Crawford School of Public Policy

CAMA

Centre for Applied Macroeconomic Analysis

Why East Asian Students Perform Better in Mathematics than Their Peers: An Investigation Using a Machine Learning Approach

CAMA Working Paper 66/2021
July 2021

Hanol Lee
Southwestern University of Finance and Economics

Jong-Wha Lee
Korea University
Centre for Applied Macroeconomic Analysis, ANU

Abstract

Using a machine learning approach, we attempt to identify the school-, student-, and country-related factors that predict East Asian students' higher PISA mathematics scores compared to their international peers. We identify student- and school-related factors, such as metacognition–assess credibility, mathematics learning time, early childhood education and care, grade repetition, school type and size, class size, and student behavior hindering learning, as important predictors of the higher average mathematics scores of East Asian students. Moreover, country-level factors, such as the proportion of youth not in education, training, or employment and the number of R&D researchers, are also found to have high predicting power. The results also highlight the nonlinear and complex relationships between educational inputs and outcomes.

Keywords

education, East Asia, machine learning, mathematics test score, PISA

JEL Classification

C53, C55, I21, J24, O1

Address for correspondence:

(E) cama.admin@anu.edu.au

ISSN 2206-0332

Why East Asian Students Perform Better in Mathematics than Their Peers: An Investigation Using a Machine Learning Approach*

Hanol Lee†
Southwestern University of Finance and Economics

and

Jong-Wha Lee‡
Korea University and Centre for Applied Macroeconomic Analysis, ANU

July 2021

Abstract

Using a machine learning approach, we attempt to identify the school-, student-, and country-related factors that predict East Asian students' higher PISA mathematics scores compared to their international peers. We identify student- and school-related factors, such as meta cognition–assess credibility, mathematics learning time, early childhood education and care, grade repetition, school type and size, class size, and student behavior hindering learning, as important predictors of the higher average mathematics scores of East Asian students. Moreover, country-level factors, such as the proportion of youth not in education, training, or employment and the number of R&D researchers, are also found to have high predicting power. The results also highlight the nonlinear and complex relationships between educational inputs and outcomes.

JEL classification: C53, C55, I21, J24, O15

Keywords: education, East Asia, machine learning, mathematics test score, PISA

* We would like to thank Noam Angrist, Wei Xiao, and Jingjing Ye for their helpful discussions and suggestions.

† Research Institute of Economics and Management, Southwestern University of Finance and Economics, Chengdu, Sichuan, People's Republic of China, 611130. E-mail: hanollee@swufe.edu.cn.

‡*Send correspondence:* Economics Department, Korea University and Centre for Applied Macroeconomic Analysis (CAMA), Australian National University. E-mail: jongwha@korea.ac.kr

1. Introduction

Over the years, East Asian students have consistently performed well in international achievement tests, especially in mathematics and science. In the recent Programme for International Student Assessment (PISA) and Third International Mathematics and Science Study (TIMSS), East Asian countries (regions) assumed the top positions in the rankings (OECD, 2019; Mullis *et a*l., 2020). Higher test scores are considered to reflect higher levels of educational quality and human capital, which significantly and positively impact workers' labor market outcomes, as well as a nation's economic growth and human development (Hanushek and Kimko, 2000; Lee and Barro, 2001; Barro and Lee, 2015; Angrist et al, 2021). Hanushek and Woessmann (2016) asserted that the better student performance in average math and science test scores, was a major factor accounting for the strong growth of East Asian economies.

The purpose of this study is to explore the possible reasons for the high performance of East Asian students in international mathematics tests using machine learning (ML) techniques. We investigate factors relating to country, school, and student characteristics that have high relevance in predicting the higher mathematics test scores of secondary East Asian students, focusing on the differences between East Asian countries (regions) and non-East Asian countries. To address this issue, we use a state-of-the-art ensemble method that combines several base ML models and produces an optimal predictive model using gradient boosting algorithm (Dietterich, 2000; Mullainathan and Spiess, 2017). The ML algorithms can fit very flexible functional forms, such as nonlinear interactions and discontinuous relationships among input and outcome variables. By applying the ML method to the PISA 2018 mathematics test scores across 76 countries, including seven East Asian countries (regions), we construct a prediction model and identify the important variables that explain the differences in students' mathematics achievements across all countries, as well as the academic achievement gap between East Asian and non-East Asian students. We not only assess the extent to which student and school background variables contribute to high mathematics scores, but also how a country's socioeconomic factors contribute to the higher average score of East Asian students. As per our knowledge, this is the first study that investigates the possible reasons for the excellent performance of East Asian students in international mathematics tests using a ML technique.

The high academic performance of East Asian countries has raised strong interest among researchers, policy makers, and the public. Thus, there have been many studies that have searched for the factors that influence East Asian children’s high achievements in mathematics and science. A number of student-, family-, and school-related factors, as well as country-specific characteristics, have been pointed out as possible reasons for East Asian students’ strong performances.

Earlier works indicate that the superior mathematics achievement of East Asian students could be attributed to home and school life, rather than intellectual abilities (Stevenson et al., 1985; Stevenson et al., 1986). East Asian pupils appear to place more value on academic achievement in general, and believe more strongly in the value of hard work than their peers in many other countries (Hess et al., 1987; Valverde and Schmidt, 2000; Leung, et al., 2015). East Asian students tend to spend more time per day on mathematics in school and homework, than students in Western countries. Some researchers also argued that a family’s strong interest in and support of a child's education, has a positive impact on student achievement (Francis and Archer, 2005; Kim, 2021).

However, other researchers claim that the outstanding intellectual abilities of East Asian students have to do with their individual success in mathematics, regardless of their efforts or parents’ support (Lynn, 1982; Ho and Hau, 2008). A few studies suggested that the counting systems of East Asian languages, which facilitate number representation, are favorable to students’ numerical abilities (Miura et al., 1988; Fuson and Kwon, 1991; Dowker and Li, 2019). For instance, some scholars pointed out that certain East Asian languages, such as Chinese and Korean, directly name the values of ten and one in two-digit numbers, while the English language is more irregular and opaquer.

Many scholars have also focused on teacher- and school-related aspects. Some emphasized that East Asian mathematics teachers have greater subject knowledge and provide clearer instruction, as supported by continuous professional training programs (Ma, 1999; Leung, 2001; Jerrim, and Vignoles, 2016). Several researchers have also suggested that in East Asian countries, the standards for math and science subjects are more consistent and focused, which can be related to their more centralized school systems (Valverde and Schmidt, 2000; Kim, 2005; Kaya and Rice, 2010). Some researchers have pointed out that the Confucian culture is an East Asian identity which has greatly influenced mathematics education and achievement (Biggs 1996; Leung 2001; Leung, et al., 2015; Wang and Lin, 2015).

Confucianism emphasizes that learning is key to personal success and national development. In East Asia, the teaching profession is highly respected and attracts highly qualified people.

Contrary to previous studies that adopt case studies and conventional statistical methods, this study adopts ML techniques to identify the factors that are closely associated with the mathematics achievement of East Asian students. The distinct difference between a ML approach and a statistical approach is that while the latter assumes an appropriate data model and uses the data to estimate parameters, the former avoids beginning with the model and instead uses an algorithm to learn about the relationship between the predictors and the responses (Mullainathan and Spiess, 2017; Athey and Imbens, 2019). Since a ML approach identifies the dominant patterns between the input and response data without the imposition of functional forms *a priori* by the researcher, it can overcome common specification errors. The ML algorithms can fit complex and very flexible functional forms to the data without overfitting, and thus can search for functions that are good predictors out of the sample. The ML models also have another merit of avoiding the multicollinearity problem that arises from correlated covariates (Sandri and Zuccolotto, 2008). Hence, the ML techniques are useful enough to identify important school, student, and country-related factors in explaining high PISA math scores for East Asian students.

Over the past decade, many studies have adopted ML techniques to identify factors predicting student performance (Masci et al., 2018). They considered it inappropriate to presume a specific functional form *a priori*, and opted for a flexible model that does not impose any functional relationships between covariates (school resources) and educational outcomes. In addition, the availability of raw data from large-scale student assessments have facilitated educational data mining, with a focus on predicting student academic performance using a set of predictors related to student and school characteristics (Gamazo and Martínez-Abad, 2020).

