on the odds ratios, in which values greater than 1 indicate that the variable increases the probability of repeating a grade, acting as a risk factor, and values below 1 suggest that this probability decreases, acting as a protective factor.

In terms of software, we used the version 4.4.0 R packages caret, stochastic gradient boosting and lme4 to compare the machine learning algorithms, select the most important variables and conduct multilevel binary logistic regression, respectively.

3. Results

Of the algorithms tested, the most accurate were stochastic gradient boosting and random forest, for both primary education (Figure 2) and secondary education (Figure 3). With the Friedman test, of note is stochastic gradient boosting, which we eventually used to select and hierarchise the most significant variables associated with grade repetition in compulsory education.

FIGURE 2. Comparison of machine learning algorithms for primary education.

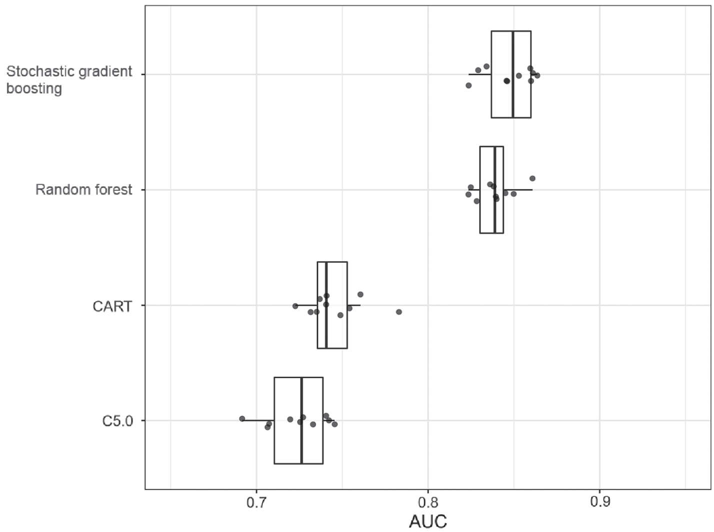

FIGURE 3. Comparison of machine learning algorithms for secondary education.

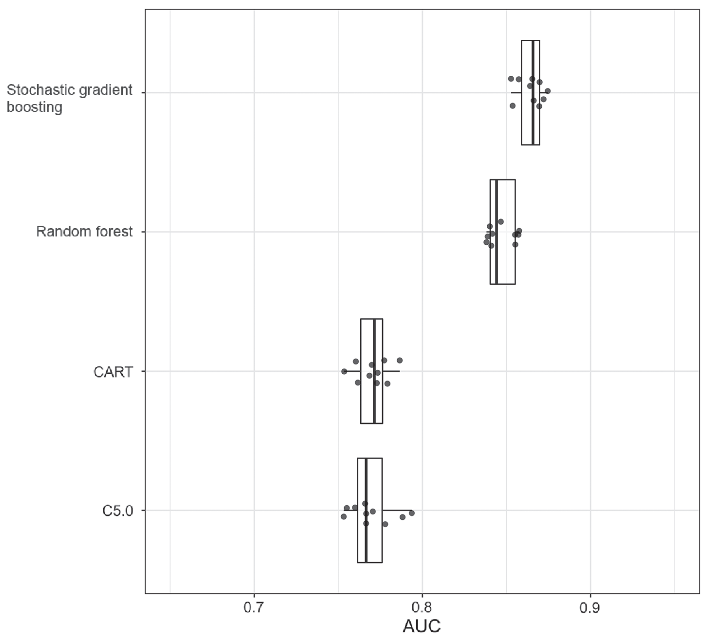

Table 2 shows the empirical results produced by stochastic gradient boosting, in which, out of the 30 most significant variables in primary education, 12 relate to background, 9 are associated with the school and 9 are NC+MCC. All of them correspond to level 1, except *the percentage of students from socio-economically disadvantaged homes*, which is from level 2, ranked 23rd. In this educational stage, the most significant predictors are the following: first, the *index of economic, social and cultural status* (ESCS); secondly, the *student's expected occupational status* (BSMJ); and thirdly, *learning time for mathematics* (MMINS).

Table 2. Most significant variables associated with grade repetition in primary education.

	Label	Order	Variables
Student background	ESCS	1st	Index of economic, social and cultural status
	HOMEPOS	4th	Home possessions
	SCCHANGE	5th	Number of school changes
	HISEI	8th	Parents' highest occupational status
	MISCED	9th	Mother's education
	IMMIG	12th	Immigration background
	ICTRES	16th	Home ICT resources
	HEDRES	17th	Home educational resources
	DURECEC	20th	Duration of early childhood education and care
	SC048Q03NA	23rd	Percentage of students from socio-economically disadvantaged homes
	BMMJ1	27th	Mother's ISEI
	WEALTH	30th	Family wealth
Related to school	MMINS	3rd	Learning time for mathematics (minutes per week)
	LMINS	6th	Learning time for language (minutes per week)
	INFOCAR	10th	Information about careers
	SMINS	11th	Learning time for sciences (minutes per week)
	DIRINS	13th	Teacher-directed instruction
	TMINS	14th	Learning time in total (minutes per week)
	PERFEED	18th	Perceived teacher feedback
	DISCLIMA	28th	Disciplinary climate in language classrooms (core classes in this PISA publication)
	BEINGBULLIED	29th	Experience of being bullied

