How much gold is in the sand? Data mining with Spain's PISA 2015 results

¿Cuánto oro hay entre la arena? Minería de datos con los resultados de España en PISA 2015

Inmaculada ASENSIO MUÑOZ, PhD. Lecturer. Complutense University of Madrid (*macu@edu.ucm.es*).
Elvira CARPINTERO MOLINA, PhD. Assistant Professor. Complutense University of Madrid (*ecarpintero@edu.ucm.es*).
Eva EXPÓSITO CASAS, PhD. Assistant Professor. National Distance Education University (*evaexpositocasas@edu.uned.es*).
Esther LÓPEZ MARTÍN, PhD. Assistant Professor. National Distance Education University (*estherlopez@edu.uned.es*).

Abstract:

Since the start of the PISA evaluations there have been numerous studies that have metaphorically tried «to separate the gold from the sand», in other words, to derive useful knowledge to guide educational practice and policy from the vast amount of data collected. However, research that uses data mining techniques to extract knowledge from the databases provided by the OECD has been less common. This paper analyses the context questionnaires from a metric perspective using a methodology based on data mining with «regression trees». Its main goal is to discover how much value (how much «gold») is in the items that compose these questionnaires, considering their use as predictors of the performance of Spanish students. The results provide a list of the items selected in the six questionnaires and their predictive value. It also provides a methodological approach to help improve the productivity of educational research derived from PISA.

Keywords: PISA 2015, regression trees, context questionnaire, Spain, validity.

Resumen:

Desde el inicio de las evaluaciones PISA abundan los estudios que pretenden, en lenguaje metafórico, «separar el oro de la arena», esto es, producir, de la cantidad ingente de datos recogidos, conocimiento útil que guíe la práctica y las políticas educativas. Pero no son frecuentes las investigaciones que usan técnicas de minería de datos para la extracción de dicho conocimiento. En este trabajo se analizan los cuestionarios de contexto desde una perspectiva métrica, con una metodología basada en «árboles de regresión» destinada a descubrir cuánto «oro» hay en los ítems que los componen, atendiendo a su uso como predictores del desempeño de los jóvenes españoles. Como resultado se obtiene un listado de los ítems más importantes en los seis cuestionarios, junto con el valor predictivo de los mis-

revista española de pedagogía
year LXXVI, n. 270, May-August 2018, 225-245

Revision accepted: 2018-01-23.

This is the English version of an article originally printed in Spanish in issue 270 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: Asensio Muñoz, I., Carpintero Molina, E., Expósito Casas, E., & López Martín, 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. doi: https://doi.org/10.22550/REP76-2-2018-02

mos. Se aporta un enfoque metodológico que puede contribuir a mejorar la productividad de la investigación pedagógica derivada de PISA.

Descriptores: PISA 2015, árboles de regresión, cuestionario de contexto, España, validez.

1. Introduction

The aims of PISA (Programme of International Student Assessment) include providing indicators of the effectiveness, efficiency, and equity of educational systems, as well as setting reference points to allow international comparisons and oversight of trends over time (OECD, 2016). More than a decade and a half since it was launched, it is a good time to take stock and reflect on whether this international evaluation is achieving its objectives and whether it is the gold mine of information that was expected. From a specifically pedagogical perspective, analysing the extent to which it contributes to increasing our knowledge of education and of educational systems is of interest. Its broad application and the metrical techniques it uses allow for comparisons of the spend on education and the results achieved, at both national and international levels and from synchronic and diachronic perspectives (Hopfenbeck et al., 2017) with a significant media impact. However, despite the efforts made, the most important question now concerns the objective of looking for simple or complex indicators of effectiveness and identifying which input, process, and output variables (non-cognitive) are most relevant, given their relationship with the performance levels evaluated. For the biggest critics, the research being carried out based on this large-scale evaluation is apparently not as productive as expected in creating useful knowledge for improving education. On these lines, Hanberger (2014) states that PISA suffers from issues with internal and external validity, and, in the best case, only works as an alarm system and as something to facilitate changes in policy at a national level. In Spain and many of its autonomous regions, the interest in participating in the programme to acquire knowledge to facilitate adopting measures to improve education has for long time been apparent (Instituto de Evaluación, 2007). However, there are arguments to support the position that PISA lacks specific value for this purpose (Carabaña, 2009, 2015), basically because the educational variables associated with the performance levels obtained are still not clearly apparent.