The studies have adopted a diverse array of ML techniques for performance prediction. Some have used decision-tree based algorithms such as classification and regression trees (CART) (Ma, 2005; Gabriel et al., 2018; She et al., 2019), chi-squared automatic interaction detection (CHAID) (Asensio-Muñoz et al., 2018), and C4.5 (Liu and Ruiz, 2008; Oskouei and Askari, 2014; Martínez-Abad, 2019, Gamazo and Martínez-Abad, 2020). Recent studies have used the ensemble learning method, including random forest (Cortez and Silva, 2008) and boosting (Masci et al. 2018) techniques. ML techniques such as least absolute shrinkage and selection operator (LASSO) are also used for variable selection and regularization (Sansone, 2019; Kerwin and Thornton, 2021). Knaus (2021) investigated the effects of musical practice on a student's academic skills, using the causal machine learning approach based on the Double Machine Learning (DML) estimator of Farrell (2015).

This study builds on this line of literature by applying boosting techniques to investigate the factors that are associated with academic performance. Our study focuses specifically on explaining the differences in mathematics achievement between East Asian and non-East Asian students. We consider a variety of student and school characteristics as predictors, as in previous literature, using the information drawn from the PISA 2018 survey. However, our approach is distinct from other studies, in that the model also includes a number of additional country-level variables as predicting variables. We consider how a country's level of development and socioeconomic environment can influence student academic achievement.

The ML results indicate that a country's student- and school-related factors, as well as its socioeconomic factors, are all important to predict the elevated PISA mathematics scores of East Asian students. We determine that a country's proportion of youth not in education, training, or employment, as well as the number of researchers in R&D, are the two most important features contributing to the higher average score of East Asian students. The student- and school-related variables — such as metacognition–assess credibility, mathematics learning time, duration in early childhood education and care, grade repetition, school type and size, class size, and student behavior hindering learning at school — also render important contributions. We also visualize the pattern between the predictors and predicted mathematics test scores across all individual students, and analyze the marginal effect of each predictor on the predicted test scores. The results highlight the nonlinear and complex relationships between educational input and outcomes.

The rest of the paper is organized as follows: Section 2 describes the data, Section 3 explains ML techniques, Section 4 presents and discusses estimation results, and Section 5 concludes the paper.

2. The Data

This paper uses the micro data of the PISA 2018, as well as several country-level macro data, to analyze the factors explaining the mathematics test scores. The PISA is a cross-national survey conducted by the Organization for Economic Co-operation and Development (OECD) with 15-year-old students that assesses students' academic performance in mathematics, reading, science, and problem solving (OECD, 2019). The PISA provides internationally comparable test scores and a lot of background information on both students and schools, which allows for the comparison and evaluation of students on a global level. Since its launch in 2000, the PISA assessment has been conducted every three years. This study uses the dataset assembled from the 2018 PISA wave, focusing on mathematics performance. Combining the available data, the dataset contains information on 590,102 students in 77 countries.[1]

In 2018, China[2] scored the best with 613 PISA points in mathematics, followed by Singapore, Macao SAR, Hong Kong SAR, and Chinese Taipei (Taiwan). Japan and Korea are also high-performing East Asian countries (regions). Among the non-East Asian countries, Estonia, the Netherlands, Poland, and Switzerland demonstrated the highest performance, while Panama, the Philippines, and the Dominican Republic scored lowest. On average, the mean score in mathematics was 554 in East Asia, compared to 454 in non-East Asia countries.[3] As can be seen from Table 1, East Asian students consistently performed better than non-East Asian students not just in terms of the mean, but rather across all corresponding percentiles. Figure 1 displays the kernel density estimates for students' mathematics test scores for East Asian versus non-East Asian countries. East Asian test scores show a distribution that is more concentrated.

Since the primary concern of this study is to explore the reasons that cause the differences in mathematics test scores between East Asian and non-East Asian countries, the output variable is the individual student test scores. The rich set of input variables includes student background information (characteristics of the students and their family) and school background information (characteristics of schools and teaching staff). The literature has shown that student and school characteristics are significantly associated with student academic achievements in international assessments (Hanushek and Woessmann, 2011; Barro and Lee

[1] The sample excludes Cyprus due to unavailability of PISA micro data.

[2] The PISA in 2018 was administered in four provinces including Beijing, Shanghai. Jiangsu, and Zhejiang of the People's Republic of China.

[3] The superior mathematics performance of East Asian countries is also shown in another major international assessment of student achievement such as the Trends in International Mathematics and Science Study (TIMSS) 2019.

2015). Appendix Table 1 provides a description of all 68 variables from the PISA dataset used in this study.

Table 1. Descriptive statistics of students' 2018 PISA test scores in mathematics.

Region (No. of country)	**Mean**	**p1**	**p5**	**p10**	**p25**	**p50**	**p75**	**p90**	**p95**	**p99**
PISA score										
Pooled (77)	462	231	292	326	387	461	536	598	633	694
East Asia (7)	554	315	389	428	493	561	622	671	701	751
non-East Asia (70)	454	229	288	322	381	453	525	586	619	678
Gap with East Asia's PISA score										
non-East Asia	100	87	100	106	111	108	96	86	81	73

Source: Authors' calculations from OECD (2019)

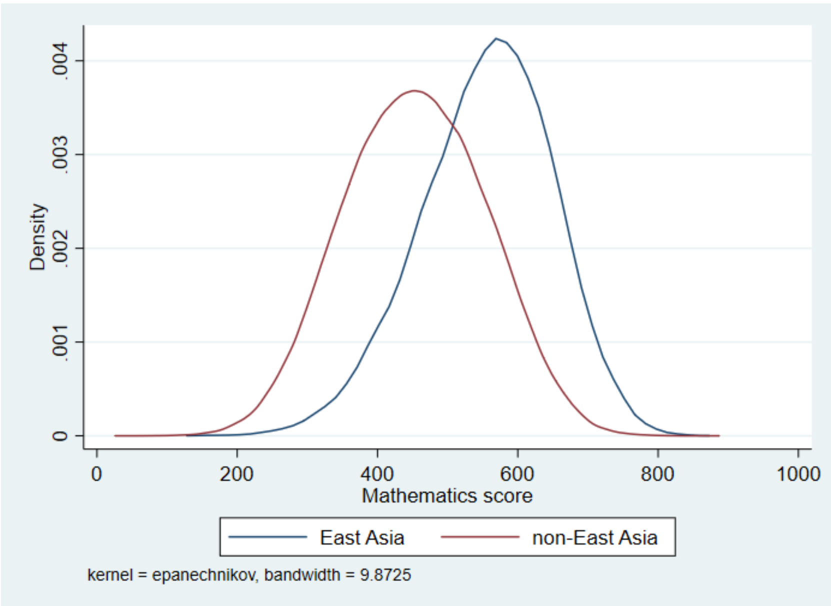


Figure 1. Kernel density distribution of PISA mathematics score: East Asia vs. non-East Asia

Source: Authors' calculations from OECD (2019)

In addition to considering the country's level of development and socioeconomic environment, we employ 10 additional variables, namely GDP per capita; life expectancy at birth; infant mortality rate; adolescent fertility rate; proportion of youth not in education, employment, or training (% of youth population); percentage of unemployed people with basic education (% of total labor force with basic education); number of researchers in R&D (per million people); R&D expenditure (% of GDP); secondary school enrolment rate; and the Gini index. Using a ML algorithm, we have selected these variables among 1,106 predictors that consist of country-level variables involving macroeconomy, population, labor market, education, health, and the environment. The data are collected from the World Bank's World Development Indicators (World Bank, 2020).[4] We have adopted recursive feature elimination and cross-validation selection (RFECV) to identify the variables providing the smallest root mean squared errors.

3. The Methodology

We employ an ensemble ML method in which the final model is a combination of simple prediction models, termed base learners. The base learners are assumed to be some class of functions such as simple linear models, spline functions, or regression trees. The gradient boosting technique, which was originally developed by Friedman (2001), adds the best new possible model sequentially to the prior base-learner models, to minimize overall prediction error. The algorithm of the gradient boosting machines or GBMs constructs the base-learners that are maximally correlated with the negative gradient of the predetermined loss function associated with the whole ensemble, thereby minimizing the expected loss (Natekin and Knoll, 2013). Hence, the learning procedure consecutively fits new models to provide a more accurate estimate of the response variable.

