

Determining the Socioeconomic and Academic Factors affecting 2018 PISA Scores in the Philippines

Paula Gonzales² Jan Roda² Alexandra Tan² Bianca Tresvalles²
Shirlee Ocampo^{1*} and Regina Tresvalles¹

¹ De La Salle University

²De La Salle University- Integrated School Manila

*Corresponding Author: regina.tresvalles@dlsu.edu.ph

Abstract: For the first time in 2018, the Philippines joined PISA (Program for International Student Assessment). It is a worldwide study by the Organisation for Economic Cooperation and Development (OECD) in member and non-member nations intended to evaluate educational systems by measuring 15-year-old school pupils' scholastic performance on mathematics, science, and reading. The Philippines' results for the 2018 PISA were considered the lowest out of 79 countries scoring 340 in reading and second to the lowest among the 79 countries scoring 357 and 353 in science and mathematics, respectively. This study is conducted to discover some of the factors that contributed significantly to the scores of the students. It aims to identify the various socioeconomic factors and academic factors that could have affected the PISA score of the students in the different regions of the country. The study used Principal Component Analysis and Multiple Linear Regression Model to find the significant correlates. Results showed that the correlates that significantly contributed to the scores are female, those whose mothers and fathers are at least junior high school graduates, those with 101-200 or 201-500 books at home, those with 3 or more cellphones at home, those who agree that they get education support from parents, those who are in grade 9 and those who know their heritage language.

Key Words: PISA; PISA results; Academic Factors; Socioeconomic Factors; Education

1. INTRODUCTION

The Programme for International Student Assessment, otherwise known as PISA, is a global evaluation that aims to analyze the reading, writing, mathematics, and science capabilities of 15-year-old students. The objective of this program is to provide an analysis and description regarding students' intelligence, socioeconomic condition, and other indicators to give the school administrators insight as to how they can improve the overall prowess of the students. The program commenced in 1997, was conducted for the first time in 2000, and has since been administered once every three years by the Organisation for Economic Cooperation and Development (OECD) in more than 80 of the world's

countries—including the Philippines. (Organisation for Economic Co-operation and Development n.d.)

According to Angel Gurría, the secretary-general of the OECD, PISA is dependable in pointing out students' abilities and skills and is a means for identifying rooms for improvement in educational policies (People for Education, 2020). PISA has been noted to bring light to issues in education such as the poor quality of education, inequity in education, and inequality in allocating school resources in different schools and places in countries (Lockheed et al., 2015).

With that in mind, this research aims to investigate the possible social, economic, and academic factors that have contributed to the PISA scores acquired by 15-year-old students in the Philippines—specifically, in the 2018 PISA results. The

reason why the study chose to undertake such a topic and period is that the Philippines participated for the first time in 2018 and it ranked last in reading among 79 countries and ranked second to the last in mathematics and science. Overall, the country achieved a score of 340 in reading, 357 in science, and 353 in mathematics—all of which fall short of the standard set by other countries participating in the assessment (Department of Education, 2019). Given the below-average score, the group thus desired to investigate some socio-economic and academic reasons that may contribute to the increase or decrease in the PISA scores. Furthermore, this research will also look into the significant factors affecting a student's performance in each of the included criteria, science, mathematics, and reading.

In a 2020 article written by Esteban Vázquez-Cano and his colleagues, they aimed to investigate the connection between a parent's educational background and their child's reading proficiency among families within Canada, Finland, and Singapore. This study found that a parent's occupation, level of education, and involvement with their child's schooling are the greatest predictors and influencers of a child's reading abilities. The study also discovered that the reading proficiency of children was likely to be more affected by the mother's academic background than the father's and that the interest that both parents have in their child's academic performance can significantly affect the child's reading performance despite the latter's educational and sociocultural environment (Vázquez-Cano et al., 2020).

This research aims to determine and assess the socio-economic and academic factors that affect the PISA scores in the Philippines by performing multiple linear regression analysis. The analysis would provide a comprehensive insight as to what of these factors significantly affected the scores. The social factors mentioned in this study include the impacts of family, friends, and different communities in school, together with the influence of educational and cultural factors. The economic factors mentioned in this study will include the impacts of the economic status of the community and family. The academic factors mentioned in this study will consist of the time spent in class per week and the student's grade level.

2. METHODOLOGY

The researchers identified from the PISA Database the variables that will be used to characterize

the factors that affect the students' Math, Science, and Reading scores per region (total of 17 regions in the Philippines). The information used in this study came from the December 2019 report from the Department of Education regarding the Philippines' 2018 PISA results. The data set and questionnaire used can be found on the PISA OECD website. The data set is available in SAS or SPSS format. Table 1 shows the variables used in the analysis.

