A top-down evaluation of factors related to the largest performance gap in reading literacy across 25 countries

Fabio Alivernini, National Institute for the Educational Evaluation of Instruction and Training (INVALSI), Italy, fabio.alivernini@invalsi.it

Abstract

The aim of the present secondary study, based on data from 25 countries, is to identify the pattern of variables (at country, school and student levels), which are typical of students performing below the Low International Benchmark compared to students performing at the Advanced Performance Benchmark, in PIRLS 2006. The dependent variable of the analysis is a dichotomous variable, the values of which represent the two different performance groups of students. The independent variables are two sets of OECD educational indicators, variables from PIRLS 2006 Reading Curriculum Questionnaire data, variables and indices based on data obtained from questionnaires for teachers, schools, parents and students. The analysis is based on classification and regression trees (CART), which is a full hierarchical non-parametric method suited to detecting and interpreting complex reciprocal influences between a large number of independent variables. Results show that pupils' performance is predicted by an interaction between country level variables such as the changes in teachers' salaries at the top of the salary scale, school context variables (e.g. percentages of students from economically affluent home) and home educational resources and variables at student level (e.g. students' reading self-concept).

Keywords: *reading literacy, classification trees, PIRLS 2006, OECD indicators, contextual factors.*

Introduction

The PIRLS 2006 benchmarks denote the range of performance of students internationally. The scores of students are divided into four performance levels, with the Advanced International Benchmark at the top and the Low International Benchmark at the bottom (Mullis, Martin, Kennedy, & Foy, 2007). Students who have reached the Advanced International Benchmark in reading comprehension can, for instance, assimilate various ideas within a text in order to provide interpretations of a fictional character's traits or behaviour, whereas students who have not even reached the Low International Benchmark cannot individuate or recognize an explicitly stated detail in a text.

The aim of the present study, based on data from 25 countries, was to identify the pattern of

variables (at country, school and student levels), which are typical of students performing below the Low International Benchmark compared to students performing at the Advanced Performance Benchmark.

The theoretical framework of the study was consistent with the PIRLS 2006 Assessment Framework (Mullis, Kennedy, Martin, & Sainsbury, 2006), which considers a large number of contextual factors influencing students' performance in reading. One should also note that the theoretical framework does not identify a precise pattern of interactions between these variables and an explorative approach is therefore appropriate. It is also important to point out that the explorative objective of the present paper includes the identification of interaction effects, which are not completely predictable *a priori*, among a large number of variables of different types (Nominal, Ordinal, Interval).

Methodology

The analysis was based on classification and regression trees (CART) (Williams, Lee, Fisher, & Dickerman, 1999), a method suited to detecting and interpreting complex interactions in large data sets that most traditional means of regression and classification analysis might ignore or find difficult to estimate and interpret (Allore, Tinetti, Araujo, Hardy, & Peduzzi, 2005). In fact tree methods can discover interactions during the growth of the tree whereas traditional regression techniques require *a priori* specification of interactions. It is important to note that CART is not affected by problems of multi-collinearity between predictors and it is a truly non-parametric method, since it makes no assumptions regarding the underlying distribution from which the subjects are sampled. In addition since classification trees derive their predictions from a few "if-then" conditions, it is possible a straightforward interpretation of results in terms of the distinctive characteristics of the different groups.

The CART algorithm proceeds by performing successive binary divisions of the subjects on the basis of a statistical criterion. Starting from the full sample (called root node or parent node) each independent variable is evaluated on the basis of the extent to which it is able to reduce the impurity of the parent node by dividing the subjects into two groups (called child nodes). The impurity consists in the degree to which the students at a node vary compared with the dependent variable: a minor impurity indicates a greater homogeneity of the subjects for the values of the dependent variable. In the case under examination a completely pure node would be one which includes students who only belong to the highest levels or only to the lowest levels of skills. A very widespread measure of the purity of a node is the Gini index which is calculated by the following

$$G = 1 - \sum_j p_j^2$$

in which p_j is the proportion of cases in class j.