Non-cognitive and metacognitive	BSMJ	2nd	Student's expected occupational status
	PISADIFF	7th	Perception of difficulty of the PISA test
	METASUM	15th	Metacognition: summarising
	STUBMI	19th	Student's body mass index
	GCSELFEFF	21st	Self-efficacy regarding global competence
	BELONG	22nd	Subjective well-being: sense of belonging at school
	SCREADDIFF	24th	Self-concept of reading: perception of difficulty
	GCAWARE	25th	Student awareness of global issues
	COGFLEX	26th	Student's cognitive flexibility/adaptability

Table 3 presents the variables that emerge as being the most important, with Stochastic Gradient Boosting, in secondary education: 8 relating to student background, 14 associated with school and 8 non-cognitive and metacognitive related to the student. In this stage, there are 29 from level 1 and only one from level 2: *student behaviour hindering learning* (ranked 11th). The most significant predictors here are revealed as: the *student's expected occupational status* (BSMJ), *learning time for sciences* (SMINS) and the index of economic, social and cultural status (ESCS). The descriptive analyses of the 30 most important variables shown in Table 2 and Table 3 can be found in the following link: https://acortar.link/hfDnvO (Tables S3 and S4).

TABLE 3. Most significant variables associated with grade repetition in secondary education.

Area	Label	Order	Variables
Student background	ESCS	3rd	Index of economic, social and cultural status
	CHANGE	5th	Number of changes in the educational biography
	HOMEPOS	9th	Home possessions
	BMMJ1	10th	Mother's ISEI
	HISEI	12th	Parents' highest occupational status
	HEDRES	20th	Home educational resources
	ICTRES	21st	Home ICT resources
	DURECEC	23rd	Duration of early childhood education and care

Related to school	SMINS	2nd	Learning time for sciences (minutes per week)
	LMINS	4th	Learning time for Language (minutes per week)
	MMINS	6th	Learning time for Mathematics (minutes per week)
	INFOCAR	7th	Information about careers
	STUBEHA	11th	Student behaviour hindering learning
	HOMESCH	13th	Use of ICT at home for school work
	TMINS	15th	Learning time in total (minutes per week)
	ICTOUTSIDE	17th	Subject-related use of ICT outside of school
	ENTUSE	18th	Use of ICT outside of school for leisure
	INFOJOB2	19th	Information about the labour market provided outside of school
	PERFEED	22nd	Perceived teacher feedback
	INFOJOB1	24th	Information about the labour market provided by the school
	DISCRIM	27th	Discriminatory school climate
	DIRINS	29th	Teacher-directed instruction
Non-cognitive and metacognitive	BSMJ	1st	Student's expected occupational status
	MASTGOAL	8th	Learning goals
	PISADIFF	14th	Perception of difficulty of the PISA test
	GCSELFEFF	16th	Self-efficacy regarding global competence
	RESILIENCE	25th	Student resilience
	METASUM	26th	Metacognition: summarising
	STUBMI	28th	Student's body mass index
	GCAWARE	30th	Student awareness of global issues

To achieve the second objective, once we had identified and sorted the variables with the strongest relationship to grade repetition, the next step was to define the predictive model, for which we dichotomised the categorical independent variables as indicated by Pardo and Ruiz (2013) (Table 4). For the predictor DURECEC, the OECD (2019) guidelines were followed.

Table 4. Recoding of independent variables for the multilevel logistic regression model.

Label	Categorical independent variable	Recoded values
SCCHANGE	Number of school changes	0 = No change 1 = Once or more changes
CHANGE	Number of changes in the educational biography	
IMMIG	Immigration background	0 = Immigrant 1 = Native
DURECEC	Duration of early childhood education and care	0 = Three or under 1 = Four or more

Table 5 summarises the regression models for grade repetition in primary education, with N1 or N2 after the variable name to identify the level to which it belongs. In model 1, all the student backgrounds are statistically significant, which explains the 62.380% variability in grade repetition. According to the odds ratio, the predictor *number of school changes* (SCCHANGE) presents the greatest probability of repetition. It should be noted that the *percentage of students from socioeconomically disadvantaged homes* (SC048Q03NA) is the only significant variable from level 2, and *mother's ISEI* (BMMJ1) has an odds ratio that is close to 1. Model 2, composed of the school-related variables, explains 19% of the criterion variance and, although they are all significant, four of them have an odds ratio of 1, with *experience of being bullied* (BEINGBULLIED) and *teacher-directed instruction* (DIRINS) being the main risk factors. Model 3 explains about 25% of the variance, containing only NC+MCC variables, of which *student awareness of global issues* (GCAWARE) is not significant and has an odds ratio that is very close to the *body mass index* (STUBMI), with *perception of difficulty of the PISA test* (PISADIFF) being the variable with the greatest effect. The last model (M4) explains 61% of the variance in grade repetition in primary education; there is no significance attached to the effects of *learning time for sciences (minutes per week)* (SMINS), or the *body mass index* (STUBMI), and furthermore, there are eight variables with an odds ratio that is very close to 1; the variables with the most effect are *number of school changes* (SCCHANGE), *teacher-directed instruction* (DIRINS) and *perceived teacher feedback* (PERFEED).