An important concern about building a machine-learning model is overfitting, which is when the final model does well in predicting only the training data itself, rather than fit out of sample. Thus, the ML algorithm aims to improve the derived final model through cross-validation. It randomly divides the original sample into subsamples (a training set and a test set) and experiments out-of-sample predictions within the original sample. Next, the training set is divided into a training set and a validation set once again. The estimation process then involves successively fitting a model on the training set and evaluating it on the validation set for a range of regularization parameters, such as the number of iterations and learning rate. Finally, the algorithm selects the hyperparameters with the best estimated average performance (Mullainathan and Spiess, 2017). In this study, we divide the sample into a training set (80%) and a test set (20%) and adopt 10-fold cross-validation.

We use the lightGBM (Light Gradient Boosting), which is a gradient boosting technique widely used in the ensemble method. The lightGBM technique uses a decision-tree

[4] Data for Taiwan are collected from its National Statistics and CEIC database.

ensemble that is a combination of multiple trees. It can handle missing values without an imputation process. Gradient boosting algorithms minimize the loss function, given the assumption that missing values contain information. The lightGBM approach is known to outperform other gradient boosting techniques in large-scale data sets (Ke et al., 2017).

We compare the performance of our model with that of other models, such as linear regression, ridge regression, and random forest, in predicting the mathematics test scores using the same set of predictors. This allows evaluating whether the lightGBM strategy is more suitable for predicting academic performance than the regression models, or other ML algorithms. The conventional linear regression approach is widely used in the literature and typically imposes a linear functional form based on the education production function. Ridge regression is designed to eliminate multicollinearity among features by imposing a penalty termed L2-norm, which is the sum of the squared coefficients. By contrast, the ML techniques assume a non-linear and complex relationship between the predicting variables and the test scores. They also avoid multicollinearity problems.

Table 2 presents four measures to compare predictive performance between the lightGBM and the alternative models. The root mean-squared error (RMSE) and the coefficient of determination or R-squared (R^2) values are two statistical measures that are typically used for model selection. The RMSE represents the root of the residual sum of squares, resulting from the comparison of the actual test score values to the predicted values. R^2 is determined by the proportion of the variance of the test scores that are explained by the predicting variables in the model. Table 2 also reports two other measures, such as mean absolute error and median absolute error, that evaluate the model fit. The results show that our lightGBM model outperforms the other three models in all four indicators.

Table 2. Comparing the predictive performance of linear regression and the lightGBM

	Performance Measures			
	Root Mean Square Error	**Mean Absolute Error**	**Median Absolute Error**	**R-squared**
The lightGBM	62.77	49.74	41.67	0.638
Random forests	81.85	64.91	54.46	0.385
Ridge regression	75.49	59.85	50.12	0.476
Linear Regression	75.65	59.96	50.22	0.474

4. The Results

Due to the complex relationships and interactions among the variables, the results from the lightGBM technique do not necessarily provide a simple explanation regarding the relationship between the predictor and the outcome variables; therefore, it is sometimes stated to be similar to being trapped inside a black box. This contrasts with the linear regression model, in which the regression coefficient shows the effect of a predictor on the outcome. In the ML model, we use the concept of “feature importance” to assess the relative importance of each input feature (predictor variables) in predicting an outcome variable. There are two approaches to measuring feature importance. One approach measures how much the prediction error increases when one of the given predictors is eliminated. The other gauges how much each feature contributes to the model’s prediction. We adopt the latter approach using the concept of SHAP (Shapley Additive exPlanations), proposed by Lundberg and Lee (Lundberg and Lee, 2017). SHAP is a game theoretic approach to explain the output of any machine learning model (Štrumbelj and Kononenko, 2014). It shows the extent to which each feature pushes the model output from the base value (the average model output of the training dataset) to the final model output (the predicted outcome). Applying the SHAP value, we can then assess the exact degree to which each predictor renders the predicted test scores of individual students either higher or lower than the global average score.

Figure 2 displays the SHAP value of each predictor for each student’s predicted score in the sample. It lists the top 15 variables that contribute most significantly to predicting individual test scores. The predictors are listed on the y-axis and individuals (instances) are on the x-axis. In the figure, f(x) represents the predicted test scores. The dotted horizontal line across the line of f(x) represents the global average test score. This graphic representation of test scores demonstrates that most East Asian students perform better compared to the global average.

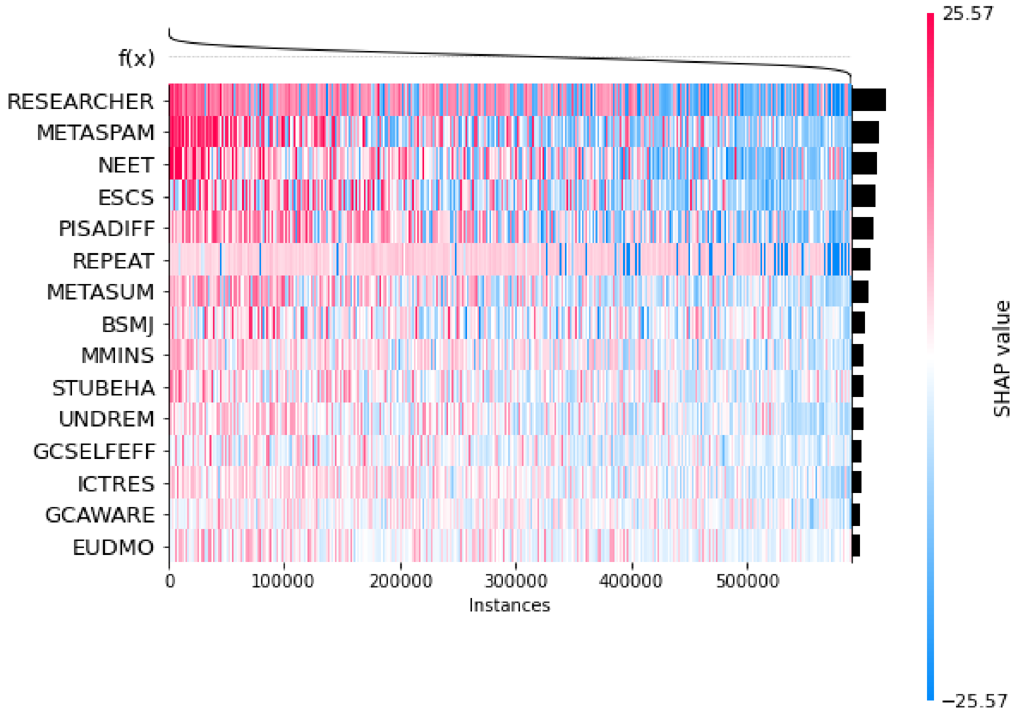


Figure 2. The SHAP values of each predictor for the predicted scores of individual students

Notes: The figure displays the SHAP value of each predictor for the predicted scores across all students in the sample. The predictors are listed on the y-axis, individuals (instances) are on the x-axis, and f(x) indicates the predicted test scores. The darker the red color, the more positively the predictor contributes to the predicted test score of the student. Conversely, the darker the blue color, the more negatively the predictor contributes. The top 15 variables that contribute most significantly to the predicted test scores are listed, as measured by the sum of absolute SHAP values for each predictor (represented by the bar on the right side of the figure). The full names of the predictors are shown in Appendix Table 1 and Appendix Figure 1.

The SHAP values are expressed in various colors. When the SHAP value of a predictor is marked in white, the predictor does not have any contribution to the predicted test score of the student compared to the global average score. Meanwhile, the dark red color of the SHAP value indicates that the predictor contributes positively to the student's predicted test score, by rendering it higher than the global average. The darker the color, the larger the positive contribution is. Conversely, the darker the blue color of the SHAP value, the more negatively the predictor contributes to the predicted test scores of the student. If the color of the SHAP value for a predictor changes from a dark red to a dark blue, uniformly across all individuals — from the highest performing student to the lowest performing student — then this predictor is considered an immensely important variable that causes significant differences in test scores across students.

