Inequality and Students' PISA 2018 Performance: a Cross-Country Study

3 authors:

Jiří Mazurek
Silesian University in Opava
102 PUBLICATIONS   **526** CITATIONS

SEE PROFILE

Carlos Fernández García
King Juan Carlos University
22 PUBLICATIONS   **63** CITATIONS

SEE PROFILE

Cristina Pérez Rico
King Juan Carlos University
30 PUBLICATIONS   **66** CITATIONS

SEE PROFILE

C|O|P|E
Member since 2018
JM13708

Inequality and Students' PISA 2018 Performance: a Cross-Country Study

Jiri Mazurek (iD) https://orcid.org/0000-0002-7965-0457
Ph.D., Silesian University in Opava, Opava, Czech Republic, e-mail: mazurek@opf.slu.cz

Carlos Fernández García (iD) https://orcid.org/0000-0003-0943-756X
Ph.D., Universidad Tecnica de Ambato, Quito, Ecuador
e-mail: garciafernandez.c@gmail.com

Cristina Pérez Rico (iD) https://orcid.org/0000-0002-4154-5483
Ph.D., Escuela Politecnica Nacional, Quito, Ecuador, e-mail: cristina.perez@epn.edu.ec

Abstract

The aim of this paper was to investigate the relationship between countries' PISA study results from 2018 and a set of indices related to socio-economic inequality, such as the Gini index, human development index, or gender inequality index, along with purely economic variables, such as GDP per capita and government expenditure on education. The study covered 70 countries, consisting of 37 OECD countries and 33 non-OECD countries. Research methods included multivariate linear regression models, k-means clustering, and hierarchical clustering. Our findings revealed that the Gini index was statistically insignificant, indicating income inequality had little effect on students' PISA performance. On the other hand, the gender inequality index was the single most statistically significant explanatory variable for both OECD and non-OECD countries. Therefore, our recommendation for policymakers is simple: increase students' PISA performance, thus enhancing countries' human capital and competitiveness, and focus on decreasing gender disparity and the associated loss of achievement due to gender inequality.

Keywords: education, gender inequality index, Gini index, inequality, PISA 2018

JEL: I20

Jiri Mazurek, Carlos Fernández García, Cristina Pérez Rico

Introduction

The second paragraph of the United States Declaration of Independence starts as follows: "We hold these truths to be self-evident, that all men are created equal…" Today, this statement means that all humans are equal regardless of their race, gender, age, religion, wealth, intellect, social status, or class. However, in real-world societies, experience teaches us that this noble premise is not satisfied, as various forms of inequality emerge again and again.

In this paper, we focus on the relationship between selected forms of inequality, which commonly occur in each society, and students' performance. Knowledge and skills attained at school are assets of each individual for his/her subsequent life and career. At a national level, the set of skills and knowledge forms a fundamental part of national human capital with an immense impact on national competitiveness in the near future. Therefore, measuring skills and knowledge attained by (formal) education has become more important than ever and has led to many national and international studies, such as TIMSS or PISA. Subsequently, the results of these studies have been extensively studied not only by academics, teachers, and other experts from the field of education, but also by policymakers who shape educational systems in each country and decide about its resources, goals, rules of operation, and other parameters.

The majority of the literature on students' performance in international studies focuses on the link between mostly economic variables, such as national GDP, government expenditure on education and teachers' salaries, and educational system parameters, such as teacher/student ratio, curriculum, and the number of schooling years, etc., on the one hand, and students' scores on the other, in order to assess the efficiency of educational systems.

However, studies on the relationship between students' performance and inequality are quite rare (with the exception of the Gini index). Therefore, the aim of this paper is to present the results of the analysis of the relationship between the PISA international study results from 2018 and a set of socio-economic indicators related to inequality on a national level. Specifically, we use income inequality expressed by the Gini index, gender inequality expressed by the gender inequality index, and educational inequality expressed by the inequality in education index. Our working hypothesis can be formulated as follows: Inequality measured by the above-mentioned indices statistically significantly relates to PISA 2018 scores. Moreover, we examine OECD countries, non-OECD countries, and Central and Eastern Europe (CEE) countries, both altogether and separately.

The data for our analysis were gathered mostly from the OECD, UNDP, and the World Bank databases, and the applied research methods include linear (multivariate) regression models and cluster analysis (k-means clustering and dendrograms).

The paper is organized as follows: a literature review is provided in Section 1, the PISA study is briefly described in Section 2, the data and research methods are described in Section 3, while results are provided in Section 4. A Discussion and Conclusions end the article.

Literature review

Recent large international studies, such as PISA (*Programme for International Student Assessment*) and TIMSS (*Trends in International Mathematics and Science Study*), of students' skills and knowledge in several areas, such as reading, mathematics, or science, provide a unique opportunity for international comparisons. Studies on the efficiency of education both on a micro-level, including individual schools and courses, and macro-level, including countries, regions, or districts, are relatively common, see, e.g., Hanushek (1986), Lockheed and Hanushek (1994), Clements (2002), Afonso and Aubyn (2005), Aubyn et al. (2009), Aristovnik and Obadic (2014), or Flores (2017). However, there are few studies that examine the direct links between students' outcomes in standardized tests (such as PISA and TIMSS) and inequality.

Clements (2002) compared selected countries in terms of their expenditure per student and teacher-to-student ratio as input variables, and scores in the international standardized test (TIMSS) as an output variable. He concluded that the resources were not used effectively. Duru-Bellat and Suchaut (2005) studied PISA 2003 and reported that a high degree of social inequality translated into a high dispersion of overall test scores. Afonso and Aubyn (2006) used the data envelopment analysis to assess the efficiency of expenditure for 25 countries participating in the PISA 2003 study. They found that expenditure was highly inefficient. Further, they showed that the GDP per capita and parents' educational attainment were highly and significantly correlated to PISA scores; hence wealthier and more cultivated environments enhanced students' performances.

Hanushek and Woessmann (2011) performed an extensive study on the influence of economic factors on educational achievements with the use of PISA and TIMSS data. They found a strong relationship between the cognitive skills of tested students and the economic growth of their respective countries. Nicoletti and Rabe (2012) found that the rise in the expenditure per pupil of £1,000 in the United Kingdom led to an increase in test scores of about 2% of a standard deviation. This effect was small, but statistically significant. Agasisti (2014) used PISA scores to compare the spending efficiency on education in 20 European countries during the period 2006–2009. He found that teachers' salaries and Internet use played a positive role in affecting educational performance, while GDP per capita was negatively related to efficiency. French, French, and Li (2015) found a strong positive association between public expenditure on education and PISA scores at the country level. Wolff (2015) found a positive effect of secondary education spending on both PISA mathematics and literacy scores (both significant at $p = 0.01$). In addition, primary school spending was also a significant factor in explaining PISA's literacy scores.