When a node is completely pure (i.e. when all the cases belong to a single class of the dependent variable), G = 0. CART performs an exhaustive search for the independent variable in order to find the split point that produces the highest reduction in the impurity of a node. The reduction is calculated by comparing the purity of the root node with the sum of the impurities of the child nodes. Independent variables can be nominal, ordinal or continuous. If X, the independent variable is nominal, each possible subset of categories of X is examined to find the best split. The independent variable that produces the highest reduction in the impurity is selected for the first splitting of the sample. Subsequently each resulting node is split by means of the same procedure and, in the continuation of the partition process, the students are progressively classified into smaller groups. The process continues until either the reduction in the impurity becomes less than a predetermined criterion (G = 0.02 in this study; Breiman, Friedman, Olshen, & Stone, 1984), or the number of individuals in a group produced by a partition is lower than a certain threshold. The end result of the process of partition is a tree organized in a hierarchical way, in which the root is the overall sample of individuals, the branches are the values of independent variables used in the analysis and the nodes are the subsets of individuals identified by some combination of values of the independent variables. If the nodes are located in a terminal position (i.e. they are not further split) they are called *leaves*. Each individual is classified by following a pathway along the tree, leading from the root to a leaf.

In the present study, the CART analysis was conducted using a hierarchical approach (Hox, 2002; Fabbris, 1997) in three stages:

1. with only country level variables included in the model;

2. with school and teacher level variables nested under the country model identified at stage 1;

3. with parent and student variables nested under the country and school model identified at stage 2.

Initially, we developed a model based only on those contextual variables which are generally considered to be logical or temporal antecedents of reading literacy (e.g. socio-economic and cultural variables, resources, the child's past abilities). Subsequently, the obtained model was integrated by analyzing potentially modulating variables at school level (e.g. teachers' satisfaction) and at student level (e.g. student's reading self concept).

The classification model was developed on the basis of a random subset of the data (training sample) and then the results were validated on a separate random sample (test sample). The accuracy model was estimated using cross validation techniques (Breinman et al., 1984).

The student sample was weighted according to the indications of the PIRLS 2006 Technical Report (Martin, Mullis, & Kennedy, 2007). The CART algorithm used is the one implemented in the SPSS Clementine software.

Sample

Data is based on the answers of pupils performing below the Low International Benchmark and performing at the Advanced Performance Benchmark, as well as those of teachers and principals who took part in the PIRLS 2006 in 25 countries (the PIRLS 2006 countries, which are also OECD countries). The data from the PIRLS 2006 Curriculum Questionnaire and information on some OECD educational indicators (OECD, 2008) were also taken into consideration. Descriptive statistics showed a large number of missing values on OECD educational indicators for Canada, so this data was excluded from subsequent analysis. The final data refers to 11,598 students.

Variables

The dependent variable of the analysis was a dichotomous variable, the values of which correspond to the two different groups of fourth grade students (i.e. pupils performing at the Advanced International Benchmark and pupils performing at the Low International Benchmark).

The independent variables at the country level were:

1. OECD educational indicators (OECD, 2008) related to financial and human resources invested in primary education and related to the learning environment and organization of primary schools. In Table 1 and Table 2 a complete list of the OECD educational indicators analyzed is presented. In the analysis all of the OECD variables were considered as ordinal.

2. The Countries' Gross Domestic Product in 2006 (considered as an ordinal variable).

3. Variables from the PIRLS 2006 Curriculum Questionnaire (Table 3). In the analysis all of the variables from the curriculum questionnaire were considered as nominal.

4. Variables from PIRLS 2006 Student and Parents Questionnaire aggregated (median) at the country level referring to cultural status (Index of Home Educational Resources).

[Please, insert Table 1, 2 and 3 about here]

The independent variables at the school level were:

1. Variables and indices from the PIRLS 2006 School and Teacher Questionnaire (Table 4 and Table 5).

2. Variables from PIRLS 2006 Student and Parents Questionnaire aggregated (median) at the school level referring to cultural status (Index of Home Educational Resources).

[Please, insert Table 4 and 5 about here]

The independent variables at the student level were from the PIRLS 2006 Parents and Student Questionnaire (Table 6 and Table 7).

[Please, insert Table 6 and 7 about here]

Finding and Discussion

Figure 1 shows the final tree (potentially modulating variables included) produced by CART in order to identify the segments of students with the greatest disparities in their reading proficiency levels.

[Please, insert Figure 1 about here]

The results show that the country level variable that makes the most significant difference (first split: node 1 and node 2) as regards pupils' performance groups is the changes in primary school teachers' salaries between 1996 and 2006. In particular this indicator refers to the changes occurring in a 10-year period in the salaries of teachers with over 15 years of experience and the minimum level of qualifications required for teaching in primary school. From an interpretative point of view it should be noted that:

a) this is a measure of the salary of teachers at a midpoint of their careers,

b) this is a measure of change over time.

