Econometric: Midterm Paper

Jumpei Kasai

March 7, 2025

1 Introduction

Education is the process in which an individual acquires or imparts basic knowledge to another. A well-educated population contributes to increased productivity, innovation, and overall national competitiveness [Canada, 2025]. However, education quality varies significantly between countries due to economic conditions, government policies, and social factors. Understanding the key determinants of student performance can help policymakers develop effective education policies.

The exam used to measure education quality is the Program for International Student Assessment (PISA), conducted by the Organization for Economic Cooperation and Development (OECD). PISA assesses the reading, mathematics, and science literacy of 15-year-old students in different countries, providing a standardized way to compare education systems globally. The PISA score serves as a benchmark for evaluating student achievement and identifying factors that influence academic performance.

This paper demonstrates how economic and social factors affect PISA scores. Specifically, it examines the relationship between economic development, unemployment, income inequality, and gender differences with student performance. Existing research suggests that higher GDP per capita leads to better education outcomes, while greater income inequality and unemployment may negatively impact student achievement. In addition, gender disparities in educational outcomes have been debated, with some studies suggesting that female students tend to outperform male students in reading but lag behind in mathematics.

To test these relationships, we use a multiple linear regression model to analyze the impact of these variables on PISA scores. By identifying the most significant determinants of student performance, this study provides insight into the economic and social conditions that contribute to education success. The findings may offer valuable implications for policy makers seeking to improve education quality and reduce disparities in learning outcomes in different demographic and economic contexts.

2 Literature Review

2.1 Economy Growth and Education

Hanushek and Woessmann (2012) found that higher levels of GDP per capita are correlated with better educational performance, as wealthier nations tend to invest more in education infrastructure, teacher quality and learning outcomes [Hanushek and Woessmann, 2012]. Also, Barro (2001) argues that economic growth enables governments to allocate more resource to education, improving student performance in standardized assessment like PISA [Barro, 2001].

2.2 Unemployment and Education

Unemployment negatively affect education. Oreopoulos et al. (2008) found that high unemployment rates can discourage students from investing in education, as uncertain job prospects reduce the perceived benefits of higher education [Oreopoulos et al., 2008]. Additionally, Reis and Stockhammer (2019) suggest that regions with persistent unemployment tend to experience underfunded public education systems, leading to lower academic performance among students [Reis and Stockhammer, 2019].

2.3 Income Inequality and Education

Income inequality, as measured by the Gini index, has been widely studied in the context of educational disparities. Checchi (2003) argues that higher income inequality leads to greater disparities in education access and quality, particularly in countries where public education is underfunded [Checchi, 2003]. Corak (2013) further supports this, showing that income inequality reduces intergenerational mobility, making it more difficult for students from low-income families to achieve high academic performance [Corak, 2013].

2.4 Gender Difference in Student Performance

Gender disparities in education have been extensively analyzed in PISA studies. OECD (2019) reports that female students tend to outperform male students in reading, while male students often perform better in mathematics and science. Stoet and Geary (2013) argue that these differences are partly due to social and cultural expectations, which shape students' academic interests and self-confidence in different subjects.

3 Data

3.1 Data Source

The dataset used in this study was obtained from Kaggle, specifically the dataset titled *"PISA Results 2000-2022 - Economics and Education"*, [Tomaz, 2022] .

This dataset contains PISA scores from multiple countries between 2000 and 2022, along with economic and social indicators [OECD, 2025].

PISA is a standardized international assessment conducted by the OECD that evaluates 15-year-old students' proficiency in reading, mathematics, and science. The PISA score serves as the dependent variable in this study, representing student performance across different education systems.

The dataset also includes macroeconomic indicators such as GDP per capita, unemployment rate, and income inequality, allowing for an analysis of how these economic conditions impact student achievement.

3.2 Variables and Descriptions

This study focuses on the following variables:

3.2.1 Dependent Variable

- `rating` - PISA score measuring student performance.

3.2.2 Independent Variables

- `gdp_per_capita_ppp` - GDP per capita in terms of purchasing power parity.
- `unemployment` - Unemployment rate (%).
- `gini_index` - Statistic that measures economic inequality
- `sex_binary` - Dummy variable for gender (Male = 0, Female = 1).

These variables were selected based on prior research, which suggests that economic prosperity, labor market stability, and social equity are crucial factors influencing education outcomes. Additionally, gender differences in academic achievement remain a relevant topic in education research.