Lastly, as Table 5 shows, regarding the model fit, model 4 has the lowest AIC and BIC score, with a significant reduction in the variance, equivalent to an R^2 of 9% compared to model 1. This is also significant in relation to the null model, with a reduction in the variance of model 1 (8%), model 2 (5%) and model 3 (9%).

Revista Española de Pedagogía
year 82, n. 289, September-December 2024, 539-562

TABLE 5. Odds ratio and standard errors for the multilevel logistic regression models in primary education.

	M0 (Null)	M1 (Backgrounds)	M2 (School)	M3 (NC+MCC)	M4 (All)
Fixed effects					
Intercept	0.10(0.03)***	0.14(0.08)***	0.07(0.09)***	0.41(0.11)***	0.38(0.17)***
ESCS_N1		0.59(0.03)***			0.67(0.03)***
SCCHANGE_N1		2.10(0.04)***			1.89(0.04)***
IMMIG_N1		0.62(0.05)***			0.58(0.05)***
ICTRES_N1		0.90(0.03)***			0.90(0.03)***
HEDRES_N1		0.89(0.02)***			0.93(0.03)***
DURECEC_N1		0.69(0.05)***			0.77(0.05)***
SC048Q03NA_N2		1.01(0.00)***			1.01(0.00)***
BMMJ1_N1		0.99(0.00)***			1.00(0.00)***
MMINS_N1			1.00(0.00)***		1.00(0.00)***
LMINS_N1			1.00(0.00)***		1.00(0.00)***
INFOCAR_N1			0.72(0.02)***		0.79(0.02)***
SMINS_N1			1.00(0.00)***		1.00(0.00)***
DIRINS_N1			1.30(0.03)***		1.24(0.02)***
TMINS_N1			1.00(0.00)***		1.00(0.00)***
PERFEED_N1			1.25(0.03)***		1.22(0.02)***
DISCLIMA_N1			0.83(0.02)***		0.89(0.02)***

BEINGBULLIED_ N1			1.30(0.02)***		1.13(0.02)***
BSMJ_N1				0.97(0.00)***	0.98(0.00)***
PISADIFF_N1				1.31(0.02)***	1.19(0.02)***
METASUM_N1				0.69(0.02)***	0.75(0.02)***
STUBMI_N1				1.01(0.00)***	1.00(0.00)***
GCSELFEFF_N1				0.87(0.02)***	0.89(0.02)***
BELONG_N1				0.87(0.02)***	0.93(0.02)***
SCREADDIFF_N1				1.89(0.02)***	1.08(0.02)***
GCAWARE_N1				0.99(0.02)	1.01(0.03)***
COGFLEX_N1				1.08(0.02)***	1.06(0.02)***
Random effects					
Variance	0.52(0.05)***	0.20(0.03)***	0.42(0.04)***	0.39(0.04)***	0.20(0.07)***
PEV (%)		62.38	18.62	25.53	0.61
Fit indices					
AIC	22198.69	19804.01	21053.05	20120.57	18072.87
BIC	22215.58	19888.48	21145.96	20213.48	18309.368
Loglik	-11097.34	-9892.01	-10515.53	-10049.29	-9008.43
Deviance	20724.84	19026.15	19749.16	18906.55	17292.51
Chisq		2410.72	1163.65	2096.13	1767.1
Pr (> Chisq)		0.00	0.00	0.00	0.00

Note: $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$. The reference category to interpret the odds ratio is the value 1 in all cases.

Lastly, Table 6 presents the regression models for grade repetition in secondary education, with N1 or N2 after the variable name to identify the level to which it belongs. Model 1 explains approximately 51% of the criterion and it is not significantly affected by *home ICT resources* (ICTRES), with the main risk factor being the *number of changes in educational biography* (CHANGE). Model 2 explains 32% of the variance; all the variables are statistically significant, although four of them have an odds ratio that is very close to 1. The greatest effect corresponds to *student behaviour hindering learning* (STUBEHA), a level 2 variable. Model 3, containing NC+MCC predictors, explains 14%. *Student resilience* (RESILIENCE) is not significant and there are two other predictors with an odds ratio that is very close to 1, with the greatest risk factor found in the variable *perception of difficulty of the PISA test* (PISADIFF). In the global model, *learning time for sciences* (SMINS), *learning time for mathematics* (MMINS) and *resilience* (RESILIENCE) are not significant and there are another six predictors with an odds ratio that is very close to 1. This set of predictors explains almost 56% of grade repetition in secondary education. The variables with the greatest effect are *number of school changes* (SCCHANGE), *student behaviour hindering learning* (STUBEHA) and *teacher-directed instruction* (DIRINS).