Validity is a complex and fundamental metrical concept (AERA, APA, and NCME, 2014) and could be the basis of this circumstance. Carabaña (2015) sees flaws in the definition of the competencies, thus raising a potential problem with the validity of the performance measurements themselves. However, there might also be weaknesses that relate to the validity of the measurements provided by the background questionnaires or context questionnaires, which until now have usually been regarded as being of «secondary importance» (González-Montesinos & Backhoff, 2010, p. 14) but are now taking on an increasingly prominent role (OECD, 2016). Despite the important role of these questionnaires in international evaluations, there is hardly any data relating to the reliability of the measurements they provide (Rutkowski & Rutkowski, 2010, 2017) nor has proof of their validity been reported (Taut & Palacios, 2016). According to De La Orden and Jornet (2012), the main problems with sample-based evaluative studies include shortcomings in the conceptual and operative definition of the context measurements «and their low metric controls» (p. 78). In PISA 2015 a theoretical effort was made concerning validity as an internal structure, involving identifying underlying constructs, defining simple and complex indicators and indices (González-Such, Sancho-Álvarez, and Sánchez-Delgado, 2016), and establishing the possible relationships between them. Nonetheless, obtaining proof of validity for the context measurements is not easy given the great quantity and complexity of the information they provide and the many uses and interpretations derived from them, ranging from imputation of missing data and estimating plausible values (Kaplan and Su, 2016) to establishing subgroups in the population of 15-year-olds evaluated, «making it possible to introduce descriptors to the results (gender, ethnicity, educational level of the parents, type of school etc.)» (Martínez Arias, 2006, p. 120). In this piece, which examines the use of the PISA results as country-level assessment information and centres on identifying the factors that are most closely related to performance (Taut & Palacios, 2016), we propose a methodology based on using knowledge extraction techniques that are collectively known as «data mining» as these can, from an empirical and exploratory focus, complement the selection of variables done by the OECD (2016) in accordance with essentially political and also theoretical criteria, as explained above. This focus is proposed as ideal for discerning how much of this mass of available information is useful for the objective of explaining differences in performance, helping us «separate the gold from the sand». Going beyond the precious-metal metaphor, in this piece we connect «data mining» to PISA as this term includes a new generation of techniques and tools that aim to extract useful knowledge from the information held in large databases (Knowledge Discovery in Databases, KDD), with the special feature that this knowledge does not necessarily fit a predetermined model but instead an emerging one (Hernández Orallo, Ramírez, and Ferri, 2004). Although use of data mining in education (Castro and Lizasoaín, 2012) has increased in recent years, especially in connection with the development of e-learning, some research also uses it predict performance levels (Alcover et al., 2007; Thai Nghe, Janecek, & Haddawy, 2007; Lizasoain, 2012; Muñoz Ledesma, 2015; Ruby & David, 2015; Thakar, Mehta, & Manisha, 2015; Lakshmipriya and Arunesh, 2017), it being especially appropriate for large-scale evaluations that study efficiency (Santín, 2006) or the variables that affect the competences evaluated (Yu, Kaprolet, Jannasch-Pennell, & DiGangi, 2012; Kiray, Gok, & Bozkir, 2015; Aksu & Güzeller, 2016; Gorostiaga & Rojo-Álvarez, 2016; Idil, Narli, & Aksoy, 2016). Blanco-Blanco, Asensio, Carpintero, Ruiz de Miguel, & Expósito (2017) illustrate the use of tree techniques to give a solid foundation to interpretations of the scores obtained in educational evaluation, using them to obtain proof of validity.

By focussing on the particular use of context questionnaires as an instrument for measuring the variables that explain performance, this study aims to explore the databases derived from the PISA study for Spain to discover how much pedagogical knowledge they contain and what items they provide. In short, taking the item as its unit of analysis, this piece will seek arguments for the validity of the measurements obtained through the PISA context questionnaires, with performance in sciences, reading, and mathematics as its criterion, and using the data mining technique with regression trees as its methodology. In this way, it will attempt to help make progress by studying the validity of the context questionnaires based on proof of the measurements taken from them by identifying, ordering, and selecting the items from them that are most relevant thanks to their value for predicting the competencies evaluated in PISA 2015 in the setting of Spain's educational system.

2. Method

The methodological approach for obtaining proof of validity depends on the type of interpretations that are hoped to be made based on the scores obtained. Most studies look for proof regarding the internal structure of the construct, and so the most commonly used approaches are exploratory factor analysis (EFA) and confirmatory factor analysis (CFA). However, in this case the questions in the context questionnaires are analysed with the aim of studying their predictive and explanatory capacity, and so it makes more sense to use a multivariate approach which introduces the answers to the items as independent variables and the performance score as the dependent variable. Accordingly, the most common methodological option in these cases is linear or logistic regression analysis, but in PISA the competencies are continuous variables and the variables measured based on the context questionnaires are of different types. Consequently, we rely here on the non-parametric option of regression trees. These work appropriately with this complexity of data types in a single analysis without needing them to be transformed and are robust when faced with the presence of outliers and missing values (Streifer & Schumann, 2005).

2.1. Sample

The six context questionnaires used in PISA 2015 are analysed, in all cases, using the performance level obtained by students from Spain in this evaluation as the validation criterion. Consequently, the study sample comprises the 15-year-olds from Spain who participated in the evaluation, the parents of these students who completed the questionnaire intended for families, and the management and teachers from the schools where the students were enrolled (Table 1). It should be noted that the data have not been weighted by the student final weights as the aim is not to make international comparisons, but instead to explore the situation in Spain (OECD, 2014).

Table 1. Number of responses in each of the questionnaires analysed.

	N	Respondent
Student questionnaire	39066	Students
Educational career questionnaire	38384	Students
ICT familiarity questionnaire	38585	Students
Parents questionnaire	4753	Parents or guardians
School questionnaire	1177	Principals
Teacher questionnaire	3894	Teachers

Source: Own elaboration.

2.2. Instruments

The theoretical framework for the PISA 2015 context questionnaires is presented in the study report (OECD, 2016).