DiCorrado, Kelly, and Wright (2015) used linear regression models with PISA 2012 scores as an output variable and GDP per capita, GDP growth, literacy, and spending on education as input variables. They found a positive and statistically significant relationship between GDP per capita and PISA scores, but spending on education was

found to be statistically insignificant. Yorulmaz, Colak, and Ekinci (2017) studied the relationship between PISA 2015 scores and income inequality. Their results suggested income inequality was one of the variables influencing the PISA achievements of students, but several countries with high-income inequality scored above average results, nonetheless. Mazurek and Mielcová (2019) examined the relationship between PISA 2015 scores and a set of socio-economic indicators, such as GDP, expenditure on education, or the democracy index. They found GDP and the democracy index to be statistically significant and directly proportional to PISA scores, while expenditure was negatively correlated with PISA scores. In addition, they showed that the relationship differed between OECD and non-OECD countries. Miningou (2019) focused on the efficiency of public spending in developing countries. She estimated that, on average, sixteen percent of the public financial resources dedicated to education in developing countries were wasted because of inefficiencies.

Programme for International Student Assessment (PISA)

The Programme for International Student Assessment, organized by the Organization for Economic Cooperation and Development (OECD), is a worldwide study of pupils between the age of 15 years and 3 months and 16 years and 2 months. The first PISA study was conducted in 2000, and after 2000 it was repeated every 3 years. The last study took place in 2018, with the results of the study being published in OECD (2019). Altogether, 78 countries or territories participated in the PISA study.

From each country, 4,000 to 8,000 pupils participated in the test. Only students attending schools (home-schoolers were excluded from the study) were tested. In less populated countries, such as Iceland or Luxembourg, the entire age cohort was selected. In total, around 600,000 students, out of a total of 32 million students of the same age in the world, took the PISA 2018 test. As for the continental representation, 42 countries were from Europe, 21 from Asia, six from North America, six from South America, two from Australia and Oceania, and one from Africa. China's sample consisted of pupils from four cities, Beijing, Shanghai, Jiangsu, and Zhejiang.

The cognitive test lasted two hours and had three main parts: mathematics, science, and reading. The test was computer-based with multiple-choice or open (full) answers. After the test, students filled in a questionnaire on their background. Detailed methodology of the pupils' sampling, formulation of the test questions, and other features of the PISA tests can be found on the PISA OECD webpage (OECD n.d.).

The best results, in mathematics, reading, and science combined, were achieved by China (579 points), followed by Singapore (556), Macao (542), Hong Kong (531), and Estonia (525).

Data and method

We used national data mainly from the United Nations Development Programme (UNDP) and World Bank (WB) databases. PISA data were retrieved from the OECD (2019). From a total of 78 countries that participated in PISA 2018, eight countries or territories were excluded due to an insufficient amount of data: Bosnia, Brunei, Hong-Kong, Kosovo, Macao, Montenegro, Morocco, and Taiwan.

The dependent variable was the national mean PISA 2018 score, the average of mathematics, reading, and science. Independent (explanatory) variables included:

- *GDP per capita* (GDP), purchase power parity, in US dollars, from WB (2020).
- *Government expenditure on education* (EXP) in % of national GDP, from WB (2020).
- *Gender inequality index* (GII), which is a composite index quantifying the loss of achievement within a country due to gender inequality. It has three dimensions: reproductive health, empowerment, and labor market participation. GII attains values 0 to 1, where 0 means perfect equality and 1 perfect inequality, from UNDP (2020).
- *Gini coefficient* (GINI), which expresses income inequality on a scale from 0 to 1, where 0 means perfect equality and 1 perfect inequality, from World Bank (2020).
- *Human Development Index* (HDI), which is a composite index of three main dimensions: life expectancy, education, and per capita income. It attains values from 0 to 1, where values close to 0 mean very low human development, while values near 1 mean very high human development, from UNDP (2020).
- *Inequality in education* (EDU), which is an index describing inequality in the distribution of years of schooling based on data from household surveys. The higher the EDU, the greater the inequality, from UNDP (2020).

Altogether, six independent variables were selected for further analysis. Two of them have an economic nature (GDP and EXP) and were found to be important (statistically significant) in previous studies. One variable, HDI, is associated with national development and well-being, and the last three variables are inequality indices (GII, GINI, and EDU), where each index 'captures' a different form of inequality. The EDU and GII indices have not been used in the context of the PISA study before. All data from the year 2018 were retrieved from respective databases with the exception of GDP per capita, which was averaged over the 2014–2018 period, and the Gini index, where the last known value for a given country was used when the 2018 value was not available. Detailed descriptions of the GII and HDI indices can be found in UNDP (2013). For an explanation of the Gini index, see Gini (1921), and for EDU, see UNDP (2020).

A linear relationship between pairs of variables in the form of a correlation matrix is shown in Table 2, along with a heat map in Figure 1. The results of the correlation analysis with respect to PISA showed that HDI was the most positively correlated variable, while GII was the most negatively correlated variable, see also Figures 3 a)–b).

Table 1. Data: descriptive statistics

Variable	Mean (st. deviation) All countries	Mean (st. deviation) OECD countries	Mean (st. deviation) non-OECD countries
PISA	457.91 (50.84)	488.24 (26.73)	423.91 (50.1)
EDU	8.51 (5.99)	6.69 (5.74)	10.55 (5.67)
EXP	4.75 (1.45)	5.21 (1.18)	4.23 (1.55)
GDP	37447 (23796)	45597 (18097)	28309 (26260)
GII	0.192 (0.128)	0.120 (0.090)	0.271 (0.117)
GINI	0.350 (0.069)	0.333 (0.059)	0.370 (0.074)
HDI	0.847 (0.071)	0.894 (0.047)	0.795 (0.054)

Source: authors compilation.

