An economic evaluation of educational interventions in the LOMLOE*: Proposals for improvement with Artificial Intelligence

Evaluación económica de intervenciones educativas en la LOMLOE: Propuestas de mejora con Inteligencia Artificial

María Teresa BALLESTAR, PhD. Associate Professor. Universidad Rey Juan Carlos *(teresa.ballestar@urjc.es).*
Jorge SAINZ, PhD. Professor. Universidad Rey Juan Carlos *(jorge.sainz@urjc.es).*
Ismael SANZ, PhD. Associate Professor. Universidad Rey Juan Carlos *(ismael.sanz@urjc.es).*

Abstract:

This research aims to demonstrate the need for an economic evaluation of the Organic Law that modifies the Organic Law of Education (LOMLOE), especially after the investment of EU Next Generation funds that open new opportunities that were lacking in the initial drafting of the law. The challenge for Public Administrations is to use this additional investment efficiently.

Our analysis shows that artificial intelligence models can predict whether educational support programmes will help increase the likelihood that students who lag behind will pass the 4th grade of ESO (Compulsory Secondary Education). In this way, we can calculate the social return of one of these programmes and contribute to their *ex-ante* design to achieve higher success rates for students.

To complement the models already used by Public Administrations, we use robust Machine Learning (ML) models such as CHAID decision trees and artificial neural networks to analyse the characteristics of the groups of students and the intervention they have been part of. The conclusions allow us to improve educational reinforcement programmes in the coming years to

revista española de pedagogía
year 80, n. 281, January-April 2022, 133-154

* Organic Law 3/2020, of 29 December, which amends Organic Law 2/2006, of 3 May, on Education (LOMLOE).

Revision accepted: 2021-12-13.
This is the English version of an article originally printed in Spanish in issue 281 of the **revista española de pedagogía**. For this reason, the abbreviation EV has been added to the page numbers. Please, cite this article as follows: Ballestar, M. T., Sainz, J., & Sanz, I. (2022). Evaluación económica de intervenciones educativas en la LOMLOE: propuestas de mejora con inteligencia artificial |*An economic evaluation of educational interventions in the LOMLOE: Proposals for improvement with artificial intelligence. Revista Española de Pedagogía, 80* (281), 133-154. https://doi.org/10.22550/REP80-1-2022-09
https://revistadepedagogia.org/ ISSN: 0034-9461 (Impreso), 2174-0909 (Online)

support students with lower chances of academic success.

Keywords: Public policy analysis, Machine Learning, educational efficiency, LOMLOE.

Resumen:

El objetivo de esta investigación es demostrar la necesidad de evaluar económicamente la LOMLOE (Ley Orgánica 3/2020, de 29 de diciembre de 2020, por la que se modifica la Ley Orgánica 2/2006, de 3 de mayo, de Educación), especialmente tras la inversión de los fondos EU Next Generation que abren nuevas oportunidades de las que carecía la ley en su redacción inicial. Las Administraciones públicas tienen el reto de emplear esa inversión adicional de forma eficiente.

Nuestro análisis demuestra que los modelos de inteligencia artificial pueden predecir si los programas de apoyo educativo ayudarán a incrementar la probabilidad de que estudiantes rezagados superen 4.º de la ESO. De esta forma, se puede calcular el retorno social de los programas de apoyo educativo y contribuir a su diseño *ex-ante* para lograr que los alumnos tengan mayores tasas de éxito.

Para completar los modelos ya utilizados por Administraciones públicas, empleamos modelos de Machine Learning (ML) robustos como árboles de decisión CHAID y redes neuronales artificiales para analizar las características de los grupos de estudiantes y la intervención en la que han formado parte. Las conclusiones permiten mejorar los programas de refuerzo educativo de los próximos años para apoyar a los alumnos con menos posibilidades de éxito académico.

Descriptores: análisis de políticas públicas, *Machine Learning*, eficiencia educativa, LOMLOE.

1. Introduction

One of the principles of educational management by international organisations such as the OECD (Organisation for Economic Co-operation and Development) is the evaluation of educational policies, both global and of specific interventions. (Golden, 2020). The result of the analysis leads to greater effectiveness and impact, both at the social and individual levels, of educational actions. This in turn has a positive impact on students and a positive dynamic for the system in general. (OECD, 2018). This informative principle appears in almost all European law, and, in the case of Spain, is included in Organic Law 2/2006 of 3 May (LOE), which states in Article 2 bis that "The functioning of the Spanish Education System is governed by the principles of [...] efficiency in the allocation of public resources, transparency and accountability".

Our aim is to demonstrate that the economic evaluation of education policies serves to optimise and prioritise the legislative objectives set by the LOMLOE. This fact would demonstrate that the law faces a gap in its construction that affects its management evaluation, an evaluation that is a requirement for the justification of *EU Next Generation* funds for education, which in 2021 alone represent 1,852.5 million euros in the general state budget.

It is this European investment that will make it possible to implement and evaluate the law, since in its Analysis and Regulatory Impact Report (known as MAIN), which includes the technical aspects of the regulation, it is stated that "...the project does not involve an increase or decrease in public spending". On the other hand, the Commission requires an assessment of whether the population effectively acquires the skills to compete at a global level. (Crescenzi et al., 2021; Porte & Jensen, 2021).

The European Commission and UNESCO recognise the importance of this type of action, both in the choice of reforms and in their subsequent analysis, and the need for data on educational interventions to be available for said analysis, despite the reluctance of some administrations. European Comission (EC, 2010; Yusuf, 2007). Although the preferred approach for this analysis is the experimental method, in the reality of education there are problems in carrying it out for various reasons: such as the economic cost, the difficulty of establishing randomised control groups, problems of management, anxiety, etc. (Golden, 2020; Slavin, 2016).

To solve this proposition, there is an arsenal of new techniques that are being used and which are in full development and can serve as a proposition, including artificial intelligence (AI) (Ballestar et al., 2019; Chassignol et al., 2018; Chatterjee & Bhattacharjee, 2020) because of the difficulty of identifying who are the main beneficiaries and what are the long-term effects. Still, new policies including financial incentives have been adopted to increase the research output at all possible levels. Little literature has been devoted to the response to those incentives. To bridge this gap, we carry out our analysis with data of a six years program developed in Madrid (Spain. Following the triangulation strategy of Ballestar et al., (2020) we evaluate the use of different AI strategies for the evaluation of the return on educational interventions. To do so, we will delve into the intervention carried out by the Regional Ministry of Education of the Junta de Comunidades de Castilla y León called *Programa para la Mejora del Éxito Educativo (Programme for the Improvement of Educational Achievement)*, which we will analyse in its 2020 edition, after the closure of schools due to the pandemic. The Junta made a call for financial aid so that schools could offer support classrooms during July to students in the sixth year of primary school, the fourth year of ESO and the second year of baccalaureate in the areas of Spanish language and literature, mathematics, or English with educational difficulties. The programme has contributed over the years to the promotion of a significant number of students in the community. All grades benefited from the programme, yielding an improvement in promotion of 5% in the 2019-20 academic year for students who participated in the programme, compared to those who were able to participate but did not.