The student questionnaire is administered during the evaluation of the students' knowledge and skills and takes around 35 minutes to answer. The questions it contains concern the students' characteristics, family and home, the students' view of their lives, their experience at school, timetable, time spent studying, study of sciences at school, and view of science. It comprises 224 items.

The educational career questionnaire, ICT skills questionnaire, family questionnaire, and teacher questionnaire are optional for the participating countries. The first of them contains 164 items and the second 81. The family questionnaire comprises 146 items concerning family-school relationship, educational career, and parents' views on science. There are two versions of the teacher questionnaire: one for science teachers (102 items) and one for teachers of other subjects (107 items). In both cases, the questionnaire is structured around context information, initial training and professional development, the school, and teaching practices, whether general or specifically relating to the sciences.

In addition, the school principal answers the school questionnaire. This comprises 229 items and makes it possible to collect information about the context and conditions of the school, school administration, teaching staff, supervision and evaluation, organisation, and the school atmosphere.

Finally, it should be noted that the scores obtained by Spanish students in the three competencies evaluated in PISA 2015, in other words, the 10 plausible estimated values for sciences, reading, and mathematics, have been taken as the dependent variables.

2.3. Procedure

One of the most popular decision tree algorithms is CART (Classification And Regression Trees) (Strobl, Malley, & Tutz, 2009), developed by Breiman, Friedman, Olshen, and Stone (1984). In this piece they are used as the main method of analysis, although CHAID (Chi Automatic Interaction Detection) is used to complement them. The CART process is frequently used as a segmentation methodology and can be used as a non-parametric supervised learning technique (Izenman, 2008). This comprises a recursive partitioning process applied to complex problems, which is based on the principle of «divide and conquer» (Hernández Orallo, Ramírez, & Ferri, 2004). It provides binary segmentation and a measure of the importance of the independent variables. Although it is used with a variety of objectives, it is often felt that tree analysis is classificatory when the dependent variable is nominal or ordinal and that it is regressive when the dependent variable is a scale. For its part, CHAID performs segmentations that can have more than two categories, allows selection of the independent variables that interact with the dependent variable (Kass, 1980) and provides p-values. The following process was used in this piece to identify, order, and select the context variables that make the greatest contribution to explaining student performance:

I. Estimating the initial models using CART, introducing the scores in the three competencies studied as dependent variables, and all of the items from the six questionnaires analysed as predictors. Independent estimates were made for each of the 10 plausible values (6 questionnaires × 3 competencies × 10 plausible values = 180 estimated models). The average risk values were then calculated for each questionnaire by subject. Using these, the joint predictive value of the items from the questionnaires was quantified. The general stopping rules set as default in the program were used.

II. Calculation with CART of the importance of each independent variable as the sum of the reduction of the impurity measure produced by the best division of said variable in each of the nodes (Breiman et al., 1984, p. 147). This calculation was also performed by subject and plausible value, so that the mean of the importance of each explanatory variable in each of the competencies evaluated could be estimated. The range is established based on the mean.

III. Estimation of the initial models, this time using the CHAID algorithm, which provides a selection of predictors. In this way, 180 different models are estimated, the results for which make it possible to identify the variables for each questionnaire that interact with the dependent variable.

IV. Selection of the variables that meet the following inclusion criteria: a) their standardised mean importance, estimated using CART, is at least 10% and b) they are included by CHAID as a significant influencing variable for at least one of the plausible values. The selection criteria established are intended to create a parsimonious list of variables that does not increase the level of risk obtained by including all of the items in the model.

V. Reestimation of the mean risk values to quantify the joint predictive value of the items from the shortened questionnaires.

The analyses were performed using the IBM SPSS Statistics version 22 program.

3. Results

All of the items that make up each of the questionnaires were included in the initial model, while the final model only included the ones that met the criteria for inclusion. CART provides a risk estimate which, if divided by the total variance of the dependent variable (S^2), tells us of the proportion of it that is not explained by the variables included in the model (Risk/S^2). The global predictive value of each questionnaire, complete and shortened, was obtained from the square root of the proportion of variance explained. The initial model (Table 2) shows that the student questionnaire is the most informative in all subjects, while the other five questionnaires have a lower predictive value. The questionnaire completed by teachers is the one that makes the smallest contribution to explaining differences in the three subjects studied.

TABLE 2. Global predictive value of the items from the different context questionnaires in the initial models obtained with CART.

SCIENCES	S^2	Risk	Risk/S^2	S^2 explained	Predictive value
Students	7549.80	3955.58	0.52	0.48	0.69
Educational career	7549.80	6007.06	0.80	0.20	0.45
ICT	7549.80	6168.04	0.82	0.18	0.43
Family	7549.80	5460.54	0.72	0.28	0.53
Principal (school)	1181.95	835.56	0.71	0.29	0.54
Teacher	1227.68	1092.97	0.89	0.11	0.33
READING	**S^2**	**Risk**	**Risk/S^2**	**S^2 explained**	**Predictive value**
Students	7643.46	4258.73	0.56	0.44	0.67
Educational career	7643.46	5995.89	0.78	0.22	0.46
ICT	7643.46	5551.23	0.73	0.27	0.52
Family	7643.46	5607.55	0.73	0.27	0.52
Principal (school)	1222.94	941.63	0.77	0.23	0.48
Teacher	1339.86	1174.27	0.88	0.12	0.35

Mathematics	S^2	Risk	Risk/S^2	S^2 explained	Predictive value
Students	6926.07	3822.95	0.55	0.45	0.67
Educational career	6926.07	5587.38	0.81	0.19	0.44
ICT	6926.07	5831.06	0.84	0.16	0.40
Family	6926.07	5177.52	0.75	0.25	0.50
Principal (school)	1130.58	835.48	0.74	0.26	0.51
Teacher	1199.21	1058.54	0.88	0.12	0.34

Source: Own elaboration.