Both the first and the second point suggest that this variable is potentially related to the level of motivation and satisfaction of teachers rather than being simply a measure of welfare. Indeed in this respect one can suppose that if higher salaries in the early or later stages of teachers' careers favour the choice of teaching as a profession over other jobs (OECD, 2008), the improvement of wages at an intermediate point of the career is more closely linked to the experience of the teacher while working in a school.

The analysis conducted by means of CART showed that factors associated with the school, the family and the student interact in a more complex way than the traditional methods of data analysis such as regression are usually able to reveal. Knowledge of these interactions sheds new light on the differences that at various levels are associated with the largest profit gap detected in PIRLS 2006. The combination of several variables underlies the identification of student groups with particular concentrations of "best performers" and "very low performers". The profile of these groups is as follows:

Groups with a high proportion of students performing at the lowest level

- Students who live in countries where changes in teachers' salaries at an intermediate level of career grew less than 109.7%. Here the group with the highest concentration of lowest performers (93.2%) corresponds to Node 16. These are students who attend schools where few pupils come from economically affluent homes (less than 10%). According to the Parents Questionnaire these students had a lower capacity than excellent students to recognize letters (Child's ability to recognize letters less than "very well") when they began primary school. At present they appear to have a low or medium reading self-concept.

- Students who live in countries where changes in teachers' salaries at an intermediate level of career grew more than 109.7%. Here the group with the highest concentration of lowest performers (73.1%) corresponds to Node 12. These are students who attend schools where few pupils are from economically affluent homes (less than 10%). As regards their families, the parents' highest occupational level corresponds to the following categories: skilled worker; general labourer; never worked outside the home for pay.

Groups with a high proportion of students performing at the highest level

- Students who live in countries where changes in teachers' salaries at an intermediate level of career grew less than 109.7%. Here the group with the highest concentration of highest performers (74.9%) corresponds to Node 10. This group occurs in schools with a percentage of more than 10% of students from economically affluent homes, living in families with average or above-average availability of educational resources (more than 25 books at home).

- Students who live in countries where changes in teachers' salaries at an intermediate level of career grew by at least 109.7%. Here the group with the highest concentration of highest performers (95%) corresponds to Node 19. These are students who attend schools with a percentage of more than 10% of pupils from economically affluent homes. Furthermore, these students have average or above-average educational resources at home

(more than 25 books at home). At present they also appear to have a high reading self-concept.

Conclusion and Implications

The analysis carried out in this study used an exploratory and hierarchical approach. In other words *a priori* assumptions or hypotheses were not tested, but a data driven model which took the multilevel structure of PIRLS 2006 data into consideration was developed. In addition a methodology was used which aimed to identify relationships between variables upon a substantial basis and not on the basis of their statistical significance. Indeed this latter criterion, when the number of subjects is very high, as was the case in this study, may lead to misleading conclusions that are more related to the statistical power of the analysis (Cohen, 1988) than to the existence of relevant relationships. The classification model was developed on a training sample and tested on a control sample, and a cross-validation was carried out with 25 independent samples.

The results showed that factors associated with the country's national educational policies, the school, the family and the student interact in a very complex way. For instance if we look at the right side of the classification tree, in countries with higher increases in teacher salaries (more than 109.7%), in schools with an economically affluent context, and in families with at least some educational resources (more than 25 books at home), we can find both reading excellence (node 19), and a substantial lack of expertise in reading (node 20). This depends on the student's level of reading self-concept: if it is high (node 19) a group with 95% performing at the Advanced Performance Benchmark is identified. Otherwise, if the student's reading self-concept level is low or medium (node 20) this percentage goes down to 46.4%.

On the left side of the tree, in countries with lower increases in teacher salaries (less than 109.7%) and in schools with an economically disadvantaged context, a child's ability to recognize letters when he/she began primary school ("very well" versus "moderately well", "not very well", and "not at all") makes an important difference to the probability of him/her reaching the Advanced International Benchmark in reading comprehension (node 7) or of not even being able to individuate an explicitly stated detail in a text (node 8).

In summary, this study provided a complete and detailed description of pupils with large reading proficiency disparities that is both informative and readily interpretable. The results are presented in a familiar question-and-answer format, providing additional insights by pointing out the order in which different questions should be asked. The study reveals the presence of different groups among students who have the highest or the lowest reading proficiency levels in PIRLS 2006 across a large number of countries. Each group has its own

peculiarities, which it may be useful to be aware of in order to plan and implement interventions and measures aiming to improve the performance of these students and to bridge the most significant gaps between them. In conclusion one should note that this study did not aim to explain the specific reasons why the identified phenomena occur. Instead, rather than speculating on the causes of the phenomena, it provides a rich and reliable description of the characteristics of these phenomena. This description is especially useful in that it suggests a number of issues for further investigations and research.