2. The intervention

To design any educational intervention Slavin (2016) y Golden (2020) recognise as fundamental that it should be based on scientific evaluations. The support classes that make up the *Programa para la Mejora del Éxito Educativo en* Castilla y León are probably one of the most scientifically grounded in recent years. Research conducted by the *Education Endowment Foundation*[1] shows that small group tutoring is among the measures for which there is empirical evidence of greatest effectiveness internationally. As such, they have proven to be a good complement to an education system that is designed to move large numbers of students from grade to grade, but which does not work for all students.

Kraft (2015) and Burgess (2020) discuss the factors contributing to its success: delivery by selected and trained staff in coordination with their regular teachers and in small groups. The impact is significant: Nickow et al., (2020) show, from the study of 96 randomised experiment items, that the effect of small-group tutoring is large and significant (37% of the standard deviation).

This intervention is one of the few that are included in the Spanish legislation. LOMLOE, Art. 4.4, states as one objective to facilitate "...access to the support that students require" and the increase of these measures throughout the entire educational process in an individualised manner (Artart. 20 bis) "... to avoid school repetition, particularly in socially disadvantaged environments". In fact, the regulation leaves a very clear wording on its objectives to repetition, for example, the new Artart. 28 reads: "Remaining in the same year [of secondary school] shall be considered an exceptional measure and shall be taken after having exhausted the ordinary reinforcement and support measures to solve learning difficulties".

These support measures have been present in educational legislation throughout this century. Thus, the unborn Organic Law 10/2002, of 23 December, on the Law on the Quality of Education (LOCE) included in its art. 2 the right of students to "...receive the necessary help and support to compensate for personal, family, economic, social, and cultural deficiencies, and disadvantages, especially in the case of special educational needs, which prevent or hinder access to and permanence in the education system". The various regulations since then have taken up similar precepts of protection in line with the recommendations of specialised bodies (Gouëdard et al., 2020; Pont & Montt, 2014; Schleicher, 2020) or the literature on educational return (Brunello & Paola, 2014; de la Fuente & Jimeno, 2009; Doncel et al., 2014) which establish the importance of the acquisition of competences, a fact which has become fundamental with the irruption of new technologies (Ballestar et al., 2020, 2022; Goos et al., 2009; Gregory et al., 2019).

Art. 9 of the LOMLOE gives the responsibility to the autonomous communities (CCAA) to reduce and prevent school failure and early school leaving through territorial cooperation programmes. The Law's Analysis and Regulatory Impact Report (MAIN) in the first draft of the law, prior to the pandemic, provides 45 million euros per year for this type of programme for the period 2020-2023.

The autonomous regions were already investing heavily in this type of project. Since the 2007-2008 academic year, the Regional Ministry of Education of Castilla y León (Castile and Leon) has been developing, among others, the *Programme for the Improvement of Educational Achievement*. Its objectives are, in line with the LOE, LOMCE (Organic Law for the Improvement of the Quality of Education, 2013) and LOMLOE "... to contribute to the improvement of the educational model in the Community of Castilla y León and to facilitate the success and continuity of students in the education system" guaranteeing "the didactic progression of students and their promotion in the education system, reducing early school leaving, promoting their effective integration into the labour market and at the same time optimising the climate of coexistence in schools".

It is the schools that request these measures to favour educational success, the integration of students and the involvement of families to increase the promotion and graduation rate of students, with special attention to the most vulnerable students in the 4th year of ESO, 6th year of Primary Education and 2nd year of Baccalaureate in publicly funded schools. The programme was is taught in 93 centres in language, mathematics, and English to reinforce the end of the cycle and with the participation of students attending voluntarily and supported by their families. The access requirements are to require educational support or to be in a situation of socio-educational vulnerability and it is taught by specialised teachers.

For the 2019-2020 academic year, students participating in the programme increase their probability of passing language by 24.6%, English by 13.8% and mathematics by 8.4% compared to those not enrolled, generating a high level of satisfaction among both teachers (85.7%) and families (88.2%).

This first analysis serves as the basis for our research on the economic efficiency of the interventions included in the LOMLOE. To this end, we propose to classify students participating in the Programme for the Improvement of Educational Success in Castile and Leon in the 4th year of ESO into groups and to predict the probability of passing this school year for each of the groups of students. This will allow individual predictions to be made for students based on their characteristics, details of the special support programme carried out and their performance in the programme. This allows the probability of success of the programme to be estimated and its social return (ROI) calculated through a triangulation model that employs different ML methodologies such as CHAID decision trees and artificial neural networks to evaluate educational policies.

3. Empirical analysis.

The information used for the analysis contains the data available for students in the 4th year of ESO participating in the Programme for the Improvement of Educational Success in the 2019-2020 academic year. There are 1,739 records that correspond to the students who participated in one of the three programmes implemented in the Autonomous Community. Of these, 47.27% are girls and 52.73% are boys. Of these students, 47.15% (820 students) have already repeated a grade.

Students can participate in three categories of programmes. The C2 consists of accompanying students throughout the academic year and accounts for 76.37% of students (1,328). The C3 consists of extra classes during the summer in July and includes 17.02% of the students (296). The C2C3 programme, a combination of the two previous programmes, represents 6.61% of the students (115). These actions are delivered in the nine provinces of Castilla y León, with 66.36% of the students concentrated in provincial capitals and 33.64% in other localities, with a success rate in the completion of 4th-year ESO studies of 85.34% for those who finished the programme.

One of the innovations of our research is the application of a triangulation methodology, which consists of the development of more than one quantitative method, applying different approaches, with the aim of enriching the results, as well as confirming the results obtained twice. (Ballestar et al., 2020). In the first phase, we develop a Machine Learning (ML) model based on CHAID (Chi-square Automatic Interaction Detector) decision trees to determine which variables are relevant when designing and predicting the expected success rate of participants in the programme. The aim is to identify the factors to be considered when designing an intervention of these characteristics, as well as to calculate its success rate and, consequently, the social return on the investment made.

In the second phase, we assess the robustness of the first model by designing a new model using an ML methodology based on artificial neural networks multilayer perceptron backward programming (ANN-MPL) to validate the results obtained previously.

3.1. Definition of variables in ML models.

CHAID decision trees and multilayer perceptron artificial neural networks (ANN-MLP) are supervised machine learning methods. These models describe and explain the underlying relationships between different input variables in order to predict the value of the target variable, through a training process using a data sample that contains both the values for the input and output variables. (Maimon & Rokach, 2005). As supervised methods, the accuracy of their classifications and predictions will depend on the quality of the sample available for training, validation and testing of the models. (Aad et al., 2012; Ballestar et al., 2018; Li & Eastman, 2006).