Tables 3 to 8 present the variables that are selected as they meet the inclusion criteria set for each questionnaire as well as the order of importance for each item by subject, calculated using CART[3].

TABLE 3. Selection of items from the student questionnaire.

DESCRIPTION OF THE ITEM	Science	Reading	Mathematics
Grade the student is in (11, 10, 9, 8, 7)	1st	1st	1st
Student expectations	2nd	2nd	2nd
Has repeated 'ISCED 2' (a secondary course)	3rd		3rd
Possesses information about the increase of greenhouse gases in the atmosphere	4th	5th	6th
Gives up easily when confronted with a problem and is often not prepared for class	5th	3rd	7th
Attends chemistry courses this year	6th	8th	8th
Attends physics courses this year	7th	12th	9th
Self-reported ease of explaining why earthquakes occur more frequently in some areas than in others.	8th	11th	10th
Science classes per week	9th		
Attends biology courses this year	10th		
Works for pay before going to school	11th	9th	
Has repeated 'ISCED 1' (a primary course)			4th
Number of books		4th	5th
Believes it is good to try experiments more than once to make sure of the findings.		6th	
Believes good answers are based on evidence from experiments.		7th	

DESCRIPTION OF THE ITEM	Science	Reading	Mathematics
Wants to get top grades at school and continues working on tasks until everything is perfect.		10th	
Number of classes per week			11th

Source: Own elaboration.

Seventeen items meet the inclusion criteria in the student questionnaire (Table 3). The two most important variables in the three subjects are the grade the student is studying followed their level of expectation. In the educational career questionnaire 30 items meet the inclusion criteria (Table 4). «Changing study programme» is the most important variable thanks to its relationship with Reading and «not needing additional mathematics instruction» is the most important given its relationship with Science and Mathematics.

Table 4. Selection of items from the educational career questionnaire.

ITEM	Science	Reading	Maths
I don't attend additional mathematics instruction in this school year because I don't need it	1st	5th	1st
Have you ever changed your 'study programme'?	2nd	1st	4th
I don't attend additional science instruction in this school year because I don't need it	3rd		2nd
Hours per week you attend additional instruction in art	4th	2nd	6th
Hours per week you attend additional instruction in science (or broad science)	5th		
Attending additional mathematics instruction at school	6th	4th	7th
Other people regularly help me with my homework or private study.	7th		8th
Did you change schools when you were attending 'ISCED 2'?	8th	9th	10th
Comparing help received from the teacher in classes at school and in additional instruction	9th	7th	
Attending additional language instruction at school	10th	10th	14th
My sister(s)/brother(s) regularly help me with my homework or private study	11th	15th	
Differences in the hints and strategies for solving mathematics tasks provided in lessons in school and in additional instruction	12th		

ITEM	Science	Reading	Maths
My grandparents regularly help me with my homework or private study	13th	13th	16th
Attending additional instruction in pre-primary education	14th	16th	18th
Hours per week you attend additional instruction in foreign languages	15th	19th	17th
The teacher for the additional language instruction is one of my regular teachers in this year's school courses	16th	20th	
Nobody regularly helps me with my homework or private study	17th		19th
Hours per week you attend additional language instruction			3rd
Hours per week you attend additional mathematics instruction		3rd	5th
Attending additional science instruction at school		6th	
Other family members regularly help me with my homework or private study			9th
How many years altogether have you attended additional instruction?			11th
The additional science instruction I attend covers chemistry			12th
Hours per week you attend additional music instruction		8th	13th
Participation in additional mathematics instruction through video recorded instruction by a person		11th	
I attend additional science instruction in this school year because I was attracted by the tutoring advert		12th	
Participation in additional science instruction through Internet tutoring with a person (including, for example, Skype)		14th	
I don't attend additional science instruction in this school year because I don't have the money			15th
The additional science instruction I attend covers physics		17th	
Participation in additional language instruction during this school year through Internet or computer tutoring with a program or application		18th	

Source: Own elaboration.

Twenty-one variables are selected from the ICT familiarity questionnaire (Table 5), the most important one in science and reading being the opinion of the Internet as a source of information. The school's equipment (projectors) is an especially important variable in mathematics.

TABLE 5. Selection of items from the ICT familiarity questionnaire.