Table 2. Correlation matrix

Variable	PISA	GINI	GII	GDP	EXP	EDU	HDI
PISA	1.0000	−0.4377	−0.8204	0.6281	0.2875	−0.4689	0.7523
GINI	−0.4377	1.0000	0.5780	−0.2100	−0.2823	0.6312	−0.3800
GII	−0.8204	0.5780	1.0000	−0.6780	−0.4309	0.5322	−0.8423
GDP	0.6281	−0.2100	−0.6780	1.0000	0.2755	−0.3642	0.8275
EXP	0.2875	−0.2823	−0.4309	0.2755	1.0000	−0.2611	0.5027
EDU	−0.4689	0.6312	0.5322	−0.3642	−0.2611	1.0000	−0.5437
HDI	0.7523	−0.3800	−0.8423	0.8275	0.5027	−0.5437	1.0000

Source: authors compilation.
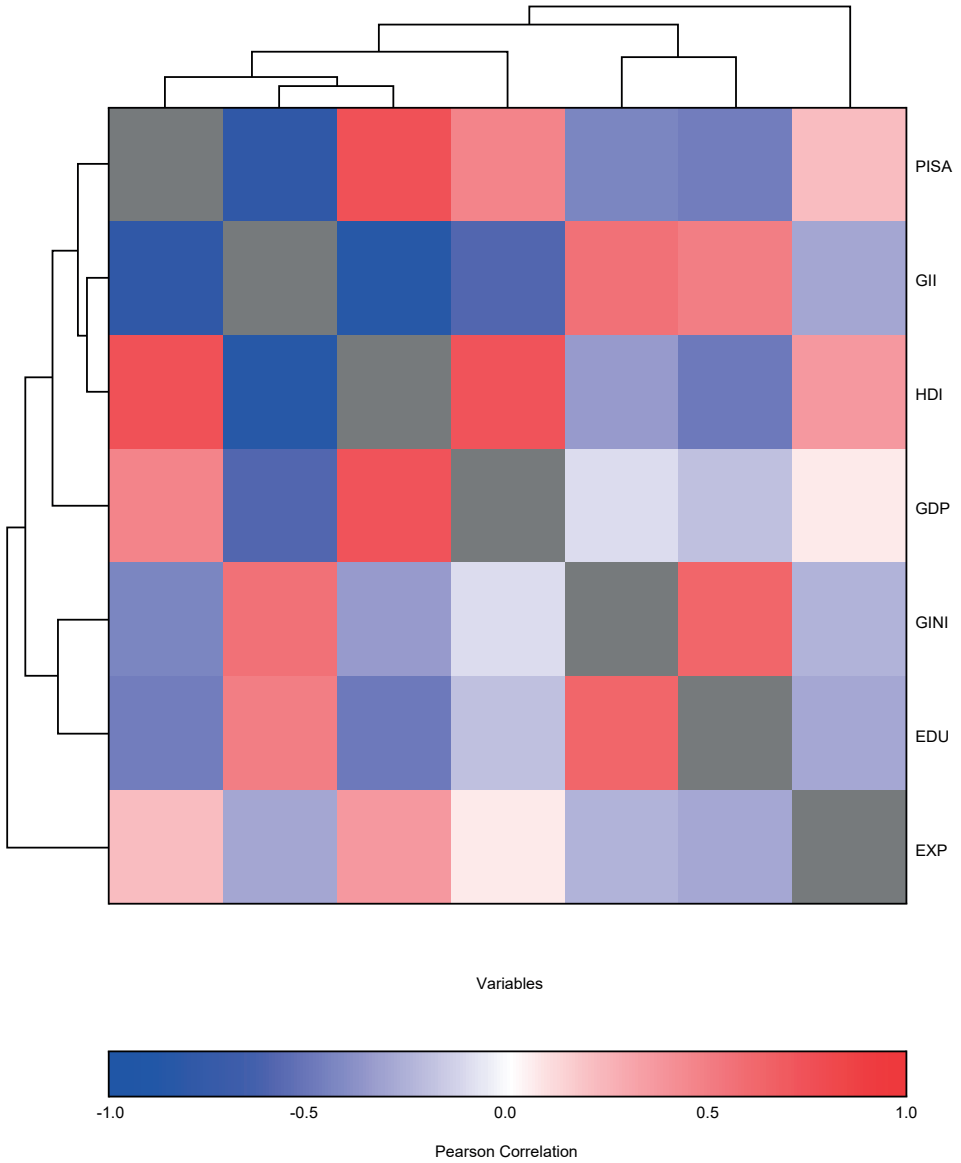
Heat Map of the Pearson Correlation Matrix

Figure 1. The heat map of the Pearson correlation coefficient for variables in Table 1.
Source: authors compilation.

To examine the relationship between the PISA 2018 scores and other variables, we used a multivariate linear model:

$$y_i = \beta_0 + \beta_1 x_{1i} + \ldots + \beta_p x_{pi} + \varepsilon_i$$

where y denotes the dependent (explained) variable, $x = (x_1, \ldots, x_p)$ is a vector of p independent (explanatory) variables, i denotes individual countries, β_i are regression coefficients, and ε is the error term.

We performed the linear regression via the free Gretl statistical software, applying the OLS method with corrected heteroscedasticity, a built-in feature of Gretl.

In addition, we used the k-means clustering method and hierarchical method (dendrograms) to examine similarities among countries via NCSS 2020 statistical software. An explanation of the k-means clustering method can be found in MacKay (2003), while the hierarchical method is well described, e.g., in Nielsen (2016).

Results

Multivariate linear regression models

At the beginning, we examined the multicollinearity of the data. According to Kennedy (2003), multicollinearity does not affect the reliability of a regression model as a whole, but it affects individual predictors. A standard measure of multicollinearity is the variance inflation factor (*VIF*), see O'Brien (2007): $VIF(i) = \dfrac{1}{1 - R_i^2}$, where

R_i^2 is the coefficient of determination of a regression of a variable (i) on other variables in the model. According to O'Brien (2007), values $VIF(i) > 10$ indicate a problem of multicollinearity. The variance inflation factor for all independent variables is provided in Table 3. No value of VIF exceeds 10; therefore, multicollinearity is not a problem.

Table 3. The variance inflation factor for independent variables

Variable	GINI	GII	GDP	EXP	EDU	HDI
VIF	2.353	5.055	2.471	1.298	2.044	6.374

Source: authors compilation.

For the linear regression, we first consider the following Model 1 with six independent variables:

$$\begin{aligned} \text{PISA}_i = \beta_0 &+ \beta_1 \text{GINI}_i + \beta_2 \text{GII}_i + \beta_3 \text{GDP}_i + \\ &+ \beta_4 \text{EXP}_i + \beta_5 \text{EDU}_i + \beta_6 \text{HDI}_i + \varepsilon_i \end{aligned} \tag{1}$$

The results are shown in Table 4. The only variable found statistically insignificant was (somewhat surprisingly) the GINI index. Negative signs for GII and EDU indi-

cate indirect proportionality between PISA results on the one hand, and gender and education inequality on the other. The positive sign for HDI means that it is directly proportional to PISA, which could be expected. However, contrary to expectations, negative signs of GDP and EXP mean that the richer the country, or the higher its expenditure on education, the lower the PISA results.