By comparing the sum of absolute SHAP values of each predictor across individual students, it is possible to identify the predictors that contribute largely to differences in students' mathematics performance. Figure 2 lists the top 15 predictors selected by the sum of absolute SHAP values, and the magnitudes are displayed in a bar on the right side.[5]

The top 15 variables consist of country-, school-, and student-related factors. More specifically, they consist of two country factors (namely the number of researchers in R&D; RESEARCHER and the proportion of youth not in education, training, or employment; NEET), one school factor (namely student behaviors that hinder learning; STUBEHA), and 12 student features. The student factors include three metacognition variables (METASPAM, METASUM, UNDREM); mathematics learning time (MMINS); grade repetition (REPEAT); index of economic, social, and cultural status (ESCS); ICT resources at home (ICTRES); perception of difficulty of the PISA test (PISADIFF); student's expected occupational status (BSMJ); self-efficacy regarding global issues (GCSELFEFF); student's awareness of global issues (GCAWARE); and eudaemonia–meaning in life (EDUMO).

The top 15 features include 12 student background variables, related to a student's cognitive abilities, study time, family background, ICT resources, and expected occupational status. Previous literature that employed multivariate regression analysis or case studies have highlighted these student background variables as important factors for student academic performance (Hanushek and Woessmann, 2011; Barro and Lee, 2015). Recent studies have also discussed the impact of ICT use on student academic achievement (Beland and Murphy, 2016). Our study confirms the significance of these variables on student math performance using the ML method. In addition, we find that country-related factors, including NEET and the number of researchers in R&D, significantly correlate with student math scores. These variables can have a significant influence on an environment or society and can impact a student's motivation to learn. However, the causal relationship may go the other way, as higher

[5] The sum of absolute SHAP values for the top 15 predictors with their full variable names are shown in Appendix Figure 1.

academic achievement increases the proportion of youth in higher education and employment, and produces more researchers in R&D.

Feature importance as a concept, is useful in determining the variables that bring about differences in academic performance among all individual students in the sample. However, we do not know which predictors are most significantly related to the difference in math test scores between East Asian and non-East Asian students. The key aim of this study is identifying the important features that predict the elevated mathematics performance of East Asian students. To answer this question, we decompose the predicted math score of an average East Asian student into the contributions from the predictors. We first construct a machine using the student data from all countries and predict the math score of an (artificial) average East Asian student. The predicted value is 557.5 points, which is 96.6 points higher than the global average of 460.9. The SHAP value of each predictor measures its contribution to the difference between the average East Asian student's score and the global average score.

Figure 3 displays the SHAP value of each predictor for the student's score across all individual East Asian students. It lists the top 15 variables that contribute most significantly to the predicted outcome. They contain three country-related factors (NEET, number of researchers in R&D, and percentage of unemployed with basic education), four school-related factors (school type–private independent, school size, class size, and student behaviors that hinder learning), and eight student-related factors (metacognition–assess credibility, mathematics learning time, grade repetition, duration in early childhood education and care, ICT use outside of school for leisure, discriminating school climate, student's attitudes toward immigrants, and work mastery).

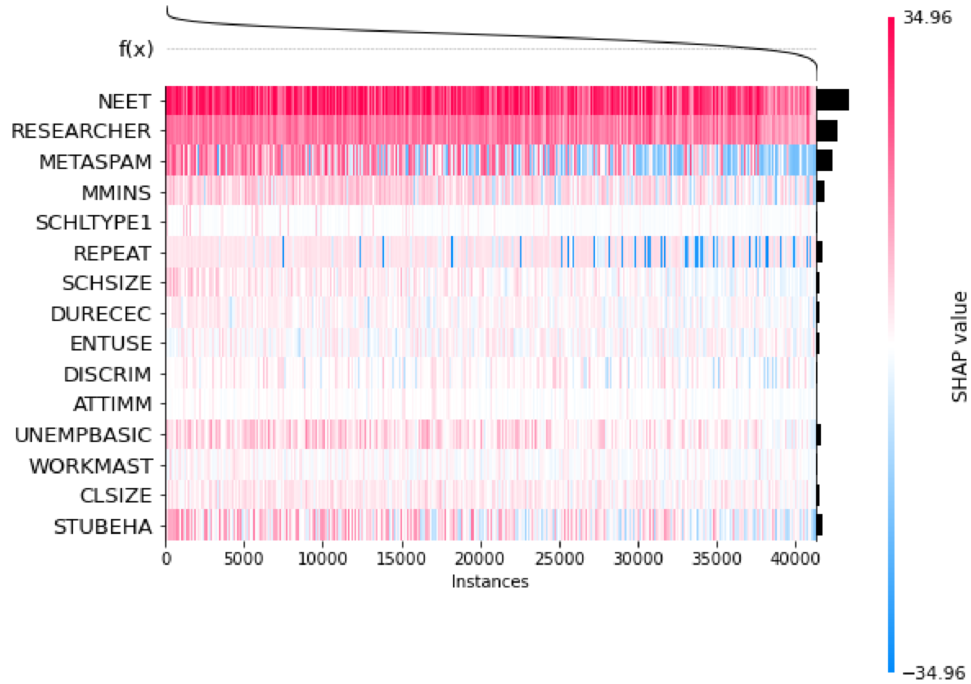


Figure 3. The SHAP values of each predictor for the predicted scores of East Asian students

Notes: The figure displays the SHAP value of each predictor for the predicted scores of East Asian students. Please also refer to the notes in Figure 2.

Compared to the top 15 variables that best predict the mathematics scores of all students in Figure 2, substantial changes occur in the new model. Although many student factors, such as metacognition–assess credibility, mathematics learning time, and grade repetition, remain important for predicting the difference between the average East Asian student's score and the global average score, some student factors — such as the index of economic, social, and cultural status; self-efficacy regarding global issues; awareness of global issues; and eudaemonia or meaning in life — do not appear important. On the other hand, many school characteristics, including school type (private/public, independent/government dependent), school size, class size, discriminating school climate, and student's attitudes toward immigrants, appear important in explaining the elevated average mathematics score of East Asian students relative to their peers. The literature has emphasized the significant impact of school characteristics and school climate (Hanushek and Woessmann, 2011; Thapa et al., 2013) on student learning outcomes. The duration of early childhood education and care is also identified as an important factor for student achievement, which is consistent with other studies' findings (for instance, Heckman, 2006). Two country-related factors, including NEET and the number researchers in R&D, as well as another percentage of unemployed people with basic education variable, are identified as being a part of the top 15 features.

Table 3 presents the decomposition results using the SHAP values. The model output, that is the predicted math score of the artificial East Asian student, is 557.5 points, compared to the actual mean of 550.4 points. Based on the concept of SHAP, we decompose the predicted average score of 557.5 into the global average score of 460.9 points and the contributions by each of the 78 predictors. The SHAP value indicates how much each predictor contributes to rendering the average East Asian student's score higher (or lower) than the global average score.

Table 3. Analysis of the predicted score of an average East Asian student

Actual average score of East Asian students	550.4
Predicted average score of East Asian students	557.5
Contribution to predicted average score	
Global average student score	460.9
(Country) Proportion of youth not in education, employment, or training	17.0
(Country) Number of researchers in R&D (per million people)	10.8
(Student) Meta-cognition: assess credibility	7.1
(Student) Mathematics learning time (minutes per week)	4.8
(School) School type: private Independent	4.6
(Student) Grade repetition	4.2
(School) School size	4.0
(Student) Duration in early childhood education and care	3.9
(Student) ICT use outside of school (leisure)	3.7
(Student) Discriminating school climate	3.6
(Student) Student's attitudes toward immigrants	3.5
(Country) Unemployment with basic education	3.4
(Student) Work mastery	3.2
(School) Class size	3.2
(School) Student behaviors that hinder learning	3.1
62 other features	-12.7

Notes: The value is constructed using the SHAP (Shapley Additive exPlanations) value. The SHAP value of each variable measures its contribution to the difference between the average East Asian student's score and the global average score. The variables consist of country-, school-, and student-related factors, as indicated in parenthesis before the variable name.

Among the top 15 variables that contribute to higher East Asian scores (compared to the global average), NEET and the number of researchers in R&D are the two most important features, contributing 17.0 and 10.8 points, respectively. Student-related features are also important, including metacognition–assess credibility which measures competence in assessing the quality and source credibility of information, mathematics learning time, and grade repetition, as these specific features contribute 7.1, 4.8, and 4.2 points, respectively, to the average score of East Asian students. School background variables, such as school type–private independent, school size, and class size, also contribute significantly to the average East Asian student score, by 4.6, 4.0, and 3.1 points respectively.