ITEM	Science	Reading	Mathematics
I believe the Internet is a great resource for obtaining information I am interested in (e.g. news, sports, dictionary)	1st	1st	5th
A data projector is available for me to use at school	2nd	6th	1st
Frequency of using social networks for communication with teachers outside of school	3rd	2nd	8th
How old were you when you first used a digital device?	4th	10th	2nd
Frequency of downloading learning apps on a mobile device outside of school	5th	3rd	10th
I feel comfortable using my digital devices at home	6th	5th	9th
I have a USB (memory) stick	7th	15th	4th
Frequency of use of digital devices to obtain practical information from the Internet outside of school	8th	9th	15th
I have available to use at school a Tablet , iPad, BlackBerry, PlayBook	9th	7th	6th
Frequency of use of email outside of school	10th		11th
An e-book reader is available for me to use at school	11th	8th	3rd
I have a printer at home	12th		7th
Frequency of checking the school's website for announcements outside of school	13th		
Frequency of downloading science learning apps on a mobile device outside of school		4th	12th
Frequency of browsing the Internet for schoolwork outside of school (for example, presentations)		11th	
Frequency of downloading, uploading or browsing material from the school's website (e.g. timetable or course materials) outside of school		12th	

ITEM	Science	Reading	Mathematics
How old were you when you first used a computer?			13th
During a typical weekday, how long do you use the Internet at school?		13th	
I have an Internet connection at home		14th	
During a typical weekday, how long do you use the outside of school?			14th
Frequency of use of digital devices to upload your own created contents for sharing outside of school			16th
Frequency of browsing the Internet to follow up lessons (for example, for finding explanations)		16th	

Source: Own elaboration.

In the questionnaire aimed at families (Table 6), 35 items were selected, of which interest in a science-related career was most important in science and reading. Family income is the most important variable in relation to performance in mathematics.

Table 6. Selection of items from the family questionnaire.

ITEM	Science	Reading	Mathematics
Does your child show an interest in working in a science-related career?	1st	1st	3rd
Has your child shown interest in studying science after completing secondary school?	2nd	2nd	2nd
Do you expect your child will study science after completing secondary school?	3rd	4th	4th
What is your annual income?	4th	3rd	1st
At what age did your child start attending 'ISCED 1'?	5th	5th	5th
Main reason your child attended pre-primary education	6th	8th	7th
During the last academic year, my participation in activities at my child's school been hindered by the way to school being unsafe	7th	6th	10th
During the last academic year, I have discussed my child's behaviour on the initiative of one of his/her teachers	8th	11th	13th

ITEM	Science	Reading	Mathematics
During the last academic year I have talked about how to support learning at home with my child's teachers	9th	13th	9th
When choosing a school for my child, it is important that the school has financial aid available	10th		
Main reason your child attended supervision or child care	11th	16th	11th
When your child was 10-years-old, how often did he/she read books about scientific discoveries?	12th		
During the last academic year, I have discussed my child's behaviour with a teacher on my own initiative	13th	15th	12th
During the last academic year, I have discussed my child's progress on the initiative of one of their teachers	14th	20th	14th
My child attended a supervision and care arrangement at the age of one	15th		16th
My child attended a supervision and care arrangement before the age of one	16th		15th
How often you help your child with his/her science homework?	17th	22nd	
How often you obtain science-related materials for your child?	18th	23rd	
Does anybody in your family (including you) work in a science-related career?		7th	6th
When your child was about ten, how often did he/she experiment with a science kit, electronics kit, or chemistry set, or use a microscope or telescope?			8th
How often you discuss science-related career options with your child		9th	
I believe science is valuable for society		10th	
When choosing a school for my child, it is important that the expenses are low		12th	
During the last academic year, I have been supportive of my child's efforts at school and his/her achievements		14th	
Type of provider offered this pre-primary education arrangement		17th	

ITEM	Science	Reading	Mathematics
When your child was 10-years-old, how often did he/she fix broken objects?		18th	
During the last academic year, I have supported my child when he/she is facing difficulties at school		19th	
I believe science is relevant to me			17th
How many parents of your child's friends at this school do you know?			18th
My child started attending pre-primary education aged one		21st	19th
My child started attending pre-primary education aged two			20th
In what country was your child's maternal grandmother born?			21st
When your child was about 10, how often would he/she watch TV programmes about science?			22nd
My child attended a supervision and care arrangement at the age of two.		24th	
Before attending school, my child was taken care of by an adult untrained in child care (not a relative).		25th	23rd

Source: Own elaboration.

After performing the analyses, 29 items were selected from the school questionnaire (Table 7). In reading, the most important variable is ownership, while for science and mathematics, it is the proportion of disadvantaged students.

Table 7. Selection of items from the school questionnaire.

ITEM	Science	Reading	Mathematics
Percentage of 15-year-old students from socioeconomically disadvantaged homes	1st	3rd	1st
Extent to which learning is hindered by student truancy	2nd	2nd	2nd
Ownership	3rd	1st	3rd
The principal is responsible for firing teachers	4th		5th
The local or regional educational agency is responsible for selecting teachers for hire	5th		7th