We performed a data mining process on the database to create a single table for the development of the ML models. This table contains 1,739 student records and 21 variables that capture the information for the characterisation of the student, as well as their performance and results both in the programme they have participated in and in the 4th year of ESO.

Through descriptive and iterative causal analyses, five of these variables were statistically significant and relevant for the empirical analysis and development of the two ML models. Of these variables, four act as input variables in the models and one as an output or target variable to be determined or predicted (Table 1). Therefore, both models use the same five variables and are trained on the same database.

The target or output variable, the percentage of students who did not pass the 4th year of ESO (0 in the variable finish_studies), is 14.66%, while those who do pass the year represent 85.34% (value 1 of the variable finish_studies), leaving an unbalanced sample.

Table 1. Description of the variables of the ML models: CHAID decision tree and multilayer perceptron artificial neural network (ANN-MLP).

Input variables	Description
years_repetition	Discrete numeric variable. Number of years the student has repeated. If the student has not repeated any year, the value will be 0.
student_repetition	Boolean variable. Value 1 if the student has repeated a course; Value 0 if the student has not repeated a course.
kind_program	Categorical variable indicating which of the three academic programmes the student has completed: C2; C3; C2C3.
finish_program	Boolean variable. Value 1 if the student has completed the academic support programme; Value 0 if the student has not completed it.

Output variable	Description
finish_studies	Boolean variable. Value 1 if the student has passed the 4th year of ESO; Value 0 if the student has not passed the course.

Source: Own elaboration.

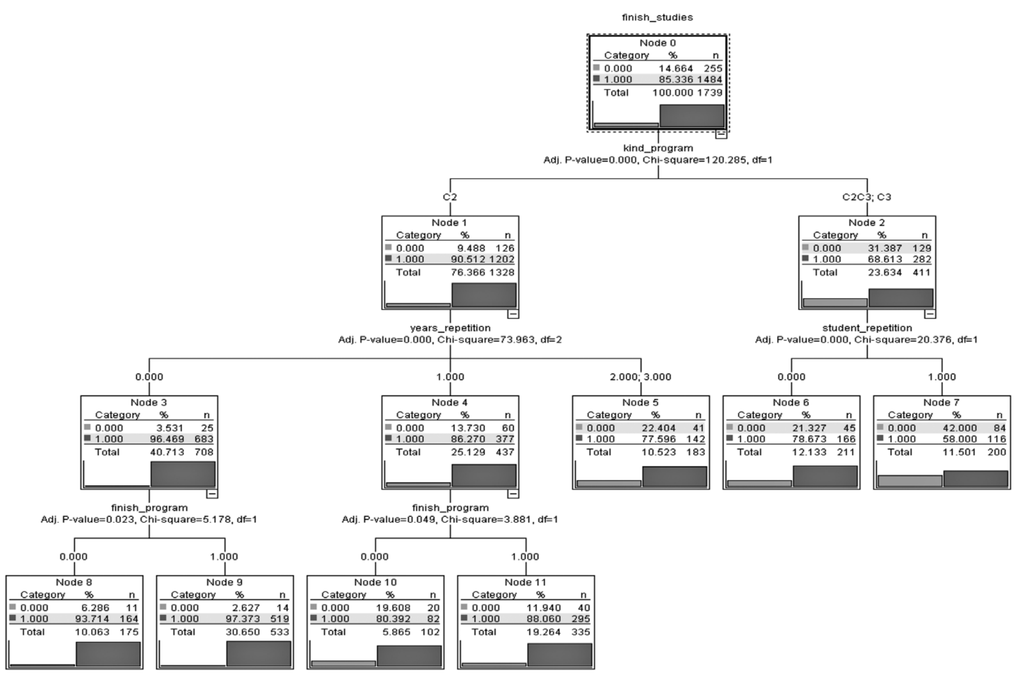
3.2. First phase of the empirical analysis: Predictive model of the success of the special education support programme.

In this phase, we developed a model consisting of a CHAID or *chi-squared automatic interaction detection (*CHAID) decision tree (Kass, 1980)CHAID, is and offshoot of AID (Automatic Interaction Detection whose objective is twofold: classification and prediction. This model will classify each of the students into homogeneous groups based on the explanatory input variables or predictors that interact significantly with the dependent output or target variable (Akin et al., 2017; Khosravi et al., 2019). The algorithm identifies the students who are likely to belong to a particular group and, at the same time, identifies the rules that it will use in predicting future cases of those students who have not participated. In our case, these groupings are constituted according to the characteristics of the student, the programme in which he/she participates and his/her performance, making it possible to fulfil the second objective of determining and predicting the pass rate in the 4th year of ESO for each of the groups and individually (Ramaswami & Bhaskaran, 2010). Finally, the prediction of this success rate makes it possible to calculate how to reinforce each of the programmes by type of student.

One of the benefits provided by CHAID decision trees over other methods is that they could generate non-binary trees. In addition, they also support continuous or categorical numeric type input variables. Being non-parametric, it can handle both linear and non-linear relationships between the explanatory variables and the output variable and can also handle very large volumes of data very efficiently, even in real time, (Chassignol et al., 2018; Khosravi et al., 2019).

For the development and training of the model, we applied a training, testing and validation (TTV) methodology which consists of training the model with 70% (1,203 records) of the sample and testing and validating it with the remaining 20% (367 records) and 10% (169 records) of the sample, respectively. (Ballestar et al., 2019)because of the difficulty of identifying who are the main beneficiaries and

Graph 1. CHAID decision tree diagram for the success model of the student support programmes for the successful completion of the 4th year of ESO.



Source: Own elaboration.

what are the long-term effects. Still, new policies including financial incentives have been adopted to increase the research output at all possible levels. Little literature has been devoted to the response to those incentives. To bridge this gap, we carry out our analysis with data of a six years program developed in Madrid (Spain). Graph 1 shows the CHAID decision tree diagram obtained for the programme success model.

3.2.1. First phase of the empirical analysis: Evaluation of the model.

The most relevant CHAID indicators are classification accuracy, sensitivity, specificity, area under the ROC curve and the GINI coefficient. These accuracy indicators and the confusion matrix, which contains the percentage of cases classified both correctly and incorrectly for the two possible values of the dependent variable, are available in Table 2. The indicators have been calculated both for the total sample and for each of the training, testing and validation subsamples of the model. Therefore, we observe that the model has an accuracy in classifying students and predicting their percentage of passing students of 70.73% (error of 29.27%). This percentage of correctly classified students is very similar between all the training, testing and validation subsamples, so we can affirm that the model has not been overtrained.

The percentage of true positives, called *sensitivity*, is 71.43%. This value explains the percentage of students who pass the course and who have been correctly classified by the CHAID decision tree based on the characteristics of the student, the support programme carried out and their result in this programme. The percentage of true negatives, also called *specificity*, is 66.67%. This value is the percentage of students who do not pass the 4th year of ESO and who have been correctly classified based on the same input variables. The complementary values are the percentage of false positives, 33.33%, which corresponds to the percentage of students who did not pass the 4th year of ESO and were classified by the model as having achieved it. Finally, the percentage of false negatives is 28.57%. This value corresponds to the percentage of students who, having passed the year, were classified by the model as not having passed it.