Differences in the key input variables between East Asian and non-East Asian students must have influenced PISA mathematics test scores and academic performance as a whole. Table 4 compares the average values of each predictor, presenting significant differences between East Asian and non-East Asian countries. For example, the average value of NEET is 3.4% in East Asia, compared to 16.7% in non-East Asian countries. East Asian countries, on average, have more researchers in R&D when adjusted for population size. Also, on average, East Asian students have better metacognition–assess credibility than non-East Asian students: 0.078 vs. -0.262. The average mathematics study time per week for East Asian and non-East Asian students is 257 and 241 minutes, respectively. In addition, the grade repetition rate is lower in East Asian countries, compared to non-East Asian counterparts.

Furthermore, East Asian students usually have a better learning environment than non-East Asian students, indicated by a less discriminating school climate, a more positive attitude towards immigrants, and fewer student behaviors that hinder learning. Generally, East Asian schools are larger in size, and are more private-independent. Appendix Table 2 compares the distribution of the top 15 variables in terms of percentiles, between the two samples of East Asian and non-East Asian students.

Table 4. Comparing the top 15 features among the global, East Asian, and non-East Asia countries

Features	Global	East Asian	Non-East
(Country) Proportion of youth not in education,	16.13	3.44	16.72
(Country) Number of researchers in R&D (per million	2252.8	4135.6	2048.8

(Student) Meta-cognition: assess credibility	-0.227	0.078	-0.262
(Student) Mathematics learning time (minutes per week)	242.9	256.8	241.1
(School) School type: private Independent	0.102	0.189	0.093
(Student) Grade repetition	0.148	0.074	0.153
(School) School size	1058.3	1324.6	1027.5
(Student) Duration in early childhood education and care	2.525	3.095	2.458
(Student) ICT use outside of school (leisure)	-0.017	-0.174	0.001
(Student) Discriminating school climate	0.151	-0.493	0.172
(Student) Student's attitudes toward immigrants	-0.109	0.408	-0.125
(Country) Unemployment with basic education	7.818	2.797	7.928
(Student) Work mastery	0.141	0.130	0.143
(School) Class size	32.23	35.95	31.80
(School) Student behavior hindering learning	0.083	-0.167	0.110

Each variable's contribution to the predicted test score can also be analyzed in more detail using partial dependency plots (PDPs). Partial dependence demonstrates the effect of a predictor on the modeled response, after marginalizing out all other input features. Figure 4 plots PDPs using the information between each predictor and its SHAP value, calculated from all available observations. A red line indicates the nonparametric relationship between the SHAP value and the predictor, using LOWESS smoothing. The solid, dotted, and (loosely) dashed vertical lines indicate the average values of each predictor in the global, East Asian, and Non-East Asian samples, respectively.

Figure 4.a shows the partial dependency plot for the most influential feature, which is the number of researchers in R&D. The overall positive relationship indicates that an increase in researchers tends to increase the expected test scores. However, as observed in the plot, a nonlinear relationship exists between these factors. For the students who receive scores just below the global average, the predicted score increases sharply with the increased number of researchers, yet this does not change for the range of math scores above the world average. Figure 4.b shows a clear negative relationship between NEET and student's predicted test scores. The overall negative relationship suggests that a lower value of this variable has a positive effect on East Asian students' test scores.

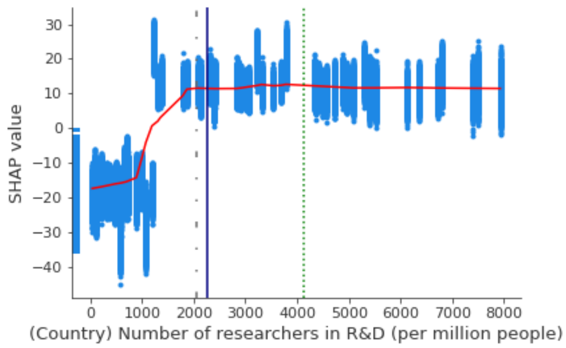
In Figure 4.c, student's metacognition–assess credibility has a clear positive relationship with predicted test scores. East Asian students' higher metacognition ability seems to contribute unequivocally to their better mathematics achievement. When the distribution of the metacognition index is examined (see Appendix Table 2), East Asian students are seen to have fewer students at the 25th percentile of the distribution, which renders the average expected test score of East Asian students higher than that of non-East Asian students.

Figure 4.d shows the importance of study time for expected test scores, especially for those students in the left tail of the distribution. The average mathematics learning time is slightly longer in East Asian students than in non-East Asian students (257 vs. 241 minutes per week). The distribution of math learning time is skewed more leftwards in the sample of non-East Asian students than the East Asian students (Appendix Table 2). The fact that a relatively large proportion of students spend less time studying mathematics in non-East Asian countries than in East Asian ones, appears to contribute to the lowering of the average math scores in non-East Asian countries.

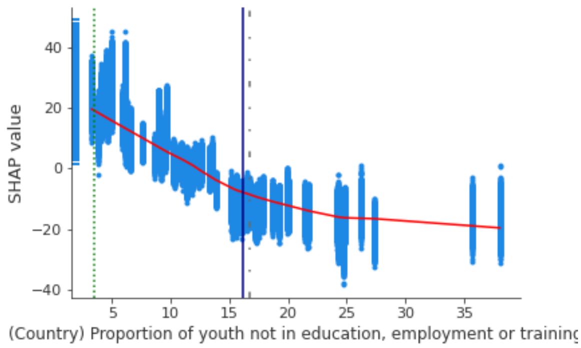
Figure 4.e shows the nonlinear relationship between student's ICT use outside of school for leisure and predicted test scores. The inverted-V type relationship indicates that expected math scores tend to be lower if a student has too much interest in ICT or too little. Figure 4.f shows school size has a positive relationship with test scores, especially in the range of school size below the world average.

Figure 4.g and 4.h presents evidence for the importance of the quality of learning environment for mathematics achievement. Both the discriminating school climate and student behaviors that hinder learning variables have significantly negative relationships with students' expected test scores. On average, East Asian students have better learning environments than their non-East Asian peers, as shown in Appendix Table 2, which presumably contributes to their better mathematics performance.

(a)



(b)



(c)

(d)