ITEM	Science	Reading	Mathematics
Inadequate or poor quality physical infrastructure (e.g., building, grounds, heating/cooling, lighting and PA system)	6th	5th	11th
Extent to which learning is hindered by students lacking respect for teachers	7th		15th
Number of girls enrolled at the school	8th	4th	9th
Part-time teachers	9th		12th
The school governing board is responsible for deciding on budget allocations within the school	10th		
Part-time fully-certified teachers	11th		14th
Data projectors in the school available to 15-year-old students	12th	7th	
The principal is responsible for establishing student assessment policies	13th		19th
The principal is responsible for selecting teachers for hire			4th
Extent to which learning is hindered by students skipping classes			6th
The local or regional educational agency is responsible for firing teachers			8th
Number of boys enrolled at the school			10th
Interactive whiteboards in the school available to students in the 10th grade		6th	13th
Location of the school			16th
Total number of 15-year-old students in the school		8th	17th
Implementation of a standardized policy for science subjects		9th	
Percentage of 15-year-old students whose heritage language is different from the language of the test		10th	
Full-time teachers with a doctoral or professional degree		11th	
Full-time teachers			18th
Full-time fully-certified teachers			20th
The school governing board is responsible for establishing student disciplinary policies			21st
Implementing teaching and learning quality measures based on internal evaluation			22nd

ITEM	Science	Reading	Mathematics
Full-time teachers with a master's degree			23rd
Computers connected to the Internet available to 15-year-old students			24th

Source: Own elaboration.

Finally, in the questionnaire aimed at teachers, there is a significant consistency for the three subjects as all of the 11 variables selected are important for sciences and 7 are shared (Table 8). The teachers' perceptions of the schools' infrastructure is the variable that is most closely related to schools' average performance in sciences, while the teachers' professional stability is most closely related to performance in reading and mathematics.

Table 8. Selection of items from the teacher questionnaire.

ITEM	Science	Reading	Mathematics
The educational capacity of your school is hindered by inadequate or poor quality physical infrastructure	1st	3rd	2nd
In how many schools have you worked over the course of your teaching career?	2nd	1st	1st
The educational capacity of your school is hindered by a lack of physical infrastructure	3rd	4th	3rd
I would recommend my school as a good place to work	4th	2nd	4th
The educational capacity of your school is hindered by a lack of teaching staff	5th	7th	6th
The educational capacity of your school is hindered by Inadequate or poor quality educational material	6th	5th	
Are you required to take part in professional development activities?	7th		8th
The educational capacity of your school is hindered by a lack of educational material	8th	6th	5th
The educational capacity of your school is hindered by a lack of assisting staff	9th		7th
Type of contract	10th	8th	9th
Type of working hours	11th		

Source: Own elaboration.

To confirm that the predictive values are maintained, the explanatory capacity of the shortened context questionnaires was quantified. That is to say, only the selected variables were inputted as independent variables and values very similar to those shown in Table 2 were obtained.

4. Conclusions

This work has considered the context measurements in relation with performance using a novel methodological perspective that allows an overview of the importance of each item on the background questionnaires in connection with the others for each competence evaluated. The research carried out provides proof of the validity of the six context questionnaires used in PISA 2015 for this purpose, although the student questionnaire, with a science coefficient of 0.68 has the greatest predictive power. This is virtually unchanged in its shortened format, despite only 17 relevant items being selected for the Spanish sample and for the use studied. Although the discussion about which variables and levels appear as most important cannot be tackled in-depth here as it goes beyond the objective of this research, the tables in the results section provide very illuminating information on this matter that leads us to reconsider the general conclusion from studies based on PISA, according to which it appears that students' socioeconomic conditions are the most important variables (Cordero, Crespo, & Pedraja, 2013). In our study, the items referring to these questions, such as «working for pay before going to school» or «number of books», come after the «grade, repeating a course, attending class, expectations», or «motivation» variables, all of which have a marked educational-psychology character. The very clear first place for the «grade» variable might sum up the student's entire educational career, their record of performance, which would explain its predictive value.

One of the clear contributions of the CART methodology is assigning numerical values to the variables according to their relative importance in explaining the variable, making it possible to quantify their «carats». The rating of the items from the context questionnaires gives an overview of which are the most (and least) important for explaining performance differences. This holistic vision is not possible with more traditional research methods as these are usually based on an intentional selection of the predictor variables with an essentially inferential objective, and so they provide information about which of the variables included in the model are significant and, at most, about their effect size taken in isolation or interacting only with the variables included in the model. However, the results of confirmatory studies in which the necessary thoroughness in including predictor variables is not achieved can lead to a representation of the educational reality of a country that inadequately informs decision makers. At this point it is worth recalling that in the confirmatory models it is vital to include all relevant variables, to minimise specification errors, which are of great importance and at the same time often neglected in studies that set out to explain educational results.

In short, while one issue with the methodology used here is its instability, as in recursive partitioning the decision about which variables to divide and the exact position of each cut-off point in the division are fundamental (Strobl, Malley, & Tutz, 2009), applying data mining techniques to studying the context questionnaires for the large scale evaluations does appear to be a useful initial exploratory tool for making an informed selection of the predictors to be considered in the secondary analyses derived from these evaluations, providing important statistical arguments that complement the necessary theoretical arguments.

We understand that education is organised in systems that learn and that their possibilities of learning depend largely on programmes such as PISA and on statistical learning tools, among which data mining techniques play an essential role (Hastie, Tibshirani, & Friedman, 2002). Data mining could be a methodological focus that will help educational researchers make better use of the information offered by PISA (Pereira, Perales, & Bakieva, 2016). This is also proving to be a programme that learns, to obtain proof that can successfully be used as a basis for making decisions concerning improvements. The use of classification and regression trees is therefore proposed as an interesting research option, not only with the items, but also with complex indicators, and international data to obtain proof of validity of the measurements in the different participating countries.