In Model 2, we omitted the least significant variable in Model 1, the GINI index:

$$\text{PISA}_i = \beta_0 + \beta_1 \text{GII}_i + \beta_2 \text{GDP}_i + \beta_3 \text{EXP}_i + \beta_4 \text{EDU}_i + \beta_5 \text{HDI}_i + \varepsilon_i \qquad (2)$$

As can be seen from Table 4, Model 2 is slightly better than Model 1, the regression coefficients are similar to those in Model 1, and all signs are preserved. Again, all independent variables are highly statistically significant.

In the following Model 3, we omit both the GINI and EDU (the least significant variables from Model 2) variables:

$$\text{PISA}_i = \beta_0 + \beta_1 \text{GII}_i + \beta_2 \text{GDP}_i + \beta_3 \text{EXP}_i + \beta_4 \text{HDI}_i + \varepsilon_i \qquad (3)$$

In this case, the adjusted coefficient of determination $R^2 = 0.819$ is slightly lower, but the explanatory power of Model 3 is still high. All signs of regression coefficients remain the same as in Models 1–2.

To illustrate that the economic variables cannot be omitted from linear models, we construct the last Model 4:

$$\text{PISA}_i = \beta_0 + \beta_1 \text{GINI}_i + \beta_2 \text{GII}_i + \beta_3 \text{EDU}_i + \beta_4 \text{HDI}_i + \varepsilon_i \qquad (4)$$

Model 4 is clearly inferior to the previous models. Interestingly, the Gini coefficient is still statistically insignificant. On the other hand, the GII was statistically significant, at least at the $p = 0.01$ level in all considered models.

Since the data sample consisted of 37 OECD countries and 33 non-OECD countries, we examined both subsamples separately. Although the subsamples are smaller, they are more homogeneous and might reveal new insights (see Mazurek and Mielcová 2019). The mean values of all variables for both groups of countries are provided in Table 5. OECD countries have, on average, higher PISA scores, lower income, gender, and educational inequality, a higher development index as well as higher GDP per capita and expenditure on education.

Linear regression was performed for both subsamples separately, and the results are reported in Table 4. Models 5–7 are associated with OECD countries, models 8–10 with non-OECD countries, and models 11–14 deal with CEE countries.

In the case of data from OECD countries, linear regression models fitted the data very well, with the adjusted coefficient of determination above 0.82. The goodness-of-fit of the data from non-OECD countries was significantly worse, with the adjusted R^2 between 0.48 and 0.61. Another noticeable difference between both samples is related

to the GDP, EXP, and HDI variables, which are highly statistically significant in the case of OECD countries, but rather statistically insignificant for the non-OECD countries. The GII is statistically highly significant in (almost) all examined models. Figures 2 a)–f) graphically illustrate selected (dis)similarities between both sets of countries.

As for the fifteen CEE countries, only two and three regressors were applied to avoid overfitting of the data. Model 13 provides the best fit in the case of two explanatory variables, while Model 14 provides the best fit in the case of three explanatory variables. In both cases, the adjusted R^2 exceeded 0.9, and the EXP and HDI variables were highly statistically significant.

Table 4. Results of multivariable linear regression models

| Model no./ sample	Sample size; adj. R^2	const.	GINI	GII	GDP	EXP	EDU	HDI
1 /All	70; 0.844	326.39***	19.87	−226.8***	−0.00045***	−6.738***	−1.601**	274.67***
2 /All	70; 0.851	310.41***	-	−211.3***	−0.00043***	−6.503***	−1.475**	293.79***
3 /All	70; 0.819	267.35***	-	−224.8***	−0.00049***	−5.357***	-	329.01***
4 /All	70; 0.629	343.22***	47.66	−210.6***	-	-	−1.446*	175.59*
5 /OECD	37; 0.899	411.30***	−121.1*	−110.8**	−0.00030***	−4.787**	−0.974	194.79*
6 /OECD	37; 0.875	291.6***	−148.1**	−82.33*	−0.00037***	−6.382***	-	342.25***
7 /OECD	37; 0.823	205.0***	-	−133.2***	−0.00045***	−5.546***	-	390.36***
8 /non O	33; 0.606	466.3***	1.498	−276.2***	−0.00048	−3.970*	0.172	73.50
9 /non O	33; 0.556	505.5***	-	−255.9***	−0.00033	−4.224	−0.803	26.19
10 /non O	33; 0.484	521.3***	-	−251.2***	−0.00003	−3.646	−1.238	-
11/CEE	15; 0.772	555.3***	-	−249.6***	-	-	−9.218***	-
12/CEE	15; 0.738	259.8***	-	-	0.00335***	25.913**	-	-
13/CEE	15; 0.960	−89.7***	-	-	-	20.556	-	556.21
14/CEE	15; 0.989	−80.5	-	-	-	20.925***	0.394	536.88***