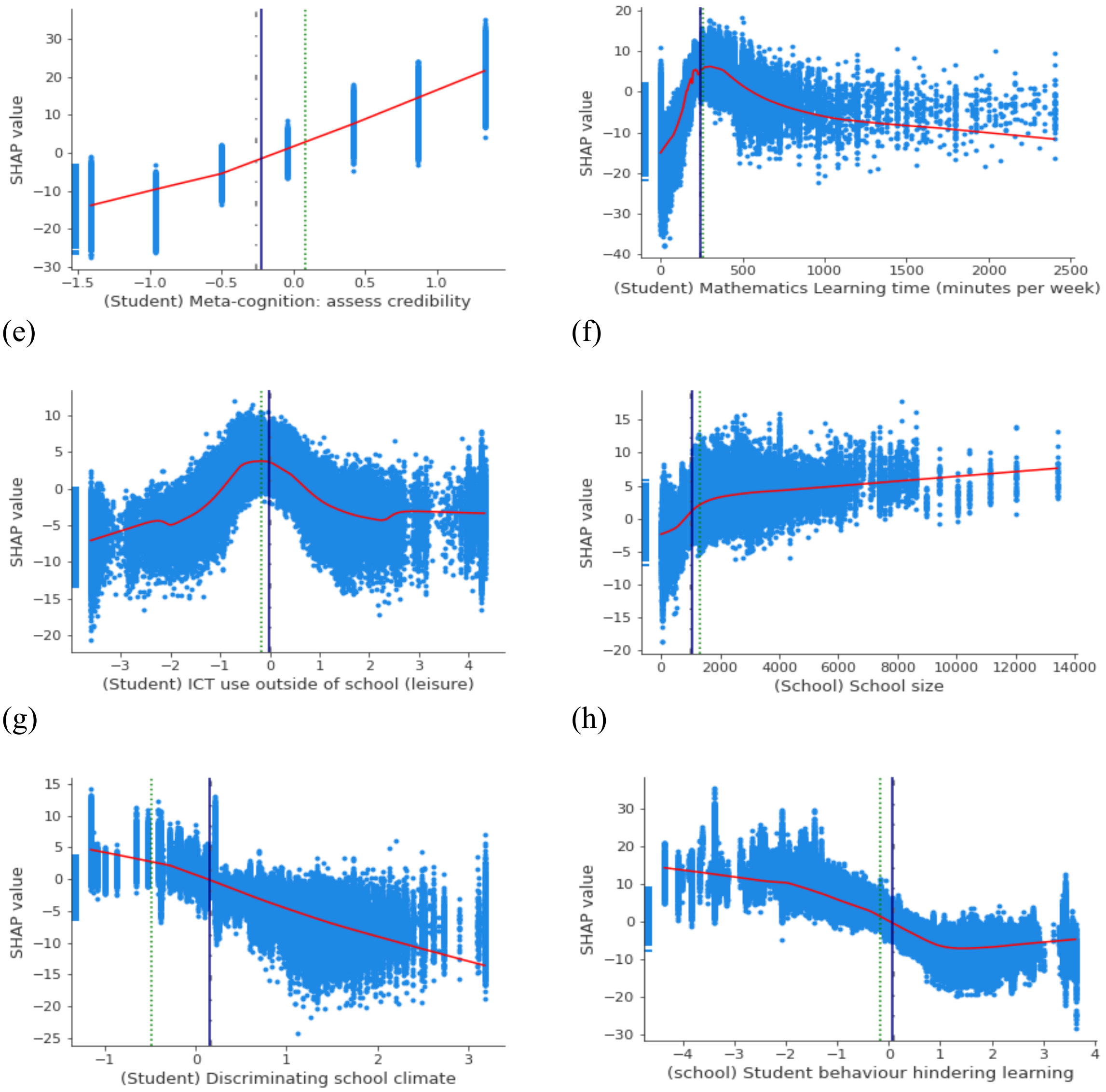


Figure 4. Marginal effects of individual features

Notes: The LOWESS line represents the nonparametric relationship between SHAP values and predictor values. The solid, dotted, and (loosely) dashed vertical lines indicate the average values of each predictor in the global, East Asian, and non-East Asian samples, respectively.

In summary, the significant differences in the key predictors between East Asian and non-East Asian students seem to contribute to the higher average mathematics test scores of East Asian students. The PDPs in Figure 4 suggest that there are nonlinear and complex relationships between predicting variables and students' mathematics achievement.

5. Concluding remarks

This study explores the relationships between student, school, and country characteristics and students' mathematics achievement using the 2018 PISA data. Adopting the ML techniques, we have identified the top 15 factors that have the highest predicting powers and assessed the extent that each factor contributes to increasing the predicted average math score of East Asian students compared to the global average score.

Our ML results demonstrate that variations in the 2018 PISA mathematics scores among students of 77 countries and regions are not only closely associated with student- and school-related factors, but also country-level socioeconomic factors. We find student- and school-related factors, such as metacognition–assess credibility, mathematics learning time, duration in early childhood education and care, grade repetition, school type and size, class size, and student behavior hindering learning, as important predictors of the elevated test scores of East Asian students. Moreover, several country-level factors — such as the proportion of youth not in education, training, or employment; the number of researchers in R&D; and the percentage of unemployed with basic education — have high predicting power for the higher average scores of East Asian students. In summary, higher math scores for East Asian students stem from their hard work, higher cognitive abilities, better school learning environments, as well as better societal labor market opportunities.

We have also visualized the pattern between the predictors and predicted test scores using the PDPs and analyzed the marginal effect of each variable to predicted test scores across all individual students. We found that the relationship between educational inputs and outcomes is not linear; therefore, non-linearities and discontinuities must be considered, to further investigate the complex and unknown process of student's learning.

The ML approach employed here is conceptually different from other modeling approaches, in that it is very strongly data-driven and imposes no specific functional form between input and outcome variables, yet the results are highly intuitive. Although ML techniques do not provide direct causal inference, they are useful in identifying important input factors that explain outcomes. In the field of ML, novel techniques are being actively developed. In this study, we used one type of algorithm, namely gradient boosting, and did not conduct an in-depth analysis of the causal interpretation for deriving policy implications. As ML is currently undergoing rapid advancements, the analytical field can help us advance the implementation of educational data mining and our understanding of the effects of certain policies on educational outcomes.

References

Angrist, N., Djankov, S., Goldberg, P. K., & Patrinos, H. A. (2021). Measuring human capital using global learning data. *Nature*, *592*(7854), 403-408.

Asensio Muñoz, I., Carpintero Molina, E., Expósito Casas, E., & López Martín, E. (2018). How much gold is in the sand? Data mining with Spain's PISA 2015 results. *Revist. Española Pedagogía 76*, 225–246.

Athey, S., & Imbens, G. W. (2019). Machine learning methods that economists should know about. *Annual Review of Economics*, *11*, 685-725.

Barro, R. J., & Lee, J.-W. (2015). *Education matters: Global schooling gains from the 19th to the 21st century*. Oxford University Press.

Beland, L.-P., & Murphy, R. (2016). Ill communication: technology, distraction & student performance. *Labour Economics*, *41*, 61-76.

Biggs, J. B. (1996). Western misconceptions of the Confucian-heritage learning culture. In D. A. Watkins & J. B. Biggs. (Eds.), *The Chinese learner: Cultural, Psychological and Contextual Influences* (pp. 45-67). Australian Council for Educational Research and the Comparative Education Research Centre, University of Hong Kong.

Cortez, P., & Silva, A. M. G. (2008). *Using data mining to predict secondary school student performance*. 5th Annual Future Business Technology Conference, Porto.

Dietterich, T. G. (2000). *Ensemble methods in machine learning*. International Workshop on Multiple Classifier Systems,

Dowker, A., & Li, A. M. (2019). English and chinese children's performance on numerical tasks. *Frontiers in Psychology*, *9*, 2731.

Farrell, M. H. (2015). Robust inference on average treatment effects with possibly more covariates than observations. *Journal of Econometrics*, *189*(1), 1-23.

Francis, B., & Archer, L. (2005). British—Chinese pupils' and parents' constructions of the value of education. *British Educational Research Journal*, *31*(1), 89-108.

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of Statistics*, 1189-1232.

Fuson, K. C., & Kwon, Y. (1991). Chinese-based regular and European irregular systems of number words: The disadvantages for English-speaking children. *Language in Mathematical Education: Research and Practice*, 211-226.

Gabriel, F., Signolet, J., & Westwell, M. (2018). A machine learning approach to investigating the effects of mathematics dispositions on mathematical literacy. *International Journal* *of Research & Method in Education*, *41*(3), 306-327.

Gamazo, A., & Martínez-Abad, F. (2020). An Exploration of Factors Linked to Academic Performance in PISA 2018 Through Data Mining Techniques. *Frontiers in Psychology*, *11*, 3365.

Hanushek, E. A., & Kimko, D. D. (2000). Schooling, labor-force quality, and the growth of nations. *American Economic Review*, *90*(5), 1184-1208.

Hanushek, E. A., & Woessmann, L. (2011). The economics of international differences in educational achievement. In Hanushek, E. A., & Woessmann (Ed.), *Handbook of the Economics of Education* (Vol. 3, pp. 89-200). Elsevier.

Hanushek, E. A., & Woessmann, L. (2016). Knowledge capital, growth, and the East Asian miracle. *Science*, *351*(6271), 344-345.

Heckman, J. J. (2006). Skill formation and the economics of investing in disadvantaged children. *Science*, *312*(5782), 1900-1902.

Hess, R. D., Chang, C.-M., & McDevitt, T. M. (1987). Cultural variations in family beliefs about children's performance in mathematics: Comparisons among People's Republic of China, Chinese-American, and Caucasian-American families. *Journal of Educational Psychology*, *79*(2), 179.

Ho, I. T., & Hau, K. T. (2008). Academic achievement in the Chinese context: The role of goals, strategies, and effort. *International Journal of Psychology*, *43*(5), 892-897.

Jerrim, J., & Vignoles, A. (2016). The link between East Asian 'mastery' teaching methods and English children's mathematics skills. *Economics of Education Review*, *50*, 29-44.

Kaya, S., & Rice, D. C. (2010). Multilevel effects of student and classroom factors on elementary science achievement in five countries. *International Journal of Science Education*, *32*(10), 1337-1363.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems*, *30*, 3146-3154.