Lastly, it should be noted that model 4 provides the best fit, with a significant reduction in the variance, equivalent to an R^2 of 10.298% compared to model 1, although the reduction in model 1 is also significant in comparison to model 0 (7%), and in models 2 and 3 compared to the null model (5% and 10%, respectively).

TABLE 6. Odds ratio and standard errors for the multilevel logistic regression models in secondary education.

	M0 (Null)	**M1 (Backgrounds)**	**M2 (School)**	**M3 (NC+MCC)**	**M4 (All)**
Fixed effects					
Intercept	0.21(0.03)***	0.27(0.06)***	0.14(0.07)***	1.00(0.10)	0.70(0.13)***
ESCS_N1		0.67(0.02)***			0.76(0.02)***
CHANGE_N1		2.44(0.03)***			2.24(0.03)***
BMMJ1_N1		0.98(0.01)***			0.99(0.00)***
HEDRES_N1		0.85(0.02)***			0.92(0.02)***
ICTRES_N1		0.91(0.02)			0.91(0.02)***
DURECEC_N1		0.69(0.05)***			0.76(0.04)***
SMINS_N1			1.00(0.00)***		1.00(0.00)***
LMINS_N1			1.01(0.00)***		1.00(0.00)***

Revista Española de Pedagogía
year 82, n. 289, September-December 2024, 539-562

MMINS_N1			1.00(0.00)***		1.00(0.00)
INFOCAR_N1			0.72(0.02)***		0.76(0.02)***
STUBEHA_N2			1.42(0.03)***		1.25(0.02)***
HOMESCH_N1			0.91(0.02)***		0.93(0.02)***
TMINS_N1			1.00(0.00)***		1(0.00)***
ICTOUTSIDE_N1			0.89(0.02)***		0.96(0.02)***
ENTUSE_N1			1.10(0.02)***		1.13(0.02)***
INFOJOB2_N1			1.13(0.02)***		1.11(0.02)***
PERFEED_N1			1.16(0.02)***		1.15(0.02)***
INFOJOB1_N1			0.84(0.02)***		0.85(0.02)***
DISCRIM_N1			1.39(0.02)***		1.20(0.02)***
DIRINS_N1			1.21(0.02)***		1.21(0.02)***
BSMJ_N1				0.97(0.00)***	0.98(0.00)***
MASTGOAL_N1				0.78(0.02)***	0.82(0.02)***
PISADIFF_N1				1.23(0.02)***	1.14(0.02)***
GCSELFEFF_N1				0.86(0.02)***	0.88(0.02)***
RESILIENCE_N1				1.00(0.02)	0.99(0.02)
METASUM_N1				0.72(0.02)***	0.79(0.02)***
STUBMI_N1				1.01(0.00)***	1.01(0.00)***
GCAWARE_N1				1.08(0.02)***	0.76(0.02)***
Random effects					
Variance	0.70(0.04)***	0.34(0.05)***	0.47(0.04)***	0.60(0.05)***	0.32(0.04)***
PEV (%)		51	32.43	14	55.69
Fit indices					
AIC	32841,56	29736,77	30809,14	29611,70	26708,20
BIC	32858.46	29804.34	30944.28	29696.16	26961.58
Loglik	-16418.78	-14860.38	-15388.57	-14795.85	-13324.10
Deviance	30681.45	28238.71	29013.64	27662.13	25330.76
Chisq		3116.8	2060.3	152.57	6215.6
Pr (> Chisq)	0.70(0.04)***	0.34(0.05)***	0.47(0.04)***	0.60(0.05)***	0.32(0.04)***

Note: *p <.05, **p <.01, ***p <.001. The reference category to interpret the odds ratio is the value 1 in all cases.

4. Discussion

In the modelling produced, the most significant contextual indices have proved to be mainly from the first level, that is to say, student characteristics predominate over school characteristics. This result coincides with that obtained by Lopez et al. (2023) in their systematic review of research conducted with Spanish students. Also with other secondary analyses in which the object of study is performance, as in the study by Choi et al. (2018) with data on Spanish students from PISA 2012 and PIRLS 2006 (Progress in International Reading Literacy Study), or by Lopes et al. (2022) with data on Portuguese students from PIRLS 2016.

It can be concluded that the three variables that are most closely linked to grade repetition throughout compulsory education are *social status*, from the student background group; *learning time*, from the group of constructs associated with school; and the *level of expectation*, from the NC+MCC set, in a different order in the two stages. In this last group, the results are in line with those found by Asensio et al. (2018) and López et al. (2018), who, without differentiating the stages, establish expectations as one of the most important variables as regards performance. In terms of the importance of the student background variable, the present study is in line with the findings of Carabaña (2013), Choi et al. (2018) or García-Pérez et al. (2014). However, here we provide specific information relating to the differential weight depending on the stage, which may be interpreted as an indicator that the educational system is fulfilling its compensatory function, as ESCS moves into third place in secondary education, behind a psychological variable and one related to the curriculum. Without a doubt, this evidence represents an interesting contribution by the present study. In the two stages, other significant psychoeducational elements emerge such as, from the set of school-related variables, *academic/professional information* (ranked tenth in primary education and seventh in secondary education), *teacher-directed instruction* (thirteenth place in primary education and twenty-ninth in secondary education), *perceived teacher feedback* and the *climate in the classroom*; and, from the NC+MCC group, *perception of the difficulty of the test*.