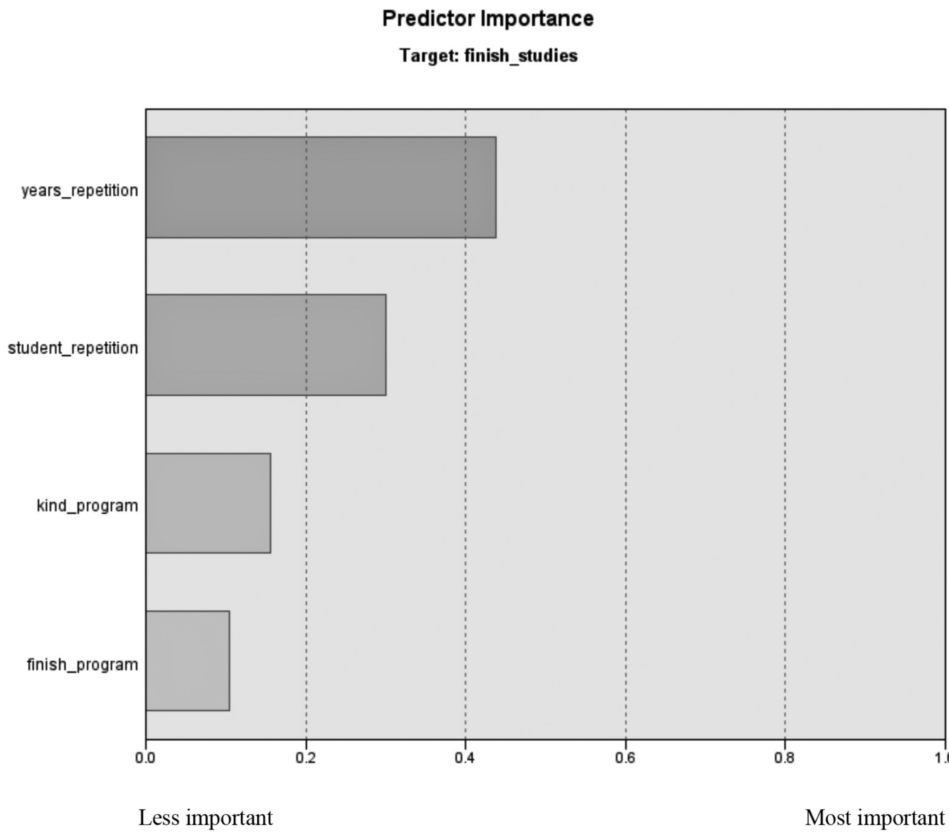
In this research we prefer to use as the main measure of accuracy of the CHAID decision tree the area under the curve (AUC) ROC indicator, as it is more robust than the classification accuracy indicator when working with unbalanced samples as in our case (Table 2). (Dželihodžić & Jonko, 2016; Yin et al., 2013). The area under the ROC curve (AUC) for the total sample has a value of 0.762, like those of the training, testing and model validation subsamples (0.763, 0.760 and 0.755 respectively). The AUC values can range from 0.5, which is the worst possible value and would imply that the model makes random classifications, to 1, which is the best value and would mean that the model makes perfect classifications. Therefore, it can be concluded that the quality of this CHAID decision tree model is good. (Hosmer Jr., et al., 2013). Complementarily, the GINI coefficient has also been calculat-

TABLE 2. Model accuracy and confusion matrix of the CHAID decision tree.

Model Accuracy					Confusion Matrix					
							Predicción			
							Sample size		Percentage	
Sample	Sample size	Percentage Correct	AUC	GINI	Sample	Observed	0	1	0	1
Total Sample	1739	70.73%	0.762	0.523	Total Sample	0	170	85	66.67%	33.33%
						1	424	1060	28.57%	71.43%
Sub-sample										
Training	1203	70.91%	0.763	0.527	Training	0	117	62	65.36%	34.64%
						1	288	736	28.13%	71.88%
Test	367	69.21%	0.760	0.519	Test	0	34	16	68.00%	32.00%
						1	97	220	30.60%	69.40%
Validation	169	72.78%	0.755	0.510	Validation	0	19	7	73.08%	26.92%
						1	39	104	27.27%	72.73%

Source: Own elaboration.

Graph 2. Relative importance of input variables in the CHAID decision tree.



Source: Own elaboration.

ed, which can be used alternatively to the AUC, as they are closely related. The GINI coefficient represents twice the area between the ROC curve and the diagonal and ranges between the values 0 and 1. In our research, the GINI coefficient for the total sample has a value of 0.523, like that of the training, testing and validation subsamples of the model (0.527, 0.519 and 0.510 respectively).

Graph 2 shows the relative importance of each of the significant predictor variables in the CHAID decision tree. The importance of these predictors is determined by calculating the reduction in variance of the target variable (completion or non-completion of the 4th year of ESO per student) that can be attributed to each predictor by performing a sensitivity analysis. (Saltelli et al., 2004). In this research, the variable that accumulates 44% of the relative importance is the number of years that the student has repeated (years_repetition), followed by the variable indicating whether the student has repeated or not (student_repetition) with an importance of 30%. Next, with

16%, is the type of programme in which the student participated (kind_program) and, finally, with 10%, whether the student completed the special support programme (finish_program).

3.2.2. First phase of the empirical analysis: Analysis of the results of the CHAID decision tree model.

The CHAID-based ML model produces a tree in which the first group, called the *root node*, represents the total set of the sample, in this case, the 1,739 students analysed. The algorithm divides this set into two or more categories called *parent or initial nodes*. Below the parent nodes, child nodes are linked to them. The categories at the last level of the decision tree are called *terminal nodes*. In terms of hierarchy, the *parent or initial nodes* exert the greatest influence on the *root node*, corresponding to the dependent variable it seeks to explain, while they exert a lesser influence on the *terminal nodes*.

The total number of nodes is 12, distributed as follows in three levels of depth: one root node (level 0), four parent nodes (distributed in levels 1 and 2) and seven terminal nodes (distributed in levels 2 and 3) (Graph 1). Additionally, based on the model evaluation analysis carried out in the previous section, it has been shown that this model has a good segmentation and prediction capacity (ROC of 0.762).

3.2.2.1. Level 0 of the CHAID Decision Tree.

At Level 0 of the CHAID decision tree, Node 0 is the variable to be predicted, which determines whether the student has successfully completed the 4th year of ESO (finish_studies). At this level it is observed that, for the total sample, 85.34% have passed the course, while 14.66% have not.