References

Allore, H., Tinetti, M. E., Araujo, K. L. B., Hardy, S., & Peduzzi, P. (2005). A case study found that a regression tree outperformed multiple linear regression in predicting the relationship between impairments and Social and Productive Activities scores. *Journal of Clinical Epidemiology, 58*, 154–161.

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees.* Belmount, CA: Wadsworth.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (2nd edition).* Hillsdale, NJ: Erlbaum.

Fabbris, L. (1997). Statistica *multivariata: analisi esplorativa dei dati.* Milano: McGraw Hill.

Hox, J. (2002). *Multilevel analysis: Techniques and applications.* Mahwah, NJ: Erbaum.

Martin, M.O., Mullis, I.V.S., & Kennedy, A.M. (2007). *PIRLS 2006 Technical Report.* Chestnut Hill, MA: TIMSS & PIRLS International Study Center Lynch School of Education, Boston College.

Mullis, I.V.S., Kennedy, A.M., Martin, M.O., & Sainsbury, M. (2006). *PIRLS 2006 Assessment Framework and specifications (2nd ed.).* Chestnut Hill, MA: TIMSS & PIRLS International Study Center Lynch School of Education, Boston College.

Mullis, I.V.S., Martin, M.O., Kennedy, A.M., & Foy P. (2007). *PIRLS 2006 International Report: IEA's Progress in International Reading Literacy Study in Primary School on 40 countries*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center Lynch School of Education, Boston College

Organisation for Economic Co-operation and Development. (2008). *Education at a glance 2008.* Paris: OECD.

Williams, C. J., Lee, S. S., Fisher, R. A. & Dickerman, L. H. (1999). A comparison of statistical methods for prenatal screening for downsyndrome. Applied Stochastic Models in Business and Industry, 15, 89-101.

Table 1

OECD educational indicators for the data analyzed - Financial resources invested in primary education

Financial resources invested in primary education	**Min**	**Max**	**Median**
Annual expenditure on primary schools per student for all services (2005) [a]	2806	14079	7532,3
Annual expenditure on primary schools per student for all services relative to GDP per capita (2005)	16	26	22
Current expenditure on primary schools (2005) as percentage of total expenditure on primary schools (2005) [b]	75,6	97,2	90,5
Capital expenditure on primary schools (2005) as percentage of total expenditure on primary schools (2005) [c]	2,8	24,4	9,5
Percentage of current expenditure on primary school (2005) for teachers' compensation	51,0	74,2	55,1
Percentage of current expenditure on primary school (2005) for other staff's compensation	10,6	27,5	25,8
Percentage of current expenditure on primary school (2005) for all staff's compensation	66,7	89,6	80,8
Percentage of other current expenditure on primary school (2005)	10,4	33,3	19,2

a Expressed in equivalent USD converted using PPPs
b expenditure on resources used each year
c Expenditure on assets that last longer than one year

Table 2

OECD educational indicators for the data analyzed - Primary school teachers salaries

Primary school teachers' salaries	**Min**	**Max**	**Median**
Primary school teachers' salaries (2006) at starting salary - Full time teachers with the minimum training necessary to be fully qualified [a]	11788	50301	34894,9
Primary school teachers' salaries (2006) after 15 years of experience - Full time teachers with the minimum training necessary to be fully qualified [a]	14976	69269	42403,6
Primary school teachers' salaries (2006) at the top of the salary scale - Full time teachers with the minimum training necessary to be fully qualified [a]	19839	102519	43057,6
Change between 1996 and 2006 in primary school teachers' salaries at starting salary - Full time teachers with the minimum training necessary to be fully qualified [b]	95	209	119,9
Change between 1996 and 2006 in primary school teachers' salaries after 15 years of experience - Full time teachers with the minimum training necessary to be fully qualified [b]	95	196	107,2
Change between 1996 and 2006 in primary school teachers' salaries after 15 years of experience and at the top of the salary scale - Full time teachers with the minimum training necessary to be fully qualified [b]	92	201	107,2

a Expressed in equivalent USD converted using PPPs

b Converted to 2006 price level using GDP deflators (1996=100)