3.3 Correlation Matrix

To examine the relationships between the independent variables and the dependent variable (rating), a correlation matrix was generated. The correlation plot in Figure 1 represents these relationships.

The correlation plot indicates several important relationships:

- **GDP per capita** (`gdp_per_capita_ppp`) has a moderate positive correlation (0.41) with `rating`, suggesting that economic prosperity is associated with higher PISA scores.
- **Income inequality** (`gini_index`) exhibits a strong negative correlation (-0.82) with `rating`, indicating that higher inequality is linked to lower student performance.

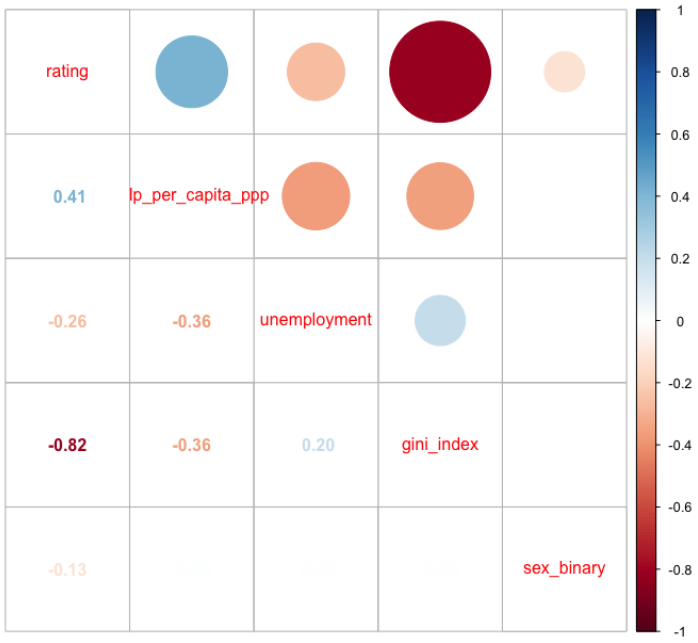

Figure 1: Correlation Matrix of Variables

- **Unemployment** (`unemployment`) shows a **weak-to-moderate negative correlation** (-0.36) with `rating`, suggesting that higher unemployment may negatively impact education outcomes.
- **The `sex_binary` variable** (Male = 0, Female = 1) does not exhibit a strong correlation with `rating`, but its impact will be further examined in the regression analysis.

This correlation matrix provides preliminary insights into the potential relationships among the variables, supporting the need for further regression analysis.

4 Results

4.1 Regression Analysis

To examine the relationship between student performance and economic and social factors, an Ordinary Least Squares (OLS) regression was conducted. The results of the regression analysis are presented in Table 1.
The regression model has an **R-squared value of 0.702**, indicating that approximately **70.2% of the variation in PISA scores** is explained by the independent variables. The adjusted R-squared of 0.698 suggests that the model is a good fit while accounting for the number of predictors.

Table 1: Regression Results

	Dependent variable:
	rating
gdp_per_capita_ppp	0.0003
	(0.0001)
	t = 3.513
unemployment	−0.641
	(0.294)
	t = −2.176
gini_index	−4.075
	(0.168)
	t = −24.300
sex_binary	−9.439
	(2.139)
	t = −4.412
Constant	623.762
	(7.568)
	t = 82.426
Observations	356
R^2	0.702
Adjusted R^2	0.698
Note:	*p<0.1; **p<0.05; ***p<0.01

4.2 Interpretation of Results

The key findings from the regression analysis are as follows:

- **GDP per capita** (`gdp_per_capita_ppp`) has a **positive and statistically significant** impact on `rating` (**p = 0.001**). A higher GDP per capita is associated with improved student performance.
- **Unemployment** (`unemployment`) has a **negative and statistically significant** effect on PISA scores (**p = 0.030**). This suggests that higher unemployment rates may negatively impact education quality.
- **Income inequality** (`gini_index`) has the **strongest negative effect** on `rating` (**p < 0.001**). A higher Gini index is associated with lower PISA scores, indicating that greater income inequality leads to worse educational outcomes.

- **Gender** (`sex_binary`) is also significant ($\mathbf{p < 0.001}$), with a coefficient of **-9.439**. This suggests that, on average, female students score 9.44 points lower than male students in PISA.