Note: * statistically significant at the 0.10 level, ** significant at the 0.05 level, *** significant at the 0.01 level
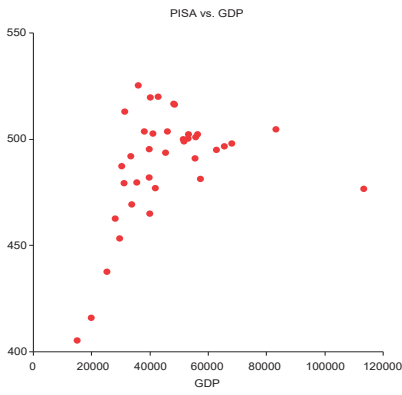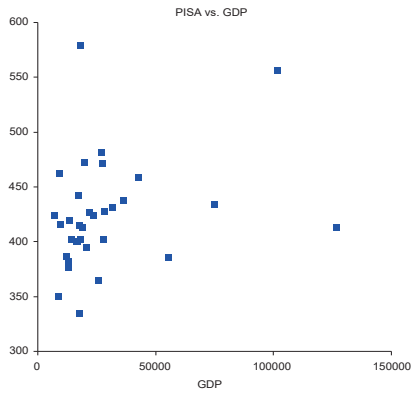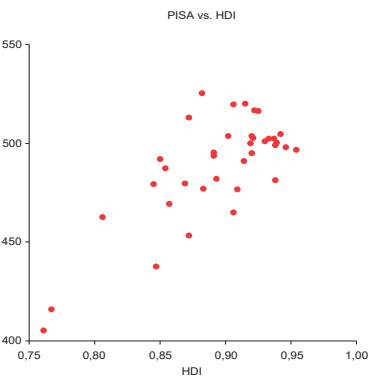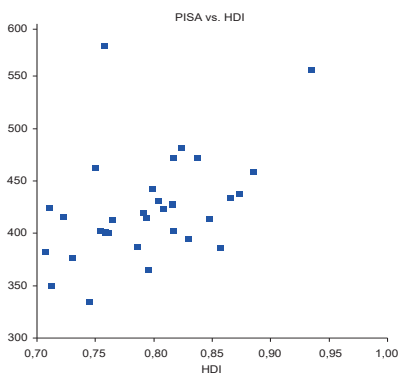Source: authors compilation.
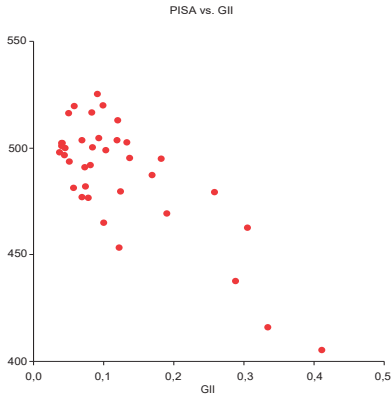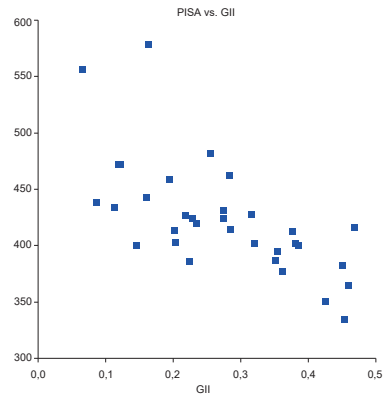
Figures 2 a)-f). Scatter plots of selected variables for OECD countries (left side, red) and non-OECD countries (right side, blue)
Source: authors compilation.

K-means clustering

K-means clustering of 70 countries was performed via NCSS 2020. For the clustering, all seven variables (PISA, EDU, EXP, HDI, GDP, GII, and GINI) were applied. The optimal number of clusters was determined by the "elbow rule." Table 5 provides a summary of all clusters with mean values of all variables.

- Cluster 1, which includes both OECD and non-OECD countries, is characterized by extremely high GDP per capita, high gender equality, and average PISA results.
- Cluster 2 consists of both OECD and non-OECD countries with relatively high PISA scores, while other variables are close to average.
- Cluster 3 contains almost exclusively OECD countries (with the only exception being Malta) with the highest PISA scores on average, with a high human development index, the highest expenditure on education, and the lowest values of inequality in terms of GINI and GII.
- Cluster 4 is formed by non-OECD countries with low PISA results, the lowest GDP per capita, and average inequality, both in terms of GINI and GII.
- Cluster 5 groups together non-OECD countries with the lowest PISA scores and the highest inequality in terms of GINI, GII, and EDU.

To illustrate the distribution of clusters graphically, see Figures 3 a)–d). For example, from Figure 3 c) it can be seen that Cluster 1 (red squares) is characterized by an aver age human development index but very high GDP per capita. Similarly, from Figure 3 d) it is obvious that countries from Cluster 5 (black diamonds) typically have a low human development index and high gender inequality.

Table 5. Clusters – summary

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Coun-tries	Luxembourg, Qatar, Singapore, UAE,	Belarus, Croatia, Cyprus, Greece, Hungary, China, Italy, Japan, S. Korea, Latvia, Lithuania, Poland, Portugal, Russia, Slovakia, Spain	Austria, Australia, Belgium, Canada, Czechia, Denmark, Estonia, Finland, France, Germany, Iceland, Ireland, Israel, Malta, Netherlands, New Zealand, Norway, Slovenia, Sweden, Switzerland, UK, USA	Albania, Argentina, Azerbaijan, Bulgaria, Georgia, Kazakhstan, Lebanon, Moldova, North Macedonia, Romania, Serbia, Ukraine, Uruguay	Brazil, Colombia, Costa Rica, Dominican Rep., Chile, Indonesia, Jordan, Malaysia, Mexico, Panama, Peru, Philippines, Saudi Arabia, Thailand, Turkey
PISA	470	488	498	415	401
GINI	0.386	0.328	0.310	0.335	0.438
GII	0.115	0.132	0.085	0.266	0.368
GDP	104210	32407	51764	17617	21208

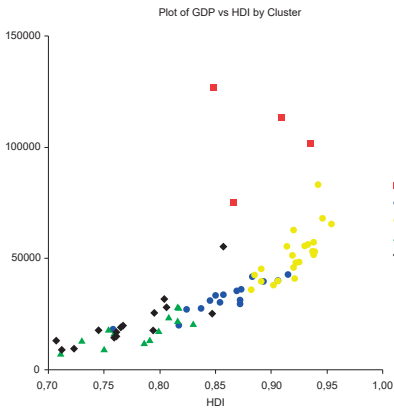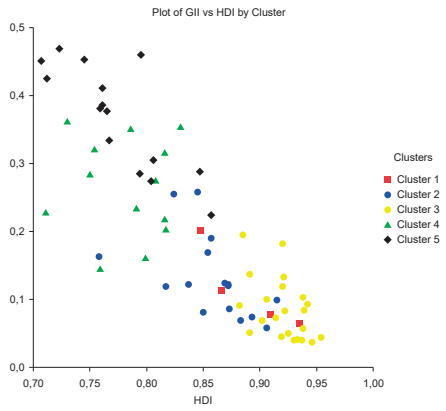|	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
EXP	2.828	4.358	5.883	4.391	4.329
EDU	12.25	8.01	3.88	6.59	16.51
HDI	0.890	0.858	0.921	0.782	0.774

Source: authors compilation.
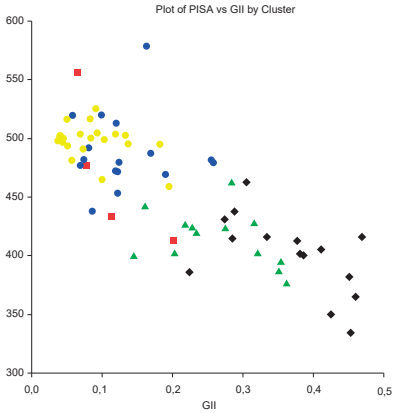
a)
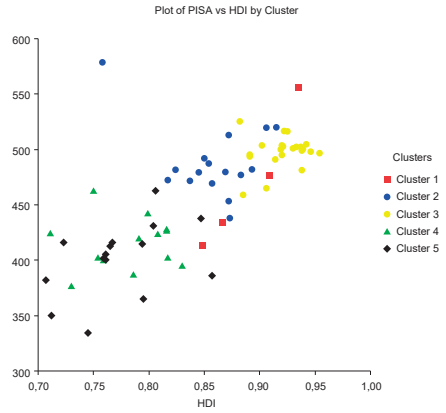
b)

c)

d)

Figures 3 a)–d). Scatter plots of selected variables
Source: authors compilation.