Notes

[1] The variables are arranged according to their importance in sciences. The ones that were not important in sciences but were in one or both of the other two competencies are presented shaded at the end.

[2] In the case of the school and teacher questionnaires, for the different estimates the average performance of the school on each of plausible values was taken as the dependent variable.

[3] The variables are arranged according to their importance in sciences. The ones that were not important in sciences but were in one or both of the other two competencies are presented shaded at the end.

References

Aksu, G. & Güzeller, C. O. (2016). Classification of PISA 2012 Mathematical Literacy Scores Using Decision-Tree Method: Turkey Sampling. *Education and Science*, *41* (185), 101-122.

Alcover, R., Benlloch, J., Blesa, P., Calduch, M. A., Celma, M., Ferri, C., … Zúnica, L. R. (2007). Análisis del rendimiento académico en los estudios de informática de la Universidad Politécnica de Valencia aplicando técnicas de minería de datos. En *Actas de las XIII Jornadas de Enseñanza Universitaria de la Informática* (pp. 163-170). Teruel: Universidad de Zaragoza.

Blanco-Blanco, A., Asensio, I. I., Carpintero, E., Ruiz De Miguel, C., & Expósito, E. (2017). Aplicaciones de la segmentación jerárquica en medición y evaluación de programas educativos. Ejemplos con un programa de educación financiera. *Educación XX1*, *20* (2), 235-257.

Breiman, L., Friedman, J. H., Olshen R. A., & Stone, C. J. (1984). *Classification and regression trees*. Boca Raton, FL: Chapman & Hall/CRC.

Carabaña, J. (2009). *Una vindicación de la escuela española. Lección inaugural del Curso Acadé-*

mico 2009-2010. Madrid: Facultad de Educación, UCM.

Carabaña, J. (2015). *La inutilidad de PISA para las escuelas*. Madrid: La Catarata.

Castro, M. & Lizasoaín, L. (2012). Las técnicas de modelización estadística en la investigación educativa: minería de datos, modelos de ecuaciones estructurales y modelos jerárquicos lineales. **revista española de pedagogía**, *70* (251), 131-148.

Cordero, J. M., Crespo, E., & Pedraja, F. (2013). Rendimiento educativo y determinantes según PISA: Una revisión de la literatura en España. *Revista de Educación, 362*, 273-297.

De La Orden, A. & Jornet, J. (2012). La utilidad de las evaluaciones de sistemas educativos: el valor de la consideración del contexto. *Bordón*, *64* (2), 69-88.

González-Montesinos, M. J. & Backhoff, E. (2010). Validación de un cuestionario de contexto para evaluar sistemas educativos con Modelos de Ecuaciones Estructurales. *RELIEVE, 16* (2), 1-17. Retrieved from: http://www.uv.es/RELIEVE/v16n2/RELIEVEv16n2_1.htm

González-Such, J., Sancho-Álvarez, C., & Sánchez-Delgado, P. (2016). Cuestionarios de contexto pisa: un estudio sobre los indicadores de evaluación. *RELIEVE, 22* (1), 1-7. doi: http://dx.doi.org/10.7203/relieve.22.1.8429

Gorostiaga, A. & Rojo-Álvarez, J. (2016). On the use of conventional and statistical-learning techniques for the analysis of PISA results in Spain. *Neurocomputing, 171*, 625-637.

Hanberger, A. (2014). What PISA intends to and can possibly achieve: A critical programme theory analysis. *European Educational Research Journal*, *13* (2), 167-180. doi: http://dx.doi.org/10.2304/eerj.2014.13.2.167

Hastie, T., Tibshirani, R., & Friedman, J. (2002). *The elements of statistical learning. Data mining, inference and prediction.* New York: Springer.

Hernández Orallo, J., Ramírez, M. J., & Ferri, C. (2004). *Introducción a la minería de datos*. Madrid: Pearson-Prentice Hall.

Hopfenbeck, T. H., Lenkeit, J., El Masri, Y., Cantrell, K., Ryan J., & Baird, J. A. (2017). Lessons Learned from PISA: A Systematic Review of Peer Reviewed Articles on the Programme for International Student Assessment, *Scandinavian Journal of Educational Research*. doi: https://doi.org/10.1080/00313831.2016.1258726

Idil, F. H., Narli, S., & Aksoy, E. (2016). Using data mining techniques examination of the middle school student attitude towards mathematics in the context of some variables. *International Journal of Education in Mathematics, Science and Technology, 4* (3), 210-228. doi: http://dx.doi.org/10.18404/ijemst.02496.

Instituto de Evaluación (IE) (2007). *PISA 2006. Informe Español*. Madrid: MEC.

Izenman, A. J. (2008). *Modern multivariate statistical techniques. Regression, classification, and manifold learning*. New York: Springer.

Kaplan, D. & Su, D. (2016). On Matrix Sampling and Imputation of Context Questionnaires with Implications for the Generation of Plausible Values in Large-Scale Assessments. *Journal of Educational and Behavioral Statistics*, *41* (1), 57-80.