4.3 Gender Differences: t-test

To examine whether there is a statistically significant difference in PISA scores between male and female students, an independent two-sample t-test was conducted. The results are shown in Table 2.

```
> t.test(rating ~ sex_binary, data = df_clean, var.equal = TRUE)

    Two Sample t-test

data: rating by sex_binary
t = 2.439, df = 354, p-value = 0.015
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
  1.829 17.050
sample estimates:
mean in group 0 mean in group 1
       492.098 482.659
```

Table 2: Two-sample t-test for Gender Differences in PISA Scores

The results indicate that the mean PISA score for male students is 492.10, while the mean for female students is 482.66. The t-test returned a t-value of 2.44 and a p-value of 0.0152, which is statistically significant at the 5% level ($p < 0.05$).

Since the 95% confidence interval for the difference in means is [1.83, 17.05], this suggests that male students tend to score between 1.83 and 17.05 points higher than female students on average. This difference may be influenced by various factors such as subject-specific performance differences, cultural expectations, or variations in education quality across genders.

4.4 F-Tests for Variable Significance

To determine whether specific variables significantly contribute to the model, F-tests were conducted. The results are shown in Tables 3, 4, and 5.

Df	F_Statistic	P_Value
353.00	12.20	7.531e-06
351.00	12.20	7.531e-06

Table 3: F-test for Joint Significance of GDP per Capita and Unemployment

Df	F_Statistic	P_Value
352.00	590.48	< 2.2e-16
351.00	590.48	< 2.2e-16

Table 4: F-test for the Effect of Income Inequality (Gini Index)

Df	F_Statistic	P_Value
352.00	19.47	1.364e-05
351.00	19.47	1.364e-05

Table 5: F-test for the Effect of Gender on PISA Scores

4.5 Summary of Findings

The regression results confirm that economic and social factors significantly influence student performance in PISA assessments. The most important takeaways are:

1. Countries with higher GDP per capita tend to have higher student achievement.
2. Higher unemployment rates are associated with lower education performance.
3. Greater income inequality negatively affects PISA scores, emphasizing the role of social equity in education.
4. Female students tend to score lower than male students in this dataset, though this may vary across different subjects and countries.

These results provide strong empirical evidence that both economic conditions and gender disparities play a crucial role in shaping student learning outcomes.

5 Conclusion, Suggestions, and Limitations

5.1 Conclusion

This study examined the impact of economic and social factors on student performance, measured by PISA scores. The results of the multiple linear regression analysis and t-test provide key insights into how GDP per capita, unemployment, income inequality, and gender affect student achievement.

Key Findings:

- Higher GDP per capita is associated with better student performance.
- Higher unemployment rates negatively affect PISA scores.

- Greater income inequality has a strong negative impact on education outcomes.
- Male students scored significantly higher than female students, with a mean difference of 9.44 points in the regression model and a significant t-test result.

These results confirm that both economic and social factors influence education quality. Reducing income inequality and improving economic stability could lead to better student outcomes globally.

5.2 Suggestions

Based on the findings, several policy recommendations can be made:

- **Investment in education**: Countries should increase funding for education, particularly in lower-income regions, to improve student outcomes.
- **Reducing unemployment**: Economic policies that reduce unemployment may indirectly enhance education by improving household stability and resources for students.
- **Addressing inequality**: Policies that close the income gap—such as scholarships and financial aid—could mitigate the negative effects of inequality on education.
- **Gender-focused education strategies**: Given the observed gender gap, targeted programs should be implemented to ensure equal learning opportunities for both male and female students.

5.3 Limitations and Future Research

While this study provides valuable insights, several limitations should be acknowledged:

- **Omitted variables**: Factors such as teacher quality, school funding, parental education, and cultural differences were not included but may significantly impact PISA scores.
- **Time effects**: This study does not account for variations in education policies or external factors that may have changed over time.

5.4 Concluding Thoughts

As a next step, further research should explore the underlying reasons for gender differences in academic performance, examining potential factors such as societal expectations, subject preferences, and educational resources.

References

[Barro, 2001] Barro, R. J. (2001). *Economic Growth*. MIT Press.

[Canada, 2025] Canada, W. V. (2025). Why is education important? Accessed: 2025-03-09.

[Checchi, 2003] Checchi, D. (2003). Inequality in incomes and access to education: A cross-country analysis. *Labour Economics*, 10(2):197–225.

[Corak, 2013] Corak, M. (2013). Income inequality, equality of opportunity, and intergenerational mobility. *Journal of Economic Perspectives*, 27(3):79–102.

[Hanushek and Woessmann, 2012] Hanushek, E. A. and Woessmann, L. (2012). The role of cognitive skills in economic development. *Journal of Economic Perspectives*, 27(2):3–24.