Specifically of note, in primary education only, is the relationship between grade repetition with *bullying* and *flexibility* or *cognitive adaptability* and in secondary education, with *use of ICT*, *learning goals* and *awareness of global issues*. Thus, this stage shows an increase in the number of school-related variables considered important, with a decrease in those related to background, on which the school has less power to act directly, thereby confirming and expanding the findings by Arroyo et al. (2019) with data from PISA 2015.

The emergence of differential models in both stages leads to the conclusion that how grade repetition is treated at the practical and political level needs to take this aspect into account. The existence of clear links between background variables and grade repetition also has implications for practice, as they are a warning of the need to persevere with compensatory approaches that improve equity, particularly in primary education. Furthermore, the importance of the time spent on learning indicates practices aimed at efficient management of this variable, without being restricted to simply increasing study time in disadvantaged classes, which is a problem raised, for example, by the policies of summer courses to recover "educational debts" (Battistin & Schizzerotto, 2019). The most equitable solution would be to support the principles of individualisation, early intervention (Choi et al., 2018) and inclusion (Tapia & Álvarez, 2022), in which teacher training and collaborative teaching or team teaching are portrayed as promising strategies in the most successful educational systems (Niemi, 2015). In terms of the level of expectation, it is important to work on motivation, at both the primary and secondary stages (Rhodes et al., 2018; Rodríguez-Rodríguez, 2022). Other specific aspects that should be developed in order to lessen the rate of grade repetition in both stages are feedback on learning, the classroom climate or academic/professional guidance.

The availability of such an extensive database as PISA has enabled the use of machine learning algorithms in this work, which is one of its main contributions. The multilevel study conducted was based on a prior selection that helped to minimise specification error, which is difficult to avoid in educational research. The use of machine learning techniques complements and increases the sensitivity of hierarchical regression models, by providing empirical criteria to maintain the important predictors and dismiss the ones that are irrelevant. Although it has not proved to be so efficient as regards the latter function due to the existence of significant odds ratios, albeit equal or very close to 1, this is a methodology that allows for a quantitative calculation not only of each predictor's worth but also, fundamentally, the value of the model's predictive validity.

In terms of limitations, it should be noted that machine learning methods may overfit models to the current sample's characteristics. Additionally, the PISA background questionnaires are based on self-reporting, which is not free of measurement errors. Moreover, it is worth mentioning that the relationships identified do not imply causality, so for this reason the results of this type of exploratory study should be complemented by confirmatory research posing hypotheses on the direct and indirect effects that apply to the factors involved. In this sense, any educational measure that is adopted to reduce grade repetition, given the relationship of this variable with performance, will have an indirect effect on it, although this aspect should be studied specifically using covariance structure modelling, or designs that are experimental or quasi-experimental. For the future, it would also be advisable to examine the replicability of the results in depth, by comparing the results in participant subgroups, for example, by gender. Likewise, it would be interesting to explore the factors associated with grade repetition by extending the hierarchy levels to include teachers (three-level model) and even autonomous regions (four-level model).

Authors' contributions

Alexander Constante-Amores: Data curation; Formal analysis; Methodology; Writing (original draft).

Delia Arroyo-Resino: Conceptualisation; Formal analysis; Writing (original draft); Writing (review and editing).

María Sánchez-Munilla: Data curation; Formal analysis; Methodology; Writing (original draft).

Inmaculada Asensio-Muñoz: Conceptualisation; Supervision; Writing (original draft); Writing (review and editing).

References

Allen, C. S., Chen, Q., Willson, V. L., & Hughes J. N. (2009). Quality of design moderates effects of grade retention on achievement: A metaanalytic, multi-level analysis. *Educational Evaluation and Policy Analysis*, *31* (4), 480-99. https://doi.org/10.3102/0162373709352239

Arroyo, D., Constante-Amores, A., & Asensio, I. (2019). La repetición de curso a debate: un estudio empírico a partir de PISA 2015 [Debate on grade repetition: An empirical study from PISA 2015]. *Educación XX1, 22* (2), 69-92. https://doi.org/10.5944/educxx1.22479

Arroyo, D., Constante-Amores, A., Castro, M., & Navarro, E. (2024a). Eficacia escolar y alto rendimiento del alumnado español en PISA 2018: un enfoque de *machine learning* [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

Arroyo, D., Constante-Amores, A., Gil, P., & Carrillo, P. J. (2024b). Student well-being and mathematical literacy performance in PISA 2018: A machine-learning approach. *Educational Psychology*, *44* (3), 340-357. https://doi.org/10.1080/01443410.2024.2359104