Hierarchical clustering

For the hierarchical clustering (a tree diagram) of 70 countries, all seven variables (PISA, EDU, EXP, HDI, GDP, GII, and GINI) were applied. The result is shown in Figure 4.

Countries are displayed on the vertical axis and are gradually, from left to right, connected by nodes into growing clusters with respect to their similarity (relative distance). The distance of a node from the vertical axis is called *height*, and it is provided on the horizontal axis. The more two objects connected by a node are similar, the lower the height of the node.

From Figure 4, it is clear that the two closest (most similar) countries are Sweden and Denmark, followed by the UK and Australia. Interestingly, China is grouped with the United Arab Emirates, while the USA is, in the first step, included in a cluster with New Zealand and Israel. In addition, it can be noticed that towards the end of the diagram, three clusters are formed. The first cluster consists of China and the UAE only, the second cluster is formed almost exclusively by OECD countries, and the last cluster (the bottom half of the diagram) is formed by mostly non-OECD members. Yet, both large clusters are closer to each other than to the cluster containing China and the UAE. These findings support the division of countries into OECD and non-OECD groups in Section 4.1.
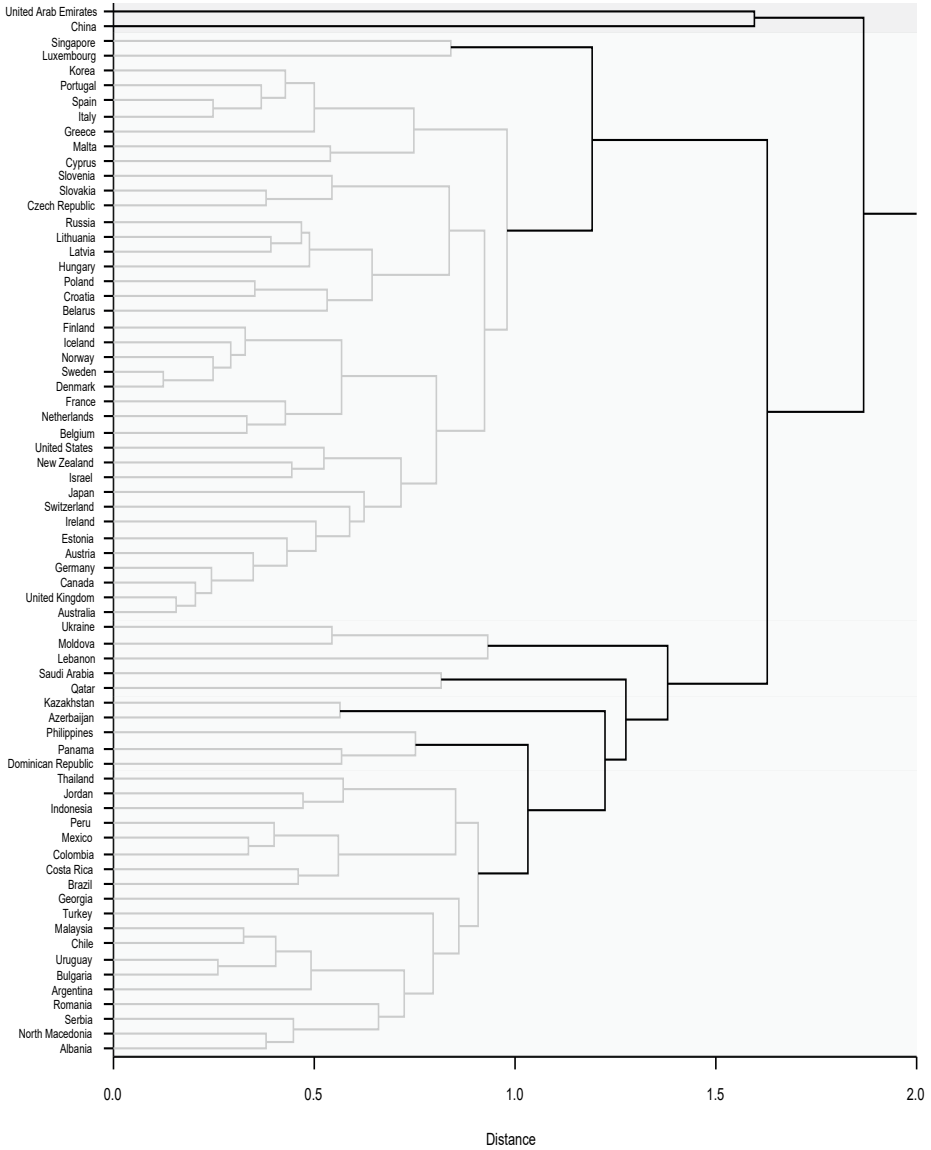
Dendrogram of Country

Figure 4. Hierarchical clustering (dendrogram)
Source: authors compilation.

Discussion

The aim of our study was to examine whether three forms of inequality (income, gender, and educational inequality) have a statistically significant impact on students' performance in the 2018 PISA international study.

Multivariate linear regression applied to data from all 70 countries revealed that gender and educational inequalities (GII and EDU) were highly statistically significant, but income inequality (GINI) was statistically insignificant. In addition, both economic variables (GDP and EXP) and the human development index (HDI) were statistically significant. Both regression coefficients of gender and educational inequality were, as expected, negative, which means the lower the inequality, the higher the PISA score. In the case of gender inequality, the estimated regression coefficient was between –226 and –210. This can be interpreted in the following way: when the gender inequality index GII decreases by 0.10, PISA scores increase by 21 to 22.6 points. Similarly, if the inequality in the education index EDU decreases by 1 point, PISA scores increase by 1.44–1.60 points.

The regression coefficients of the economic variables GDP and EXP attained negative values as well, while being highly statistically significant. This means the higher the GDP per capita or expenditure on education, the lower the PISA scores. In the case of expenditure, a similar result was reported in several previous studies, see, e.g., Afonso and Aubyn (2006) or Mazurek and Mielcová (2019), and it is related to expenditure inefficiency. As for GDP per capita, the wealth of a nation does not give obvious evidence of the quality of its educational system. Estonia achieved scores well above average in PISA 2018, though its GDP per capita is well below the average of OECD countries. On the other hand, the PISA scores of very rich countries, such as Qatar or the UAE, are situated in the bottom half of the list. Finally, a multicollinearity effect in linear regression models cannot be ruled out either.