3.2.2.2. Level 1 of the CHAID Decision Tree.

At this level, the most relevant variable to define the first two parent nodes (Node 1 and Node 2) is the type of programme in which the student has participated (kind_program). Each of these nodes will give rise to a sub-tree that will use different predictor variables to define the following levels. At this level, the most relevant variable is the programme in which the student has participated, whether he/she has completed it or not. Of the students who have participated in the C2 programme (accompaniment of the student throughout the 4th year of ESO), 90.51% have passed the 4th year of ESO, compared to 68.61% of those who have participated in the C3 and C2C3 programmes. Therefore, the programme with the greatest success in its objectives is C2.

3.2.2.3. Level 2 of CHAID Decision Subtree 1 (Left).

The next most relevant variable having participated in the C2 programme is the number of times they have repeated a year (years_repetition). In Node 1 90.51% of the students pass, but this percentage can vary greatly depending on whether the student has repeated a year or not previously: students who have not repeated a year increase their pass rate to 96.46% (Node 3), while those who have repeated once see it reduced to 86.27% (Node 4) and those who have repeated two or more times to 77.59% (Node

5). Non-repeaters obtain better results in the 4th year of ESO compared to those who have repeated a year at least once.

In this Level 2 we find the nodes Node 3, 4 and 5. On the one hand, the nodes Node 3 and 4, in which the students have not repeated a grade or have done so only once, give rise to the new subtrees of Level 3. In this Level 3 the most discriminant variable is the one that determines whether the student managed to finish the C2 programme in which he/she participated (finish_program). On the other hand, Node 5 of Level 2 is a terminal node (it has no more sublevels), which implies that the probability of passing the 4th year of ESO for the students in this group is independent of their performance in the C2 programme, contrary to Node 3 and 4.

3.2.2.4. Level 3 of CHAID Decision Subtree 1 (Left).

In Level 3 there are 4 terminal nodes: Node 8 and Node 9 correspond to students who, having participated in C2, have not repeated previously, and come from the sub-tree of Node 3. The students in Node 9 have completed the C2 programme and therefore 97.37% have managed to pass the 4th year of ESO, the most successful figure of all the segments of the CHAID decision tree. While in Node 8 they have not managed to complete the C2 programme, and their success rate drops to 93.71%.

Node 10 and Node 11 correspond to students in C2 who have repeated a grade once and come from the sub-tree of Node 4. Students in Node 11 have completed the C2 programme and therefore 88.06% have managed to pass the 4th year of ESO, while those in Node 10 see this figure reduced to 80.39% because they have not completed the C2 programme. Therefore, students who complete the C2 programme obtain better results in the 4th year of ESO compared to those who have not completed the programme.

3.2.2.5. Level 2 of CHAID Decision Subtree 2 (Right).

In level 2 of subtree 2 (right, having been part of C2C3 or C3) is whether the student has ever repeated a year (student_repetition). We know that in Node 2 only 68.61% of the students pass, but this percentage can vary a lot depending on whether the student is a repeater or not. Students who have not repeated any year increase their pass rate in 4th ESO to 78.67% (Node 6), while those who have repeated a year see this percentage reduced to 58% (Node 7). Therefore, non-repeaters obtain better results in the 4th year of ESO compared to those who have repeated a grade. In this sub-tree 2 (right) there are no more levels, and this variable is the last relevant variable to generate more groups of students.

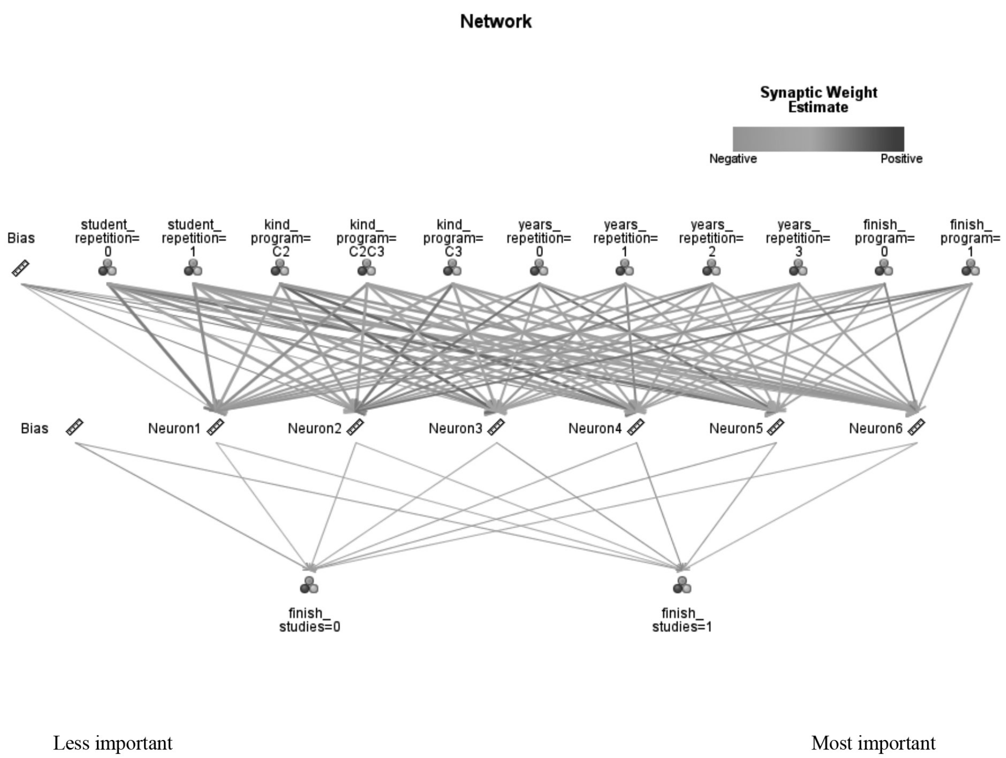
3.3. Second phase of the empirical analysis: Analysis of the robustness and predictive quality of the model.

In this research, a triangulation methodology is applied to test the robustness of the results obtained in the first phase. For this purpose, an ML predictive model based on a multilayer perceptron artificial neural network (ANN-MLP) is developed. Both models use the same database with the same input variables (student characterisation, performance and type of programme carried out) and output (student performance in the 4th year of ESO) (Table 1), but they are built using completely different ML methods, with the aim of verifying whether they obtain confirmatory results. (Wolszczak-Derlacz & Parteka, 2011).

ANN-MLP has the ability of neural networks to handle complex relationships between variables, both linear and non-linear, which makes them an alternative to more traditional predictive models such as logistic regressions where the variable to be predicted is dichotomous. (Paliwal & Kumar, 2009). This ensures that no relationships are left unidentified.

While training the ANN-MLP it should be considered whether the distribution of the dependent variable is unbalanced. This bias in our sample could have a negative impact on the ANN-MLP training process. To avoid this, we applied an *oversampling method* on the underrepresented group to train the model with 70% (2,098 records) of the sample already balanced, ensuring that both groups of students are equally represented when training the model.