Asensio, I., Carpintero, E., Expósito, E., & López, E. (2018). ¿Cuánto oro hay entre la arena? Minería de datos con los resultados de España en PISA 2015 [How much gold is in the sand? Data mining with Spain's PISA 2015 results]. **Revista Española de Pedagogía**, *76* (270), 225-245. https://doi.org/10.22550/REP76-2-2018-02

Battistin, E., & Schizzerotto, A. (2019). Threat of grade retention, remedial education and student achievement: Evidence from upper secondary schools in Italy. *Empirical Economics*, *56*, 651-678. https://doi.org/10.1007/s00181-018-1443-6

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. https://doi.org/10.1016/S0304-4076(96)01818-0

Carabaña, J. (2013). Repetición de curso y puntuaciones PISA, ¿cuál causa cuál? [Grade repetition and PISA scores, which causes which?] In Instituto Nacional de Evaluación Educativa (Ed.), *PISA 2012: Programa para la evaluación de los alumnos, informe español (volumen II: análisis secundario) [PISA 2012: Programme for international student assessment, Spanish report (volume II: secondary analysis)]*. INEE.

Choi, A., Gil, M., Mediavilla, M., & Valbuena, J. (2018). Predictors and effects of grade repetition. *Revista de Economía Mundial*, (48), 21-42.

Constante, A., Florenciano, E., Navarro, E., & Fernández, M. (2021). Factores asociados al abandono universitario [Factors associated with university dropout]. *Educación XX1, 24* (1), 17-44. https://doi.org/10.5944/educxx1.26889

Cordero, J., Manchón, C., & Simancas, R. (2014). La repetición de curso y sus factores condicionantes en España [Repetition and explanatory factors in Spain]. *Revista de Educación*, (365), 12-37. https://doi.org/10.4438/1988-592X-RE-2014-365-263

Ferla, J., Valcke, M., & Cai, Y. (2009). Academic self-efficacy and academic self- concept: Reconsidering structural relationships. *Learning and Individual Differences, 19* (4), 499-505. https://doi.org/10.1016/j.lindif.2009.05.004

Fernández-Lasarte, O., Ramos-Díaz, E., & Axpe, I. (2019). Rendimiento académico, apoyo social percibido e inteligencia emocional en la universidad [Academic performance, perceived social support and emotional intelligence at the university]. *European Journal of Investigation in Health*, *9* (1), 39-49.

Gamazo, A., & 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

García-Pérez, J., Hidalgo, M., & Robles, J. (2014). Does grade retention affect students' achievement? Some evidence from Spain. *Applied Economics*, *46* (12), 1373-1392. https://doi.org/10.1080/00036846.2013.872761

Gaviria, J. L., & Castro, M. (2005). *Modelos jerárquicos lineales [Linear hierarchical models].* La Muralla.

González-Nuevo, C., Postigo, Á., García-Cueto, E. Menéndez-Aller, Á., Muñiz, J., Cuesta, M., Álvarez-Díaz, M., & Fernández-Alonso, R. (2023). Grade retention impact on academic self-concept: a longitudinal perspective. *School Mental Health, 15*, 600-610. https://doi.org/10.1007/s12310-023-09573-2

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. https://doi.org/10.1016/j.neucom.2015.07.001

Hattie, J. (2017). *Aprendizaje visible para profesores. Maximizando el impacto en el aprendizaje [Visible learning for teachers. Maximising learning impact]*. Paraninfo.

Hornstra, L., Majoor, M., & Peetsma, T. (2017). Achievement goal profiles and developments in effort and achievement in upper elementary school. *British Journal of Educational Psychology, 87* (4), 606-629. https://doi.org/10.1111/bjep.12167

Ikeda, M., & García, E. (2014). Grade repetition: A comparative study of academic and non-academic consequences. *OECD Journal: Economic Studies*, (1), 269-315. https://doi.org/10.1787/eco_studies-2013-5k3w65mx3hnx

Jimerson, S. R. (2001). Meta-analysis of grade retention research: new directions for practice in the 21st century. *School Psychology Review*, *30* (3), 420-437. http://dx.doi.org/10.1080/02796015.2001.12086124

Jimerson, S. R., & Ferguson, P. (2007). A longitudinal study of grade retention: Academic and behavioral outcomes of retained students through adolescence. *School psychology quarterly*, *22* (3), 314.

Kılıç, D., Aşkın, Ö. E., & Öz, E. (2017). Identifying the classification performances of educational data mining methods: A case study for TIMSS. *Kuram ve Uygulamada Egitim Bilimleri*, *17* (5), 1605-1623. https://jestp.com/menuscript/index.php/estp/article/view/426

Kiray, S. A., Gok, B., & Bozkir, A. S. (2015). Identifying the factors affecting science and mathematics achievement using data mining methods. *Journal of Education in Science, Environment and Health*, *1* (1), 28-48.