Table 3

Variables from the PIRLS 2006 Curriculum Questionnaire

Variables	**Possible values**
Presence of national reading curriculum	Yes/No
Presence of local authorities' role in curriculum development	Yes/No
Revision of curriculum	Yes/No
Reading in curriculum	Presented as a separate curriculum area/Presented as part of the curriculum for language instruction
Reading curriculum's prescription	Goals and objectives (Yes/No) [a]; Processes or methods; Materials; Other
How the issue of students with different levels of ability is addressed by reading curriculum	The same curriculum is prescribed for all students/The same curriculum is prescribed for students of different ability levels, but at different levels of difficulty/Different curricula are prescribed for students of different ability levels
Reading curriculum forms available	Official publication containing the curriculum (Yes/No); Ministry notes and directives; Mandated or recommended textbooks; Instructional or pedagogical guide; Specifically development or recommended instructional activities; Other educational authorities
Current requirements for being a primary teacher	Pre-practicum before teacher education program; Pre-practicum during teacher education program; Supervised practicum in the field; Passing an examination; Completion of a probationary teaching period; Completion of a mentoring or induction program; Other
Primary teachers certifiers	Ministry of education (Yes/No); National or state licensing board; Universities; Teacher organization; Other
Teachers receive specific preparation in how to teach curriculum as part of pre-service education	Yes/No
Methods used to help teachers to implement the reading curriculum	In-service training (Yes/No); Expert teacher; Reduced teaching load for new teachers; Other
Methods used to evaluate the reading curriculum implementation	Visits by inspectors (Yes/No); Research programs; School self-evaluation; National or regional assessments; Other

a When the same values apply to each option of a variable, these values are indicated only near the first option of the variable, without repeating them for each of the next options

Table 4

Variables from the PIRLS 2006 School Questionnaire [a]

Indices and variables	Possible values
Availability of School Resources (ASR) - [Ordinal]	High/Medium/Low
Parents' Perceptions of School Environment (PPSE) - [Ordinal]	High/Medium/Low
Student Safety in School (SSS) - [Ordinal]	High/Medium/Low
Principals' Perception of School Safety (PPSS) - [Ordinal]	High/Medium/Low
Principal's Perception of School Climate (PPSC) - [Ordinal]	High/Medium/Low
Home-School Involvement (HSI) - [Ordinal]	High/Medium/Low
Schools' locations - [Nominal]	Urban/Sub-Urban/Rural
Students coming from economically affluent homes - [Ordinal]	0–10%/11–25%/26–50%/More than 50%
Availability of computers for instruction - [Ordinal]	Fewer than 5 students/5-10 students/11-20 students/More than 20 students/Schools without computers
Seriousness of absenteeism in school- [Ordinal]	Not a Problem/Minor Problem/Moderate Problem/Serious Problem.

a In square brackets the level of measurement defined for the variable in the analysis

Table 5

Variables from the PIRLS 2006 Teacher Questionnaire [a]

Indices and variables	Possible values
Reading for Homework (RFH) - [Ordinal]	High/Medium/Low
Teacher Career Satisfaction (TSC) - [Ordinal]	High/Medium/Low
Organization of students for reading instruction - [Ordinal]	Teaching reading as a whole-class activity (Always or almost always/Often/Sometimes/Never) [b]; Creating same-ability groups; Creating mixed-ability groups; Using individualized instruction for reading; Having students work independently on an assigned plan or goal; Having students work independently on an a goal they choose themselves
Materials used by teachers for reading instruction - [Ordinal]	Textbooks (Every day or almost every day/Once or twice a week/Once or twice a month/Never or almost never); Reading series; Workbooks or worksheets; Children's newspapers and/or magazines; Computer software for reading instruction; Reading material on the Internet; A variety of children's books; Materials from other subjects; Materials written by students
The use of instructional materials for students at different reading levels - [Nominal]	The same materials with all students because all students are at the same reading level/The same materials with students at different reading levels - students work at different speeds/The same materials with all students regardless of reading level - students work at the same speed/Different materials with students at different reading levels
Reading by teachers and student in the classroom - [Ordinal]	Read aloud to the class (Every day or almost every day/Once or twice a week/Once or twice a month/Never or almost never); Students read aloud to the whole class; Students read aloud in small groups or pairs; Students read silently on their own; Students read along silently while other students read; Give students time to read books of their own choosing; Teach or model for students different reading strategies; Teach students strategies for decoding sounds and words; Teach students new vocabulary; Help students understand new vocabulary
Reading comprehension activities - [Ordinal]	Answer reading comprehension questions in a workbook or on a worksheet (Every day or almost every day/Once or twice a week/Once or twice a month/Never or almost never); Write something about what they have read; Answer oral questions or give oral summary; Talk with each other; Do a project; Take a written quiz or test
The use of reading comprehension skills or strategies - [Ordinal]	Identify the main ideas (Every day or almost every day/Once or twice a week/Once or twice a month/Never or almost never); Explain or support their understanding; Compare reading with own experience; Compare with other things read; Make prediction about what will happen next; Make generalizations and draw inferences; Describe text style or structure
Emphasis on sources to monitor progress in reading- [Ordinal]	Diagnostic tests (Major emphasis/Some emphasis/Little or no emphasis); Classroom tests; National or Regional achievement tests; Teacher's own professional judgment