Kerwin, J. T., & Thornton, R. L. (2021). Making the grade: The sensitivity of education program effectiveness to input choices and outcome measures. *Review of Economics and Statistics*, *103*(2), 251-264.

Kim, K. H. (2005). Learning from each other: Creativity in East Asian and American education. *Creativity Research Journal*, *17*(4), 337-347.

Kim, Y. (2021). Home educational contexts of Asian American children: disentangling the effects of structural and cultural factors. *Early Childhood Research Quarterly*, *54*, 307-320.

Knaus, M. C. (2021). A double machine learning approach to estimate the effects of musical practice on student's skills. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *184*(1), 282-300.

Lee, J. W., & Barro, R. J. (2001). Schooling quality in a cross–section of countries. *Economica*, *68*(272), 465-488.

Leung, F. K. (2001). In search of an East Asian identity in mathematics education. *Educational Studies in Mathematics*, *47*(1), 35-51.

Leung, F. K., Park, K., Shimizu, Y., & Xu, B. (2015). *Mathematics education in East Asia*. The Proceedings of the 12th International Congress on Mathematical Education,

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

Lundberg, S. M., & Lee, S.-I. (2017). *A unified approach to interpreting model predictions*. Proceedings of the 31st International Conference on Neural Information Processing Systems,

Lynn, R. (1982). IQ in Japan and the United States shows a growing disparity. *Nature*, *297*(5863), 222-223.

Ma, L. (1999). *Knowing and Teaching Elementary Mathematics*. Lawrence Erlbaum Associates.

Ma, X. (2005). Growth in mathematics achievement: Analysis with classification and regression trees. *The Journal of Educational Research*, *99*(2), 78-86.

Martínez-Abad, F. (2019). Identification of factors associated with school effectiveness with data mining techniques: testing a new approach. *Frontiers in Psychology*, *10*, 2583.

Masci, C., Johnes, G., & Agasisti, T. (2018). Student and school performance across countries: A machine learning approach. *European Journal of Operational Research*, *269*(3), 1072-1085.

Miura, I. T., Kim, C. C., Chang, C.-M., & Okamoto, Y. (1988). Effects of language characteristics on children's cognitive representation of number: Cross-national comparisons. *Child Development*, 1445-1450.

Mullainathan, S., & Spiess, J. (2017). Machine learning: an applied econometric approach. *Journal of Economic Perspectives*, *31*(2), 87-106.

Mullis, I. V. S., Martin, M. O., Foy, P., Kelly, D. L., & Fishbein, B. (2020). *TIMSS 2019 International Results in Mathematics and Science*. Retrieved from Boston College, TIMSS & PIRLS International Study Center. https://timssandpirls.bc.edu/timss2019/international-results/ (access December 27, 2020).

Natekin, A., & Knoll, A. (2013). Gradient boosting machines, a tutorial. *Frontiers in Neurorobotics*, *7*, 21.

OECD. (2019). *PISA 2018 results (volume I): What students know and can do*. OECD.

Oskouei, R. J., & Askari, M. (2014). Predicting academic performance with applying data mining techniques (generalizing the results of two different case studies). *Computer Engineering and Applications Journal*, *3*(2), 79-88.

Sandri, M., & Zuccolotto, P. (2008). A bias correction algorithm for the Gini variable importance measure in classification trees. *Journal of Computational and Graphical Statistics*, *17*(3), 611-628.

Sansone, D. (2019). Beyond early warning indicators: high school dropout and machine learning. *Oxford Bulletin of Economics and Statistics*, *81*(2), 456-485.

She, H. C., Lin, H. s., & Huang, L. Y. (2019). Reflections on and implications of the Programme for International Student Assessment 2015 (PISA 2015) performance of students in Taiwan: The role of epistemic beliefs about science in scientific literacy. *Journal of Research in Science Teaching*, *56*(10), 1309-1340.

Stevenson, H. W., Lee, S.-Y., & Stigler, J. W. (1986). Mathematics achievement of Chinese, Japanese, and American children. *Science*, *231*(4739), 693-699.

Stevenson, H. W., Stigler, J. W., Lee, S.-y., Lucker, G. W., Kitamura, S., & Hsu, C.-c. (1985). Cognitive performance and academic achievement of Japanese, Chinese, and American children. *Child Development*, 718-734.

Štrumbelj, E., & Kononenko, I. (2014). Explaining prediction models and individual predictions with feature contributions. *Knowledge and Information Systems*, *41*(3), 647-665.

Thapa, A., Cohen, J., Guffey, S., & Higgins-D'Alessandro, A. (2013). A review of school climate research. *Review of Educational Research*, *83*(3), 357-385.

Valverde, G. A., & Schmidt, W. H. (2000). Greater expectations: Learning from other nations in the quest for'world-class standards' in US school mathematics and science. *Journal of Curriculum Studies*, *32*(5), 651-687.

Wang, J., & Lin, E. (2005). Comparative studies on US and Chinese mathematics learning and the implications for standards-based mathematics teaching reform. *Educational Researcher*, *34*(5), 3-13.

World Bank. (2020). World Development Indicators, Washington, DC: World Bank, https://datacatalog.worldbank.org/dataset/world-development-indicators. (last accessed: 4 April, 2020)

FOR ONLINE PUBLICATION

Appendix Table 1. List of variables

Variable	Description	No. of country	No. of Obs.
MATH	Mathematics	77	590,102
Student-level predictors (52)			
AGE	Age	77	590,102
ATTIMM	Student's attitudes towards immigrants	58	360,801
ATTLNACT	Attitude towards school: learning activities	77	545,716
AUTICT	Perceived autonomy related to ICT use	51	292,278
AWACOM	Awareness of intercultural communication	63	414,745
BEINGBULLIED	Student's experience of being bullied	73	447,071
BELONG	Subjective well-being: Sense of belonging to school	74	514,859
BSMJ	Student's expected occupational status	77	456,604
CHANGE	Number of changes in educational biography	31	223,353
COGFLEX	Cognitive flexibility/adaptability	64	422,122
COMPETE	Competitiveness	76	536,324
COMPICT	Perceived ICT competence	52	299,508
CULTPOSS	Cultural possessions at home	77	567,381
DISCRIM	Discriminating school climate	57	331,090
DURECEC	Duration in early childhood education and care	76	439,275
EMOSUPS	Parents' emotional support perceived by student	75	464,491
ENTUSE	ICT use outside of school (leisure)	52	326,871
ESCS	Index of economic, social and cultural status	77	576,168
EUDMO	Eudaemonia: meaning in life	72	487,115
GCAWARE	Student's awareness of global issues	66	436,380
GCSELFEFF	Self-efficacy regarding global issues	64	427,046
GFOFAIL	General fear of failure	75	522,614
GLOBMIND	Global-mindedness	62	385,660
HEDRES	Home educational resources	77	572,942
ICTHOME	ICT available at home	51	348,250
ICTRES	ICT resources	77	574,929
IMMIG1	Immigration status: Native	76	558,280
IMMIG2	Immigration status: Second-Generation	76	558,280
IMMIG3	Immigration status: First-Generation	76	558,280
INFOCAR	Information about careers	32	222,628
INTCULT	Student's interest in learning about other cultures	62	393,730
INTICT	Interest in ICT	52	305,682
JOYREAD	Joy/Like reading	75	553,154
MASTGOAL	Mastery goal orientation	75	521,195
METASPAM	Meta-cognition: assess credibility	75	508,444
METASUM	Meta-cognition: summarizing	75	516,029
MMINS	Mathematics Learning time (minutes per week)	75	450,778
PERCOMP	Perception of competitiveness at school	76	461,586
PERCOOP	Perception of cooperation at school	76	447,365
PERSPECT	Perspective-taking	64	423,445
PISADIFF	Perception of difficulty of the PISA test	74	532,480
REPEAT	Grade Repetition	74	554,010
RESILIENCE	Resilience	75	518,203
RESPECT	Respect for people from other cultures	63	406,347
SCCHANGE	Number of school changes	31	225,011