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

Liu, X., & Ruiz, M. E. (2008). Using data mining to predict K-12 students' performance on large-scale assessment items related to energy. *Journal of Research in Science Teaching*, *45* (5), 554-573. https://doi.org/10.1002/tea.20232

Lopes, J., Oliveira, C. & Costa, P. (2022). Determinantes escolares y de los estudiantes en el rendimiento lector: un análisis multinivel con estudiantes portugueses. *Revista de Psicodidáctica*, *27*, 29-37. http://doi.org/10.1016/j.psicod.2021.05.001

López, E., Expósito, E., Carpintero, E., & Asensio, I. (2018). ¿Qué nos dice PISA sobre la enseñanza y el aprendizaje de las ciencias? Una aproximación a través de árboles de decisión [What does PISA tell us about science teaching and learning? A decision tree approach]. *Revista de Educación*, (382), 133-161.

López, F., & García, I. (2020). *A vueltas con la equidad en educación: una aproximación empírica en la perspectiva de las consecuencias [The equity in education: An empirical approach in the perspective of consequences]*. Universidad Camilo José Cela. http://hdl.handle.net/20.500.12020/901

López, L., González-Rodríguez, D., & Vieira, M. J. (2023). Variables que afectan la repetición en la educación obligatoria en España [Variables affecting repetition in compulsory education in Spain]. *Revista Electrónica de Investigación Educativa*, *25*, 1-15. https://doi.org/10.24320/redie.2023.25.e17.4942

López-Rupérez, F., García-García, I., & Expósito-Casas, E. (2021). La repetición de curso y la graduación en Educación Secundaria Obligatoria en España. Análisis empíricos y recomendaciones políticas [Grade repetition and graduation in compulsory secondary education in Spain. Empirical analysis and policy recommendations]. *Revista de Educación*, (394), 325-353. https://doi.org/10.4438/1988-592X-RE-2021-394-510

Marsh, H. W., Parker, P. D., & Pekrun, R. (2018). Three paradoxical effects on academic self-concept across countries, schools, and students. *European Psychologist*, *24* (3), 231-242. http://doi.org/10.1027/1016-9040/a000332

Mathys, C., Véronneau, M., & Lecocq, A. (2019). Grade retention at the transition to secondary school: Using propensity score matching to identify consequences on psychosocial adjustment. *Journal of Early Adolescence*, *39* (1), 97-133. https://doi.org/10.1177/0272431617735651

Martínez-Abad, F., & Chaparro, A. A. (2017). Data-mining techniques in detecting factors linked to academic achievement. *School Effectiveness and School Improvement*, *28* (1), 39-55. https://doi.org/10.1080/09243453.2016.1235591

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*, 100875. https://doi.org/10.1016/j.stueduc.2020.100875

Nieto-Isidro, S., & Martínez-Abad, F. (2023). Repetición de curso y su relación con variables socioeconómicas y educativas en España [Grade retention and its relationship with socioeconomic and educative variables in Spain]. *Revista de Educación*, *1* (402), 207-236. https://doi.org/10.4438/1988-592X-RE-2023-402-600

Ministerio de Educación y Formación Profesional. (2019). *PISA 2018. Programa para la evaluación internacional de los estudiantes. Informe español (vol. I) [PISA 2018. Programme for international student assessment. Spanish report (vol. I)].* https://www.educacionyfp.gob.es/inee/evaluaciones-internacionales/pisa/pisa-2018/pisa-2018-informes-es.html

Ministerio de Educación y Formación Profesional. (2022). *Sistema estatal de indicadores de la educación 2022 [State system of education indicators 2022].* https://www.libreria.educacion.gob.es/libro/sistema-estatal-de-indicadores-de-la-educacion-2022_184171/

Niemi, H. (2015). Teacher professional development in Finland: Towards a more holistic approach. *Psychology, Society and Education*, *7* (3), 279-294.

OCDE (2019). *PISA 2018 assessment and analytical framework*. https://doi.org/10.1787/b25efab8-en

OCDE (2020). *PISA 2018 results (volume VI). Are students ready to thrive in an interconnected world?* https://www.oecd.org/publications/pisa-2018-results-volume-vi-d5f68679-en.htm

Pardo, M., & Ruiz, M. (2013). *Análisis de datos en ciencias sociales y de la salud III [Data analysis in health and social sciences III]*. Síntesis.

Peixoto, F., Monteiro, V., Mata, L., Sanches, C., Pipa, J., & Navas, L. (2016). *To be or not to be retained… That's the question!* Retention, self-esteem, self-concept, achievement goals, and grades. *Frontiers in Psychology*, 7, (1550). https://doi.org/10.3389/fpsyg.2016.01550

Raschka, S. (2015). *Python machine learning*. Packt.

Rhodes, J., Thomas, J. M., & Liles, A. R. (2018). Predictors of grade retention among children in an elementary school truancy intervention. *Journal of At-Risk Issues*, *21* (1), 1-10.

Rivero, G. (2011*). Análisis de datos incompletos en ciencias sociales [Incomplete data analysis in social sciences]*. CIS.