Kass, G. V. (1980). An exploratory technique for investigating large quantities of categorical data. *Applied Statistics, 29,* 119-127.

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 (JESEH)*, *1* (1), 28-48.

Lakshmipriya, K. & Arunesh, P. K. (2017). Predicting student performance using data mining classification techniques. *International Journal of Innovative Research in Science and Engineering*, *3* (2), 54-60.

Lizasoain, L. (2012). Las técnicas de minería de datos aplicadas a la investigación educativa. Arboles estadísticos de decisión. In M. Castro (Ed.), *Elogio a la Pedagogía Científica* (pp. 101-121). Madrid: Grafidridma.

Martínez Arias, R. (2006). La metodología de los estudios PISA. *Revista de Educación*, extraordinary number 2006, 111-129.

Muñoz Ledesma, D. (2015). *Modelos para la mejora del rendimiento académico de alumnos de la E.S.O. mediante técnicas de minería de datos* (Tesis doctoral). Universidad de Murcia.

OECD (2014). *PISA 2012 Technical Report*. Paris: OECD. Retrieved from https://www.oecd.org/pisa/pisaproducts/PISA%202012%20Technical%20Report_Chapter%2019.pdf

OECD (2016). *PISA 2015 Context Questionnaires Framework in PISA 2015 Assessment and Analytical Framework: Science, Reading, Mathematic and Financial Literacy*. Paris: OECD Publishing, 101-127. doi: http://dx.doi.org/10.1787/9789264255425-7-en

Pereira, D., Perales, M. J., & Bakieva, M. (2016). Análisis de tendencias en las investigaciones realizadas a partir de los datos del Proyecto PISA. *RELIEVE*, *22* (1), 1-18. doi: http://dx.doi.org/10.7203/relieve.22.1.8248

Ruby, J. & David, K. (2015). Analysis of Influencing Factors in Predicting Students Performance Using MLP. A Comparative Study. *International Journal of Innovative Research in Computer and Communication Engineering*, *3* (2), 1085-1092.

Rutkowski, L. & Rutkowski, D. (2010). Getting it 'better'. The importance of improving background questionnaires in international large-scale assessment. *Journal of Curriculum Studies, 42* (3), 411-430. doi: http://dx.doi.org/10.1080/00220272.2010.487546

Rutkowski, L. & Rutkowski, D. (2017). Improving the comparability and local usefulness of international assessments: A look back and a way forward. *Scandinavian Journal of Educational Research,* 1-14. doi: http://dx.doi.org/10.1080/00313831.2016.1261044

Santín, D. (2006). La medición de la eficiencia de las escuelas: una revisión crítica. *Hacienda Pública Española / Revista de Economía Pública*, *177* (2/2006), 57-82.

Streifer, P. A. & Schumann, J. A. (2005). Using data mining to identify actionable information: breaking new ground in data-driven decision making. *Journal of Education for Students Placed at Risk (JESPAR)*, *10*, 281-293.

Strobl, C., Malley, J., & Tutz, G. (2009). An Introduction to Recursive Partitioning: Rationale, Application, and Characteristics of Classification and Regression Trees, Bagging, and Random Forests. *Psychological Methods*, *14* (4), 323-348.

Taut, S. & Palacios, D. (2016). Interpretaciones no intencionadas e intencionadas y usos de los resultados de PISA: Una perspectiva de validez consecuencial. *RELIEVE, 22* (1), M8. doi: http://dx.doi.org/10.7203/relieve22.1.8294

Thai Nghe, N., Janecek, P., & Haddawy, P. (2007). A Comparative Analysis of Techniques for Predicting Academic Performance (Conference Paper, Session T2G). In *37th ASEE/IEEE Frontiers in Education Conference*. Milwaukee, WI.

Thakar, P., Mehta, A., & Manisha (2015). Performance Analysis and Prediction in Educational Data Mining: A Research Travelogue. *International Journal of Computer Applications, 110* (15), 60-68.

Yu, C. H., Kaprolet, C., Jannasch-Pennell, A., & Digangi. S. (2012). A Data Mining Approach to Comparing American and Canadian Grade 10 Students' PISA Science Test Performance. *Journal of Data Science, 10*, 441-464.

Authors' biographies

Inmaculada Asensio Muñoz has a PhD in Pedagogy from the Universidad Complutense of Madrid with special doctoral prize. Lecturer in the Department of Educational Research Methods and Assessment in the Faculty of Education of the Universidad Complutense of Madrid, and member of the Measuring and Evaluating Educational Systems research group.

Elvira Carpintero Molina has a PhD in Educational Psychology from the Universidad Complutense of Madrid and Assistant Professor in the Department of Educational Research Methods and Assessment. Member of the Measuring and Evaluating Educational Systems research group and the Adaptive Pedagogy research group at the Universidad Complutense of Madrid.

Eva Expósito Casas has a PhD in Education from the Universidad Complutense of Madrid and Assistant Professor in the Department of Research Methods and Assessment in Education II at the Universidad Nacional de Educación a Distancia. She is a member of the Complutense's Measuring and Evaluating Educational Systems research group (MESE) and the Educational Psychology Counselling and Counsellor Skills research group (GRISOP).