It is important to note that the data sample of countries participating in the PISA 2018 study was very non-homogeneous. The sample contained very developed countries, at level IV according to Rosling's classification (see Rosling 2018), but also rather poor countries, such as the Philippines (at level II-III) with GDP per capita approximately 14 times lower than that of the richest countries. That is why dividing the whole set of countries into two subsets, OECD countries (on average richer and more developed) and non-OECD countries, might lead to more conclusive results. This approach can also be supported by the dendrogram in Figure 4, where countries "naturally" formed two large clusters (and one mini-cluster of China and the UAE), with 33 out of 37 OECD countries grouped together with six non-OECD countries in one of the two large clusters.

After the division, a different pattern emerges – see Table 4, Model 5. In the case of OECD countries, the most significant variables are GDP per capita, the gender inequality index, and expenditure on education. Inequality in education is insignificant. OECD countries are characterized by very low inequality in general; hence students' PISA performance most likely depended more strongly on economic variables. In the case of non-OECD countries – see Model 8, Table 4 – the only statistically significant

variable at the $p = 0.01$ level is gender inequality. Expenditure on education is significant at the $p = 0.10$ level, and all other variables, including GDP per capita, are insignificant. This might be at least partially explained by the fact that non-OECD countries differ significantly in inequality – see Table 1 – which was most likely reflected in Model 8. Additionally, the sample of non-OECD countries is still very diverse (Qatar or the UAE versus the Philippines or Jordan) in terms of culture, economy, or human development, hence the predictive power of linear models is weaker, and conclusions drawn from the models are less clear and conclusive.

In the case of the CEE countries (former socialist countries of the post-World War II Eastern bloc), the expenditure on education and human development indicators were the two most statistically significant ($p = 0.01$ level) explanatory variables, see Models 13–14. The higher the expenditures on education and human development indicators, the higher the PISA scores. Income and gender inequality were statistically insignificant, which might be due to historical reasons –socialism was known for its egalitarianism concerning income and gender, and this feature likely persists up today. On the other hand, former socialist countries significantly differ by their economic power and their overall level of development (consider, e.g., the Czech Republic or Slovenia versus Romania or Bulgaria). This non-homogeneity is corroborated by the cluster analysis in section 4.2, where the CEE countries were divided into Clusters 3, 4, and 5, with different levels of GDP per capita. That is why this kind of inequality is probably reflected in the expenditures on education and human development indicators.

Therefore, we conclude that our findings indicate that there might be no universal answer to the question of how to enhance students' PISA performance (or students' performance in general) for all countries because each country is unique in terms of its education system, economic level, and culture, etc. In countries at Rosling's level I and II, where people struggle with basic needs, such as food supply or shelter, and pupils and students can hardly afford textbooks, the financial aspect will likely prevail. In richer societies (levels III and IV), students' skills and knowledge can be boosted by other means than just in the form of finances invested into an educational system. Nevertheless, if a general answer to the question stated above has to be provided, then it is gender inequality. Our analysis indicates that decreasing gender inequality in both OECD and non-OECD countries would lead to substantially better PISA results.

The limitations of the study rest mainly in the fact that data were not available for all 78 countries or territories that participated in the PISA 2018 study, as eight countries had to be excluded through a lack of data. Africa was strongly under-represented, as only Morocco took part in the PISA 2018 study, but, unfortunately, it had to be excluded from this study due to the lack of other data. Additionally, the Gini index in the year 2018 was not available for several countries, and was replaced by older estimates from 2013. Moreover, it should be noted that inequality is a complex phenomenon that cannot be completely reduced to the three (currently existing and widely available) inequality indices applied in this study, and further progress on the quantitative measuring of inequality in society is necessary.

Conclusions

The aim of the paper was to present the results of the analysis of the relationship between inequality and students' performance in the PISA 2018 study.

Our results indicate that inequality, described in terms of income inequality, gender inequality, and education inequality, can indeed explain (along with three other, rather economic variables) the majority of observed data variance across countries. The most important and most statistically significant variable was the gender inequality index GII. The higher the GII, the lower the PISA scores. By contrast, the Gini index expressing income inequality was found to be insignificant as an explanatory variable. The inequality expressed by the education index EDU was statistically significant only in models encompassing all countries, but not in separate models for OECD countries and non-OECD countries, respectively. In the case of the CEE countries, only economic inequality expressed by the government expenditures and human development index was found to be highly statistically significant.

Moreover, we found that the relationship between GDP per capita and expenditure on education on the one hand, and PISA scores on the other, was statistically significant, but had a negative sign. Hence, the more money that is spent on education, the worse the educational outcomes. These somewhat counter-intuitive findings are in accord with previous studies, which reported that financial resources for education are expended inefficiently in many countries, and are prone to being wasted, see, e.g., Afonso and Aubyn (2006), Agasisti (2014), or Mazurek and Mielcová (2019).

Therefore, our recommendation to policymakers (which we consider the most important added value of the presented study) is simple: focus on decreasing gender disparity and the associated loss of achievement due to gender inequality, and PISA scores, as well as the associated skills and knowledge of students, will likely significantly improve in the majority of countries.

Acknowledgments

References

Afonso, A., Aubyn, M.S. (2005), *Non-parametric Approaches to Education and Health Expenditure Efficiency in the OECD*, "Journal of Applied Economics", 8 (2), pp. 227–246, https://doi.org/10.1080/15140326.2005.12040626

Afonso, A., Aubyn, M.S. (2006), *Cross-country efficiency of secondary education provision: a semi-parametric analysis with non-discretionary inputs*, "Economic Modeling", 23 (3), pp. 476–491, https://doi.org/10.1016/j.econmod.2006.02.003

Agasisti, T. (2014), *The Efficiency of Public Spending on Education: an empirical comparison of EU countries*, "European Journal of Education", 49 (4), pp. 543–557, https://doi.org/10.1111/ejed.12069

Aristovnik, A., Obadic, A. (2014), *Measuring relative efficiency of secondary education in selected EU and OECD countries: the case of Slovenia and Croatia*, "Technological and Economic Development of Economy", 20 (3), pp. 419–433, https://doi.org/10.3846/20294913.2014.880085

Aubyn, M.S., Pina, Á., Garcia, F., Pais, J. (2009), *Study on the efficiency and effectiveness of public spending on tertiary education*, "Economics Papers", No. 390, European Committee, Brussels, https://doi.org/10.2765/30348

Clements, B. (2002), *How efficient is education spending in Europe?*, "European Review of Economics and Finance", 1 (1), pp. 3–26.