Graph 3. ANN-MLP architecture for predicting the success of programmes to support students to pass the 4th year of ESO.



Source: Own elaboration.

On the other hand, the validation and testing of the ANN-MLP model is carried out respectively with 20% (367 records) and 10% (169 records) of the sample, this time, without balancing. We also follow a training, testing and validation (TTV) methodology.

Graph 3 shows the architecture used to predict the success of student support programmes for passing the 4th grade of ESO, consisting of three layers: the first input layer consists of eleven units, corresponding to the possible values of the four input variables related to the characteristics of the students, their performance and the programme carried out (Table 1). The middle layer, called the hidden layer, consists of six units or neurons. Finally, the last one consists of two units, corresponding to the prediction of student performance in 4th ESO (Table 1). The type of activation function of the hidden layer and the output layer correspond to a hyperbolic tangent and *softmax* respectively.

Table 3 describes the accuracy indicators and the confusion matrix of the ANN-MLP which show that the quality of this model is good: correct classification of 71.19% of the students and an AUC of 0.763. (Hosmer Jr., et al., 2013). Furthermore, it is also verified that there is no over-training as the indicators obtained for the total sample, as well as for the training, testing and validation subsamples are similar.

4. Discussion

The indicators we have used to assess the accuracy of the ANN-MLP as a predictor of student success are the same as those used for the CHAID, thus facilitating the comparison between results, predictive quality, and robustness of both models.

Table 4 compares the classification accuracy and confusion matrix indicators of both models (Table 3 vs. Table 2). The predictions and accuracy of the ANN-MLP are consistent with those of the CHAID decision tree. The ANN-MLP correctly predicts 71.19% (8.81% error) of the students who will pass the 4th year of ESO. This figure represents a 0.46% (8 students) higher predictive capacity compared to the 70.73% of correct predictions obtained by the CHAID decision tree. For the rest of the indicators, such as sensitivity, specificity, false positives and false negatives, the differences are also statistically non-significant, ranging from -1.57% (difference in the classification of four students) to 0.81% (difference in the classification of 12 students). The same is true for the differences between the areas under the ROC curve (AUC) and the GINI coefficient. The AUC for the ANN-MLP and the CHAID decision tree, 0.763 and 0.762 respectively and the GINI coefficient 0.525 and 0.523 respectively. Graph 4 shows the relative importance of the ANN-MLP input variables or predictors following the variance method and in total they add up to 1.

Although it is true that both the CHAID decision tree and the ANN-MLP use the same database and variables in the model and manage to achieve very similar results, predictive capacity and robustness, the way of doing so is very different, both in the methods used

TABLE 3. Model accuracy and confusion matrix of the artificial neural network multilayer perceptron (ANN-MLP).

Model Accuracy					Confusion Matrix					
							Prediction			
							Sample Size		Percentage	
Sample	Sample Size	Percentage Correct	AUC	GINI	Sample	Observed	0	1	0	1
Total Sample	1739	71.19%	0.763	0.525	**Total Muestra**	0	166	89	65.10%	34.90%
						1	412	1072	27.76%	72.24%
Submuestra										
Training	2098*	69.78%	0.767	0.535	**Training**	0	720	354	67.04%	32.96%
						1	280	744	27.34%	72.66%
Test	367	70.03%	0.755	0.510	**Test**	0	31	19	62.00%	38.00%
						1	91	226	28.71%	71.29%
Validation	169	69.23%	0.744	0,488	**Validation**	0	15	11	57.69%	42.31%
						1	41	102	28.67%	71.33%

*Oversampled to balance the sample.

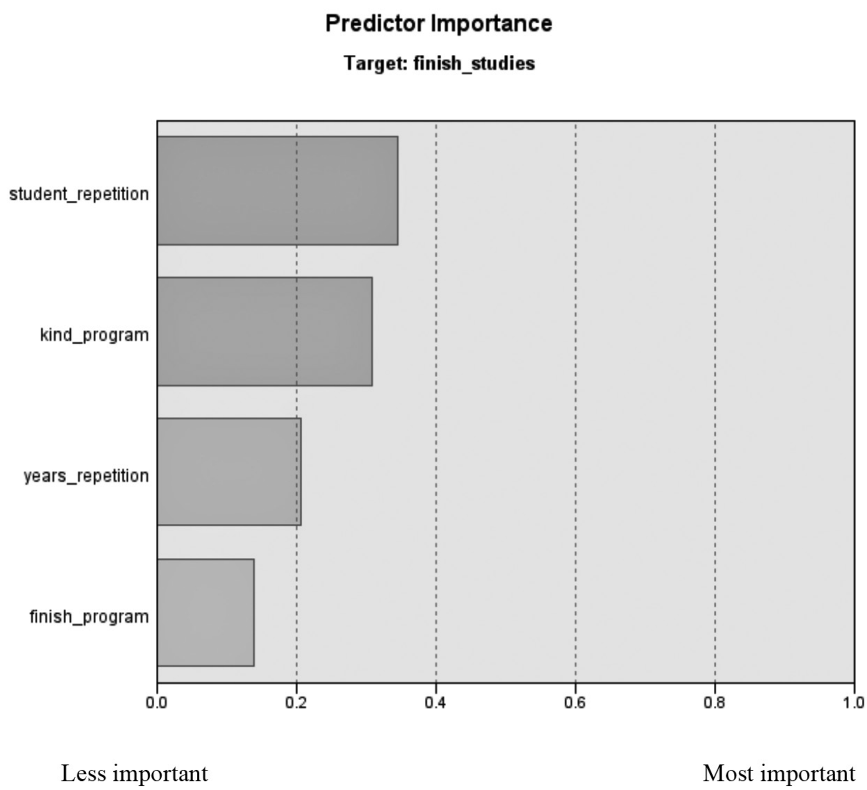
Source: Own elaboration.

Table 4. Comparison of indicators: Classification accuracy and confusion matrix of the ANN-MLP compared to the CHAID decision tree.

	ANN-MLP compared to CHAID Tree	
Indicators	**% change**	**Classified variation**
Percentage Correct	0.46%	8
Sensitivity	0.81%	12
Specificity	-1.57%	-4
False positives	1.57%	4
False negatives	-0.81%	-12
AUC	0.001	-
GINI	0.002	-

Source: Own elaboration.

Graph 4. Relative importance of the input variables of the artificial neural network multilayer perceptron (ANN-MLP).



Source: Own elaboration.

and, in the handling, and importance of the variables that contribute to the model. In the case of the ANN-MLP, the variable with the highest relative importance accumulates up to 35% and indicates whether the student has repeated a year (student_repetition), followed by the type of programme in which the student participated with an importance of 31% (kind_program), next with an importance of 21% is the variable indicating the number of years repeated by the student (years_repetition) and, finally, with an importance of 13% is the variable indicating whether the student finished the special support programme (finish_program).