Approaches to assessing students' performance in reading - [Ordinal]	Multiple-choice questions (At least once a week/Once or twice a month/Once or twice a year/Never); Short-answer written questions; Paragraph-length written responses; Listening to students read aloud; Oral questioning; Students give an oral summary; Meeting with students to discuss what they have been reading
Teachers' Age - [Ordinal]	Under 25/25-29/30-39/40-49/50-59/60 or more
Teachers' Gender - [Nominal]	Female/Male
Teachers' highest level of formal education - [Nominal]	Did not complete ISCED Level 3/Finished ISCED Level 3/Finished ISCED Level 4/Finished ISCED Level 5B/Finished ISCED Level 5A first degree/ Finished ISCED Level 5A second degree or higher
Certified teacher- [Nominal]	Yes/No
Hours spent by teachers in workshops or seminars that dealt directly with reading or teaching reading - [Ordinal]	None/Less than 6 hours/6-15 hours/16-35 hours/More than 35 hours
Teachers reading for enjoyment - [Ordinal]	Every day or almost every day/Once or twice a week/Once or twice a month/Never or almost never

a In square brackets the level of measurement defined for the variable in the analysis.
b When the same values apply to each option of a variable, these values are indicated only near the first option of the variable, without repeating them for each of the next options.

Table 6

Variables from the PIRLS 2006 Parents Questionnaire [a]

Indices and variables	Possible values
Parents' Perceptions of School Environment (PPSE) - [Ordinal]	High/Medium/Low
Parents' Attitudes Toward Reading (PATR) - [Ordinal]	High/Medium/Low
Home Educational Resources (HER) - [Ordinal]	High/Medium/Low
Early Home Literacy Activities (HELA) - [Ordinal]	High/Medium/Low
How well students could perform beginning literacy activities when they entered school - [Ordinal]	Recognize most of the alphabet (Very well/Moderately well/Not very well/Not at all) [b]; Read some words; Read sentences; Write letters of the alphabet; Write some words
Number of books at home - [Ordinal]	0-10/11-25/26-100/101-200/More than 200
Number of children's books at home - [Ordinal]	0-10/11-25/26-50/51-100/More than 100
Parents' highest level of education - [Ordinal]	Some ISCED Level 1 or 2, or did not go to school/ISCED Level 2/ISCED Level 3/ISCED Level 4/ISCED Level 5B/ISCED Level 5A first degree/Beyond ISCED Level 5A first degree/not applicable
Parents' highest occupation level - [Nominal]	Never worked outside the home for pay/Small business owner/Clerk/Service or sales worker/Skilled agricultural of fishery worker/Craft or trade worker/Plant or machine operator/General laborer/Corporate manager or senior official/ Professional/Technician or associate professional/Not applicable)

a In square brackets the level of measurement defined for the variable in the analysis.
b When the same values apply to each option of a variable, these values are indicated only near the first option of the variable, without repeating them for each of the next options.

Table 7

Variables from the PIRLS 2006 Student Questionnaire [a]

Indices and variables	Possible values
Student Safety in School (SSS) - [Ordinal]	High/Medium/Low
Students' Reading Attitudes (SATR) - [Ordinal]	High/Medium/Low
Students' Reading Self Concept (SRSC) - [Ordinal]	High/Medium/Low
Students' Gender - [Nominal]	Girl/Boy
Student born in the country - [Nominal]	Yes/No
Mother born in the country - [Nominal]	Yes/No/I don't know
Father born in the country - [Nominal]	Yes/No/I don't know

a In square brackets the level of measurement defined for the variable in the analysis.

Figure 1. *Classification tree - Final model. Bold character is used for category which increases compared to the proportions of the precedent node. Below L.= students performing below the Low International Benchmark; Advanced = students performing at the Advanced Performance Benchmark*