[OECD, 2025] OECD (2025). What is the pisa test? Accessed: 2025-03-09.

[Oreopoulos et al., 2008] Oreopoulos, P., von Wachter, T., and Heisz, A. (2008). The short- and long-term career effects of graduating in a recession. *American Economic Journal: Applied Economics*, 121(3):1–29.

[Reis and Stockhammer, 2019] Reis, R. and Stockhammer, E. (2019). Unemployment, public spending, and educational outcomes. *Economic Policy*, 34(98):205–239.

[Tomaz, 2022] Tomaz, W. (2022). Pisa results 2000-2022 - economics and education dataset. Accessed: 2025-03-09.

A R Code for Analysis

Listing 1: R Code for Regression Analysis

```
# MidtermPaper Econometrics
# Mar7 2025
# Jumpei Kasai

library(stargazer)
library(tidyverse)
library(corrplot)
library(car)
library(xtable)

econedu <- read_csv('~/Documents/2025 Spring/Econometrics/MidtermPaper/
    economics_and_education_dataset_CSV.csv')
df <- econedu %>%
  mutate(sex_binary = ifelse(sex == "GIRL", 1, 0)) %>%
  filter(sex != "TOT")
```

```r
df_clean <- subset(df, select = c(rating, gdp_per_capita_ppp,
    unemployment, gini_index, sex_binary))
df_clean <- na.omit(df_clean)

stargazer(as.data.frame(df_clean), type = 'text')

# Scatter polot gdp per capita vs rating
ggplot(df_clean, aes(x = gdp_per_capita_ppp, y = rating)) +
  geom_point() +
  theme_minimal() +
  ggtitle("GDP per Capita vs Rating") +
  xlab("GDP per Capita") +
  ylab("Rating")

# Unemployment vs Rating
ggplot(df_clean, aes(x = unemployment, y = rating)) +
  geom_point() +
  theme_minimal()

# Unemployment vs gdp
ggplot(df_clean, aes(x = gdp_per_capita_ppp, y = unemployment))+
  geom_point()+
  theme_minimal()

# gini vs. Rating
ggplot(df_clean, aes(x = gini_index, y = rating)) +
  geom_point() +
  theme_minimal()

# sex vs Rating
ggplot(df_clean, aes(x = sex_binary, y = rating)) +
  geom_point() +
  theme_minimal()



# Corr plot
C <- cor(df_clean)
corrplot.mixed(C)

reg <- lm(rating ~ gdp_per_capita_ppp + unemployment + gini_index +
    sex_binary, data = df_clean)
summary(df_clean)

vif(reg)

stargazer(reg, report = 'vcst',omit.stat = c('ser', 'f'),type = "latex")

qt(.95, nrow(df_clean) - 4)
qt(.975, nrow(df_clean) - 4)
```

```r
t_test_result <- t.test(rating ~ sex_binary, data = df_clean, var.equal =
    TRUE)
print(t_test_result)


library(car)
library(xtable)

# Run F-tests
test1 <- linearHypothesis(reg, c("gdp_per_capita_ppp = 0", "unemployment
    = 0"))
test2 <- linearHypothesis(reg, c("gini_index = 0"))
test3 <- linearHypothesis(reg, c("sex_binary = 0"))

# Function to extract and format test results
extract_test_results <- function(test_result) {
  data.frame(
    Df = test_result$Res.Df, # Degrees of freedom
    F_Statistic = round(test_result$F[2], 3), # Round F-statistic
    P_Value = ifelse(test_result$`Pr(>F)`[2] < 2.2e-16, "< 2.2e-16",
        formatC(test_result$`Pr(>F)`[2], format = "e", digits = 3)) #
        Format small p-values
  )
}

# Convert results into LaTeX tables
test1_df <- extract_test_results(test1)
test2_df <- extract_test_results(test2)
test3_df <- extract_test_results(test3)

# Print LaTeX tables
print(xtable(test1_df, caption = "F-test for Joint Significance of GDP
    per Capita and Unemployment", label = "tab:f_test_econ"),
     type = "latex", include.rownames = FALSE)

print(xtable(test2_df, caption = "F-test for the Effect of Income
    Inequality (Gini Index)", label = "tab:f_test_gini"),
     type = "latex", include.rownames = FALSE)

print(xtable(test3_df, caption = "F-test for the Effect of Gender on PISA
     Scores", label = "tab:f_test_gender"),
     type = "latex", include.rownames = FALSE)
```