The application of the triangulation methodology concludes that the CHAID decision tree models and the ANN-MLP obtain similar results in terms of classification accuracy and prediction of the success of the student support programmes for passing the 4th year of ESO and, therefore, they are robust models that confirm their results. Of the three programmes analysed, the one with the highest success rate is C2, which accompanies the student throughout the entire 4th year of ESO, reaching a 90.51% pass rate, compared to 68.61% for the other two options C2 (extraordinary classes in July) and C2C3 (combination of the previous options).

In terms of educational evaluation, the results are striking: the economic and social return of an intervention can be evaluated according to the different characteristics of the students, such as course performance, the type of course or, of course, whether they are repeaters or not.

5. Conclusions

Through our research we have contrasted the need for an economic evaluation of the LOMLOE. The law started with the same budget allocation as the LOMCE for interventions such as the one analysed (45 million euros). The arrival of European funds opens up new possibilities not contemplated in its initial design, which in turn raises the challenge of using this additional investment efficiently.

Our analysis shows that it is possible to predict the probability of at-risk students passing the 4th year of ESO depending on the support programme they participate in. This can contribute to the design of more efficient programmes with higher success rates.

To complement the models already used by Public Administrations, we employ robust ML models such as CHAID decision trees and artificial neural networks to analyse the characteristics of the students, the intervention they have taken part in and the final outcomes. This allows us to maximise the social return of each programme and to support those students who are less likely to succeed.

For example, our results show that, in the *Programme for the Improvement of Educational Success* in Castile and Leon, regardless of the programme in which the student has participated, those who have never repeated a grade are more likely to pass the 4th year of ESO than those who have. The probability of passing the year increases when students, in addition to participating in special programmes, stay for the full support program. For this reason, the group of students who achieve the best results is the group of non-repeaters who participate in and complete the C2 programme, with 97.37% passing the 4th year of ESO. While the group with the worst results is the group of repeaters who have participated in the C2C3 and C3 programmes, regardless of whether they have completed them or not, with a 58% pass rate in the 4th year of ESO. Our analysis also shows that providing educational support in July does not constitute a great added value since the C1 programme with accompaniment during the school year obtains better results than the C3 programme which combines this reinforcement during the school year with an additional reinforcement in the month of July. In the next phase of research, we will check whether these different results are due to the pupils themselves or to the characteristics of the programme.

As a corollary, we also show that ML models such as CHAID and artificial neural networks are good candidates for building models to answer the educational questions posed, such as the importance of mentoring the student to complete the programme and not drop out, especially in the case of young people who have repeated, for whom the programme is a very efficient tool to avoid school failure and an efficient way of investing public money.

Note

1 https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/the-impact-of-covid-19-on-school-starters/

References

Aad, G., Abbott, B., Abdallah, J., Abdelalim, A. A., Abdesselam, A., Abdinov, O., Abi, B., Abolins, M., Abramowicz, H., Abreu, H., Acerbi, E., Acharya, B. S., Adams, D. L., Addy, T. N., Adelman, J., Aderholz, M., Adomeit, S., Adragna, P., Adye, T., ... Zwalinski, L. (2012). Performance of missing transverse momentum reconstruction in proton-proton collisions at$\sqrt{s}$ = 7 TeV with atlas. *European Physical Journal C*, *72* (1), 1-35. https://doi.org/10.1140/epjc/s10052-011-1844-6

Akin, M., Eyduran, E., & Reed, B. M. (2017). Use of RSM and CHAID data mining algorithm for predicting mineral nutrition of hazelnut. *Plant Cell, Tissue and Organ Culture (PCTOC)*, *128* (2), 303-316.

Ballestar, M. T., Doncel, L. M., Sainz, J., & Ortigosa-Blanch, A. (2019). A novel machine learning approach for evaluation of public policies: An application in relation to the performance of university researchers. *Technological Forecasting and Social Change*, *149*. https://doi.org/10.1016/j.techfore.2019.119756

zBallestar, M. T., Garcia-Lazaro, A., Sainz, J., & Sanz, I. (2022). Why is your company not robotic? The technology and human capital needed by firms to become robotic. *Journal of Business Research, 142,* 328-343. https://doi.org/10.1016/j.jbusres.2021.12.061

Ballestar, M. T., Grau-Carles, P., & Sainz, J. (2018). Customer segmentation in e-commerce: Applications to the cashback business model. *Journal of Business Research*, *88*, 407-414. https://doi.org/10.1016/j.jbusres.2017.11.047

Brunello, G., & Paola, M. De. (2014). The costs of early school leaving in Europe. *IZA Journal of Labor Policy*, *3* (1), 1-31. https://doi.org/10.1186/2193-9004-3-22In

Burgess, S. (2020). *How we should deal with the lockdown learning loss in England's schools*. VOX Cepr Policy Portal. https://bit.ly/3Eom12M

Chassignol, M., Khoroshavin, A., Klimova, A., & Bilyatdinova, A. (2018). Artificial Intelligence trends in education: A narrative overview. *Procedia Computer Science*, *136*, 16-24. https://doi.org/10.1016/j.procs.2018.08.233.

Chatterjee, S., & Bhattacharjee, K. K. (2020). Adoption of artificial intelligence in higher education: a quantitative analysis using structural equation modelling. *Education and Information Technologies*, *25* (5), 3443-3463. https://doi.org/10.1007/s10639-020-10159-7

Crescenzi, R., Giua, M., & Sonzogno, G. V. (2021). Mind the Covid-19 crisis: An evidence-based implementation of Next Generation EU. *Journal of Policy Modeling*, *43* (2), 278-297. https://doi.org/10.1016/J.JPOLMOD.2021.03.002

De la Fuente, A., & Jimeno, J. F. (2009). The private and fiscal returns to schooling in the European Union. *Journal of the European Economic Association*, *7* (6), 1319-1360. https://doi.org/10.1162/JEEA.2009.7.6.1319

Doncel, L. M., Durá, P., Grau, P., & Sainz, J. (2014). Fiscal sustainability and immigration in the Madrid Region. *International Migration, 52* (6), 180-196. https://doi.org/10.1111/imig.12004

Dželihodžić, A., & Jonko, D. (2016). Comparison of ensemble classification techniques and single classifiers performance for customer credit assessment. *Modeling of Artificial Intelligence*, *3*, 140-150.

European Comission (2010). *EUROPE 2020: A strategy for smart, sustainable and inclusive growth*. http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2010:2020:FIN:EN:PDF

Golden, G. (2020). *OECD iLibrary | Education policy evaluation: Surveying the OECD landscape*. (*OECD Education Working Papers, 236*). OECD Publishing. https://www.oecd-ilibrary.org/education/education-policy-evaluation_9f127490-en

Goos, M., Manning, A., & Salomons, A. (2009). Job polarization in Europe. *American Economic Review*, *99* (2), 58-63. https://doi.org/10.1257/aer.99.2.58

Gouëdard, P., Pont, B., & Viennet, R. (2020). Education responses to COVID-19: Implementing a way forward. *OECD Education Working Papers, 224*. OECD Publishing. https://doi.org/https://doi.org/10.1787/8e95f977-en

Gregory, T., Salomons, A., & Zierahn, U. (2019). Racing With or Against the Machine ? Evidence from Europe. *IZA Institute of Labor Economics, 2019* (12063).