Rodríguez-Rodríguez, D. (2022). Repetición de curso, rendimiento académico y variables motivacionales en Educación Secundaria Obligatoria: un estudio longitudinal [Grade retention, academic performance and motivational variables in compulsory secondary education: A longitudinal study]. *Psicothema*, *34* (3), 429-436. https://doi.org/10.7334/psicothema2021.582

Rosenthal, R., & Rosnow, R. L. (2008). *Essentials of behavioral research: Methods and data analysis*. McGraw-Hill

Sarkar, D., Bali, R., & Sharma, T. (2018). *Practical machine learning with Python. A problem-solvers guide to building real-world intelligent systems*. Apress.

Schwerdt, G., West, M. R., & Winters, M. A. (2017). The effects of testbased retention on student outcomes over time: Regression discontinuity evidence from Florida. *Journal of Public Economics*, *152*, 154-169. https://doi.org/10.1016/j.jpubeco.2017.06.004

Seabra, D. A., & Ferrão, M. E. (2016). Repetência e indisciplina: Evidências de Brasil e Portugal no PISA 2012 [Grade repetition and indiscipline: Evidence from Brazil and Portugal in Pisa, 2012]. *Cadernos de Pesquisa*, *46* (161), 614-636. https://publicacoes.fcc.org.br/cp/article/view/3669

Tapia, J. G., & Álvarez, C. A. (2022). La repetición de curso en educación secundaria en clave organizacional [Grade repetition in secondary education in an organizational perspective]. *Avances en Supervisión Educativa*, (38). https://doi.org/10.23824/ase.v0i38.772

Urbina, A. B., & De la Calleja, J. (2017). Breve revisión de aplicaciones educativas utilizando minería de datos y aprendizaje automático [Brief review of educational applications using data mining and machine learning]. *Revista Electrónica de Investigación Educativa*, *19* (4), 84-96. https://doi.org/10.24320/redie.2017.19.4.1305

Valbuena, J., Mediavilla, M., Choi, Á., & Gil, M. (2021). Effects of grade retention policies: A literature review of empirical studies applying causal inference. *Journal of Economic Surveys*, *35* (2), 408-451. https://doi.org/10.1111/joes.12406

Van Canegem, T., Van Houtte, M., & Demanet, J. (2021). Grade retention and academic self-concept: A multilevel analysis of the effects of schools' retention composition. *British Educational Research Journal*, *47* (5), 1340-1360. https://doi.org/10.1002/berj.3729

Warren, J. R., Hoffman, E., & Andrew, M. (2014). Patterns and trends in grade retention rates in the United States, 1995-2010. *Educational Researcher*, *43* (9), 433-443. https://dx.doi.org/10.3102/0013189X14563599

Authors' biographies

Alexander Constante-Amore. Bachelor's Degree in Pedagogy and Master's Degree in Educational Research from the Universidad Complutense de Madrid (UCM). He is currently completing the Doctoral Programme in Education at the UCM. He is a lecturer at the Universidad Camilo José Cela and the Universidad Europea de Madrid, where he teaches subjects in the areas of Statistics, Biostatistics and Research Methods. He was an intern at the Instituto Nacional de Estadística [Spanish National Statistics Institute] and he is a member of the research group Medida y Evaluación de Sistemas Educativos (MESE) [Measurement and Evaluation of Educational Systems]. His line of research focuses on the evaluation of educational systems.

https://orcid.org/0000-0003-2024-9615

Delia Arroyo-Resino. Lecturer at the Universidad Complutense de Madrid. She holds an International Doctorate with an Outstanding Award from the Department of Research and Diagnostic Methods in Education for the Universidad Complutense de Madrid. She is a member of the research group Medida y Evaluación de Sistemas Educativos (MESE).

https://orcid.org/0000-0002-3784-7745

María Sánchez-Munilla. Bachelor's Degree in Pedagogy and a Master's Degree in Methodologies in Behavioural Health Sciences at the Universidad Nacional de Educación a Distancia (UNED). She is currently completing the UCM Doctoral Programme in Education on a predoctoral contract named Formación del Profesorado Universitario (FPU) [University Teacher Training]. She is a member of the research group Medida y Evaluación

de Sistemas Educativos (MESE) and the Servicio de Evaluación y Diagnóstico en Educación (SEDE) [Educational Evaluation and Diagnosis Service] of the Faculty of Education at the UCM

https://orcid.org/0000-0001-6689-7170

Inmaculada Asensio-Muñoz. Doctorate with an outstanding award from the Universidad Complutense de Madrid. Senior lecturer in the area of Research and Diagnostic Methods in Education, at the Department of Research and Psychology in Education in the Faculty of Education at this University. She has extensive teaching experience in pedagogy and teacher training, both at undergraduate and master's or doctorate level. She is a member of the research group Medida y Evaluación de Sistemas Educativos (MESE) and, as an expert in educational research methodology, during her professional career, she has worked on several funded projects on research and innovation related to this subject. Her publications focus on improving education and learning, and, in general, on the role of teaching and guidance.

https://orcid.org/0000-0003-3523-570X