DiCorrado, E., Kelly, K., Wright, M. (2015), *The Relationship Between Mathematical Performance and GDP per Capita*, https://smartech.gatech.edu/bitstream/handle/1853/54222/the_relationship_between_mathematical_performance_and_gdp_per_capita_1.bk-2.pdf (accessed: 10.02.2020).

Duru-Bellat, M., Suchaut, B. (2005), *Organisation and Context, Efficiency and Equity of Educational Systems: What PISA Tells Us*, "European Educational Research Journal", 4 (3), pp. 181–194, https://doi.org/10.2304/eerj.2005.4.3.3

Flores, I. (2017), *Modelling efficiency in education: how are European countries spending their budgets and what relation between money and performance*, "Sociologia, Problemas e Práticas", 83, pp. 157–170, https://doi.org/10.7458/SPP2017836496

French, J.J., French, A., Li, W.-X. (2015), *The relationship among cultural dimensions, education expenditure, and PISA performance*, "International Journal of Educational Development", 42, pp. 25–34, https://doi.org/10.1016/j.ijedudev.2015.02.010

Gini, C. (1921), *Measurement of Inequality of Incomes*, "The Economic Journal", 31 (121), pp. 124–126, https://doi.org/10.2307/2223319

Hanushek, E.A. (1986), *The economics of schooling: production and efficiency in the public schools*, "Journal of Economic Literature", 24 (3), pp. 1141–1178.

Hanushek, E.A., Woessmann, L. (2011), *The economics of international differences in educational achievement*, [in:] E.A. Hanushek, S. Machin, L. Woessmann (eds.), *Handbook of the Economics of Education*, 3, North Holland, Amsterdam, pp. 89–200, https://doi.org/10.1016/B978-0-444-53429-3.00002-8

Kennedy, P. (2003), *A Guide to Econometrics*, 5th edition, MIT Press, Cambridge.

Lockheed, M., Hanushek, E.A. (1994), *Concepts of educational efficiency and effectiveness*, "Human Resources Development and Operations Policy Working Papers", No. HRO 24, World Bank, Washington.

MacKay, D. (2003), *An Example Inference Task: Clustering*, [in:] D. MacKay, *Information Theory, Inference and Learning Algorithms*, Cambridge University Press, Cambridge, pp. 284–292.

Mazurek, J., Mielcová, E. (2019), *On the relationship between selected socio-economic indicators and student performances in the PISA 2015 study*, "E+M", 22 (2), pp. 22–39, https://doi.org/10.15240/tul/001/2019-2-002

Miningou, E.W. (2019), *Quality Education and the Efficiency of Public Expenditure: A Cross-Country Comparative Analysis*, "Policy Research Working Paper", 9077, pp. 1–22, https://doi.org/10.1596/1813-9450-9077

Nicoletti, C., Rabe, B. (2012), *The effect of school resources on test scores in England*, "Research Paper: Institute for Economic and Social Research", https://www.iser.essex.ac.uk/publications/working-papers/iser/2012-13.pdf (accessed: 19.12.2019).

Nielsen, F. (2016), *Hierarchical Clustering*, [in:] F. Nielsen, *Introduction to HPC with MPI for Data Science*, Springer, Cham, pp. 195–211.

O'Brien, R.M. (2007), *A Caution Regarding Rules of Thumb for Variance Inflation Factors*, "Quality and Quantity", 41, pp. 673–690, https://doi.org/10.1007/s11135-006-9018-6

OECD (2019), *PISA 2018 Results. Combined executive summaries. Volume I , II & III*, https://www.oecd.org/pisa/Combined_Executive_Summaries_PISA_2018.pdf (accessed: 12.02.2020).

Rosling, H., Rosling Ronnlund, A., Rosling, O. (2018), *Factfulness: Reasons We're Wrong About the World – and Why Things Are Better Than You Think*, Flatiron Books, New York.

United Nations Development Programme (2013), *Technical notes*, http://hdr.undp.org/sites/default/files/hdr_2013_en_technotes.pdf (accessed: 10.02.2020).

United Nations Development Programme (2020), http://hdr.undp.org/en/2020-report (accessed: 10.02.2020).

Wolff, E.N. (2015), *Educational expenditures and student performance among OECD countries*, "Structural Change and Economic Dynamics", 33, pp. 37–57, https://doi.org/10.1016/j.strueco.2015.02.003

World Bank (2020), *World Bank Open Data*, https://data.worldbank.org/ (accessed: 11.02.2020).

Yorulmaz, Y.I., Colak, I., Ekinci, C.E. (2017), *An evaluation of PISA 2015 achievements of OECD countries within income distribution and education expenditures*, "Turkish Journal of Education", 6 (4), pp. 169–185, https://doi.org/10.19128/turje.329755

Nierówności a wyniki badania umiejętności uczniów PISA 2018: porównanie międzykrajowe

Celem tego artykułu było zbadanie związku między wynikami badania PISA przeprowadzonego w poszczególnych krajach w 2018 r. a zestawem wskaźników związanych z nierównościami społeczno-ekonomicznymi, takimi jak indeks Giniego, wskaźnik rozwoju społecznego czy wskaźnik nierówności płci, oraz ze zmiennymi czysto ekonomicznymi, takimi jak PKB per capita i wydatki rządowe na edukację. Badaniem objęto 70 krajów, w tym 37 krajów OECD i 33 kraje spoza OECD. Metody badawcze obejmowały wielowymiarowe modele regresji liniowej, grupowanie k-średnich i grupowanie hierarchiczne. Wyniki przeprowadzonej analizy wykazały, że wskaźnik Giniego był statystycznie nieistotny, co wskazuje, że nierówności dochodowe miały niewielki wpływ na wyniki uczniów w badaniu PISA. Z drugiej strony, wskaźnik nierówności płci był jedyną najbardziej istotną statystycznie zmienną objaśniającą zarówno dla krajów OECD, jak i spoza OECD. Dlatego nasza rekomendacja dla decydentów jest prosta: należy zwiększyć wyniki uczniów w badaniu PISA, a tym samym osiągnąć poprawę w obszarze kapitału ludzkiego i konkurencyjności krajów, oraz skupić się na zmniejszaniu nierówności płci i związanej z tym utraty osiągnięć edukacyjnych wynikających z nierówności płci.