Hosmer Jr., D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression*. John Wiley & Sons.

Kass, G. V. (1980). An exploratory technique for investigating large quantities of categorical data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *29* (2), 119-127. https://doi.org/10.2307/2986296

Khosravi, P., Kazemi, E., Zhan, Q., Malmsten, J. E., Toschi, M., Zisimopoulos, P., Sigaras, A., Lavery, S., Cooper, L. A. D., & Hickman, C. (2019). Deep learning enables robust assessment and selection of human blastocysts after in vitro fertilization. *NPJ Digital Medicine*, *2* (1), 1-9.

Kraft, M. (2015, August 20). What If Every Struggling Student Had a Tutor? *Education Week*. https://www.edweek.org/education/opinion-what-if-every-struggling-student-had-a-tutor/2015/08

Li, Z., & Eastman, J. R. (2006). Commitment and typicality measurements for fuzzy ARTMAP neural network. *Geoinformatics 2006: Geospatial Information Science*, *6420*. https://doi.org/10.1117/12.7129986420, 64201I.

Maimon, O., & Rokach, L. (2005). *Data mining and knowledge discovery handbook*. Springer.

Nickow, A., Oreopoulos, P., & Quan, V. (2020). The impressive effects of tutoring on prek-12 learning: A systematic review and meta-analysis of the experimental evidence. *National Bureau of Economic Research Working Papers Series, 27476*. https://doi.org/10.3386/w27476

OECD. (2018). *Education policy outlook 2018: Putting student learning at the centre*. OECD.

Organic Law 10/2002, of 23 December 2002, on the Quality of Education (LOCE). *Spanish Official State Gazette, 307*, of 24 December 2002, pp. 45188 to 45220. https://www.boe.es/eli/es/lo/2002/12/23/10

Organic Law 2/2006, of 3 May, on Education (LOE). *Spanish Official State Gazette, 106*. https://www.boe.es/eli/es/lo/2006/05/03/2/con

Organic Law 8/2013, of 9 December, for the Improvement of the Quality of Education (LOMCE). *Spanish Official State Gazette, 295*, pages 97858 to 97921. https://www.boe.es/eli/es/lo/2013/12/09/8

Organic Law 3/2020, of 29 December, which amends Organic Law 2/2006, of 3 May, on Education (LOMLOE). *Spanish Official State Gazette*, *340*, 30 December 2020, pages 122868 to 122888. https://www.boe.es/diario_boe/txt.php?id=BOE-A-2020-17264

Organic Law 3/2020, of 29 December, which amends Organic Law 2/2006, of 3 May, on Education (LOMLOE). *Spanish Official State Gazette*, *340*, 30 December 2020, pages 122868 to 122888. https://www.boe.es/diario_boe/txt.php?id=BOE-A-2020-17264

Paliwal, M., & Kumar, U. A. (2009). Neural networks and statistical techniques: A review of applications. *Expert Systems with Applications*, *36* (1), 2-17.

Pont, B., & Montt, G. (2014). *Education policy outlook. Spain*. OECD.

Porte, C. de la, & Jensen, M. D. (2021). The next generation EU: An analysis of the dimensions of conflict behind the deal. *Social Policy & Administration*, *55* (2), 388-402. https://doi.org/10.1111/SPOL.12709

Ramaswami, M., & Bhaskaran, R. (2010). A CHAID based performance prediction model in educational data mining. *International Journal of Computer Science*, 7 (1), 10-18.

Saltelli, A., Tarantola, S., Campolongo, F., & Ratto, M. (2004). *Sensitivity analysis in practice: a guide to assessing scientific models (Vol. 1)*. Wiley Online Library.

Schleicher, A. (2020). The impact of COVID-19 on education: Insights from education at a glance 2020. *OECD Journal: Economic Studies*. https://www.oecd.org/education/the-impact-of-covid-19-on-education-insights-education-at-a-glance-2020.pdf

Slavin, R. E. (2016). Evidence-Based Education Policies: Transforming Educational Practice and Research. *Educational Researcher, 31* (7), 15-21. https://doi. org/10.3102/0013189X031007015

Wolszczak-Derlacz, J., & Parteka, A. (2011). Efficiency of European public higher education institutions: A two-stage multicountry approach. *Scientometrics*, *89* (3), 887-917. https://doi.org/10.1007/s11192-011-0484-9

Yin, L., Ge, Y., Xiao, K., Wang, X., & Quan, X. (2013). Feature selection for high-dimensional imbalanced data. *Neurocomputing*, *105*, 3-11.

Yusuf, A. (2007). *Standard-Setting at UNESCO: Conventions, recommendations, declarations and charters adopted by UNESCO, 1948-2006 (Vol. 2).* Martinus Nijhoff Publishers.

Authors' biographies

María Teresa Ballestar holds a degree in Statistics from the University of Zaragoza, a degree in Market Research and Techniques and a Master's degree in Information and Knowledge Society from the Universitat Oberta de Catalunya. PhD in Big Data methodologies and technologies applied to Economics at the Universidad Rey Juan Carlos. She is also an Associate Professor at the Universidad Rey Juan Carlos. In recent years she has held management positions and led projects in digital transformation, innovation, data analytics and data science. She has published more than a dozen scientific papers and collaborated with various media.

iD https://orcid.org/0000-0001-8526-7561

Jorge Sainz Orcid holds a degree in Economics from the Universidad Complutense de Madrid and in Law from the UNED. He holds a PhD in Economics (URJC) and an MBA (specialising in finance and public policy) from the Simon School, University of Rochester. He is Professor of Applied Economics at the Universidad Rey Juan Carlos de Madrid and Visiting Fellow at the Institute for Policy Research, University of Bath.

He has been Advisor to the Cabinet of the Regional Minister of Education and General Deputy Director of Research at the Regional Ministry of Education of the Community of Madrid. In the Ministry of

Education, Culture and Sport he has been General Director of University Policy and General Secretary of Universities.

https://orcid.org/0000-0001-8491-3154

Ismael Sanz is Associate Professor in the Department of Applied Economics I at the Universidad Rey Juan Carlos. He holds a PhD in Applied Economics from the Faculty of Economics and Business Studies at the Universidad Complutense de Madrid (UCM). He has been General Director of Innovation, Scholarships and Grants of the Ministry of Education of the Community of Madrid, Director of the National Institute for Educational Evaluation (INEE) of the Ministry of Education, Culture and Sport (2012-2015) and Chair of the Strategic Development Group of PISA of the OECD. He is currently Vice-Rector for Quality at the URJC. Author of articles in high impact research journals.

https://orcid.org/0000-0003-1286-4124