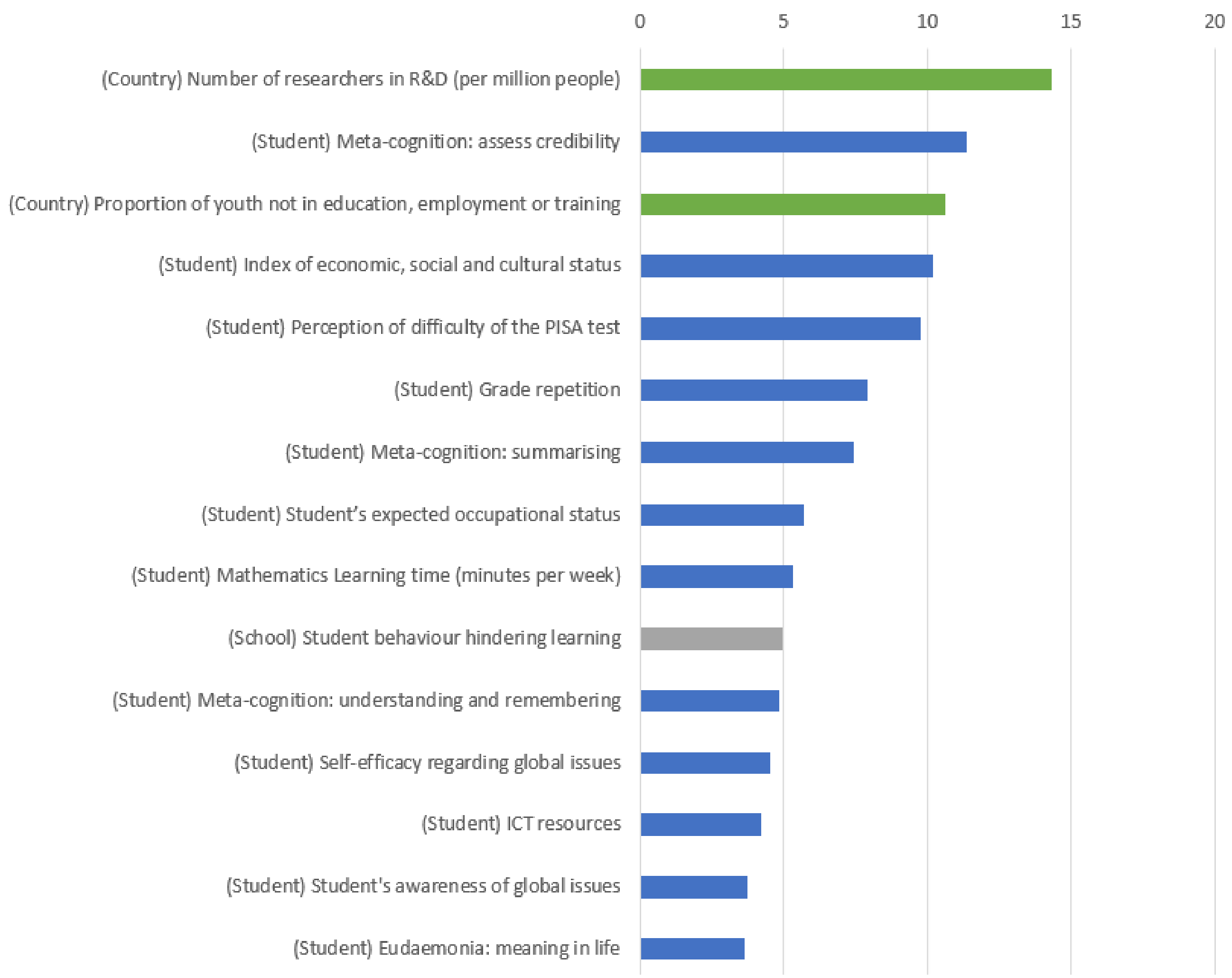
SCREADCOMP	Self-concept of reading: Perception of competence	73	521,386
SCREADDIFF	Self-concept of reading: Perception of difficulty	71	518.226
SOIAICT	ICT as a topic in social interaction	50	288.199
SWBP	Subjective well-being: Positive affect	68	467,548
UNDREM	Meta-cognition: understanding and remembering	75	515,609
WEALTH	Family wealth	77	576,799
WORKMAST	Work mastery	76	526,542
School-level predictors (16)			
CLSIZE	Class Size	74	528,444
CREACTIV	Creative extra-curricular activities	77	560,820
PROATCE	Index proportion of all teachers fully certified	73	500,668
RATCMP1	Number of available computers per student at modal grade	77	518,850
RATCMP2	Proportion of available computers connected to the Internet	77	523,744
SCMCEG	School principal's view on teachers' multicultural and egalitarian belief	63	460,923
SCHSIZE	School Size	73	500,033
SCHLTYPE1	School type: Private Independent	73	549,024
SCHLTYPE2	School type: Private Government-dependent	73	549,024
SCHLTYPE3	School type: Public	73	549,024
EDUSHORT	Shortage of educational material	77	564,233
STAFFSHORT	Shortage of educational staff	77	564,749
STUBEHA	Student behavior hindering learning	77	567,251
STRATIO	Student-Teacher ratio	73	492,283
TEACHBEHA	Teacher behavior hindering learning	77	566,864
TOTAT	Total number of all teachers at school	72	511,696
Country-level predictors (10)			
ADFERT	Adolescent fertility rate (births per 1,000 women ages 15-19)	75	585,044
ENROLNET	School enrollment, secondary (% net)	69	536,118
GDPPC	GDP per capita, PPP (current international $)	76	590,102
GINI	Gini index	71	553,577
LIFEEXP	Life expectancy at birth, total (years)	76	590,102
MORT	Mortality rate, infant (per 1,000 live births)	73	575,232
NEET	Proportion of youth not in education, employment or training (% of youth population)	70	539,335
RESEARCHER	Number of researchers in R&D (per million people)	66	521,648
RNDEXP	Research and development expenditure (% of GDP)	70	553,755
UNEMPBASIC	Unemployment with basic education (% of total labor force with basic education)	72	559,507

Source: Author's calculation from PISA 2018 Technical Report

(https://www.oecd.org/pisa/data/pisa2018technicalreport/)

Appendix Table 2. Comparison of the percentiles of the top 15 features between the samples of East Asian and non-East Asian students

	East Asia					Non-East Asia				
Features	p10	p25	p50	p75	p90	p10	p25	p50	p75	p90
(Country) Proportion of youth not in education, employment or training	3.27	3.27	4.50	6.10	6.10	6.54	9.49	13.32	19.25	24.33
(Country) Number of researchers in R&D (per million people)	1224.8	1224.8	3782.3	6802.5	7497.6	485.4	707.7	2383.1	4325.7	5387.9
(Student) Meta-cognition: assess credibility	-1.41	-0.96	-0.04	0.87	1.33	-1.41	-1.41	-0.04	0.42	1.33
(Student) Mathematics learning time (minutes per week)	150	200	240	300	385	120	180	220	250	360
(School) School type: Private independent	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00
(Student) Grade repetition	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00
(School) School size	455	708	979	1449	2382	218	404	677	1062	1613
(Student) Duration in early childhood education and care	2.00	3.00	3.00	4.00	4.00	1.00	2.00	3.00	3.00	4.00
(Student) ICT use outside of school (leisure)	-0.82	-0.44	-0.08	0.34	0.77	-0.90	-0.44	-0.05	0.43	0.93
(Student) Discriminating school climate	-1.15	-1.15	-0.42	0.59	0.94	-1.15	-1.15	0.10	0.59	1.42
(Student) Student's attitudes towards immigrants	-0.64	-0.20	-0.20	0.94	1.50	-1.18	-0.62	-0.20	0.64	1.50
(Country) Unemployment with basic education	2.28	2.61	3.53	3.87	3.87	1.17	6.50	10.44	15.36	23.85
(Student) Work mastery	-1.03	-0.23	-0.10	0.71	1.82	-1.07	-0.65	-0.10	0.88	1.82
(School) Class size	23.00	28.00	33.00	38.00	43.00	18.00	23.00	28.00	33.00	43.00
(School) Student behavior hindering learning	-2.07	-1.19	-0.34	0.47	1.76	-1.38	-0.62	0.08	0.82	1.41



Appendix Figure 1. Top 15 most important features for predicting student test score

Note: This feature importance is measured by the sum of absolute SHAP values for each predictor, which is also shown by the bar on the right side of the figure 3.