Table 1. Variables Used in the Study

| Variable Code | Variable Description |
|---------------------------|----------------------------------|
| Dependent Variable | |
| TotScore_Read | Total score in Reading |
| TotScore_math | Total score in Math |
| TotScore_Sci | Total score in Science |
| Correlates | |
| grl_7 – grl_12 | grade levels 7-12 |
| gen_Female | female |
| gen_male | Male |
| mom_1 – mom_5 | 1-pre primary, 2-primary. 3-JHS, |
| dad_1 – dad_5 | 4-SHS, 5- no education |
| schcomp_1 | schools with comp |
| schcomp_2 | schools without comp |
| wifiacc_1 | schools with wifi |
| wifiacc_2 | schools without wifi |
| numcp_1 -numcp_4 | Number of cellphone/ |
| numcomp_1 - | computer /tablet/ ebook at |
| numcomp_4 | home: |
| numtab_1 - numtab_4 | 1- 0 |
| numebook_1 - | 2- 1 |
| numebook_4 | 3- 2 |
| | 4- 3 or more |
| numbook_1 - numbook_6 | 1- 0-10 2- 11-25 3-26-100 |
| | 4-101-200 5 – 201-500 6 |
| | more than 500 books |
| langhome_1 | Home Language same as the |
| | test |
| langhome_2 | Home Language- Other |
| | Language |
| langsch_1 | Sch Lang - heritage |
| langsch_2 | Sch Lang - equal heritage |
| | and test |

| | |
|------------------------------------|---------------------------------------|
| langsch_3 | Sch Lang - mostly test language |
| langsch_4 | N/A |
| PSDiff(Too many words...)_1 - 4 | PISA Difficulty: 1-strongly disagree |
| PSDiff(Many texts...)_1 - 4 | 2-disagree |
| PSDiff(Navigating pages...)_3 | 3- agree |
| | 4-strongly agree |
| SOCBEL (Making friends...)_1 | Social Belongingness 1-strongly agree |
| SOCBEL (Belongness...)_3 | 2-agree |
| SOCBEL (Loneliness...)_2 | 3- disagree |
| | 4-strongly disagree |
| PARSUPP (Educ efforts...)_1 | Parental Support: 1-strongly disagree |
| | 2-disagree |
| PARSUPP (Facing difficulties...)_1 | 3- agree |
| | 4-strongly agree |
| PARSUPP (Encouragement)_1 | |
| ENGLAS | English classes per week |
| MATHCLAS | Math classes per week |
| SCICLAS | Science classes per week |
| CLASSWEEK | Ave class per week |

2.1 Principal Component Analysis (PCA)

Principal Component Analysis is a statistical method that is used to summarize information from a large data set. It does this by making large variables in a data set into smaller ones while still including most of the information from the original data set. Principal Component Analysis is usually used with datasets that consider a large set of variables. Some of the variables are focused on to “reduce the dimension” of the data set (Brems, 2017). In this way, there are fewer considered relationships between variables, which makes analyzing the relationships easier.

2.2 Multiple Linear Regression Model

The model that this study will utilize is the multiple linear regression model. This is inextricably tied to the concept of regression analysis, whose purpose is to display the linear relationship between independent

variables, denoted by X_i , and a dependent variable, denoted as Y , along with the random error term epsilon, ϵ . In a multiple regression model, given by $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_t X_t + \epsilon$, there can be two or more independent variables. The error term ϵ is normally distributed with mean 0 and variance σ^2 . Multiple linear regression analysis has several key assumptions to be satisfied. These assumptions are linearity, homoscedasticity, normality, absence of correlated errors and absence of multicollinearity. Linearity means that the dependent variable should exhibit a linear relationship with each of the independent variables. Homoscedasticity suggests that the variance of residuals is constant. Moreover, multivariate normality will be satisfied when residuals are normally distributed. Absence of multiple collinearity means that the independent variables in the model are not highly correlated with one another. Absence of correlated errors means there should be no correlation among the residual terms.

3. RESULTS AND DISCUSSION

3.1 Science

Principal component analysis was used on the original set of variables. Multiple linear regression is used on the set of independent variables that resulted after applying principal component analysis. The total score in science in the region is used as the dependent variable. The analysis of variance (F-test) result showed that the model is significant with a p-value of 0.00000. The coefficient of determination (R^2) is 1.000000.

For diagnostic checking, the assumptions for this MLRM were checked and found to be satisfied.

For the increase in the science score, the significant correlates are female, 3 or more home computers, those who strongly disagree that they feel lonely, those with 201-500 books, and those whose father's highest level of schooling is at junior high school and pre-primary, those whose mother's highest level of schooling is junior high school, those who disagree that there are too many words that can't be understood in PISA, those who strongly disagree that they get support from parents when facing difficulty, and those who strongly agree that there are many texts that were difficult in PISA. For the decrease in the score, the significant correlates are those who strongly disagree

that their parents encourage them, those with 2 e-book readers, those who strongly disagree and strongly agree that they make friends easily, and those who strongly agree that there are too many words that can't be understood in PISA. Table 2 shows the summary of the results.

Table 2. Regression Summary for Dependent Variable: Total Score in Science

| | b | Std.Err. of b | t(2) | p-value |
|--------------------------|---------|------------------|--------|---------|
| Intercept | -871.33 | 0.04448 | -19590 | 0.00003 |
| gen_Female | 482.03 | 0.00163 | 294688 | 0.00000 |
| numcomp_4 | 284.40 | 0.00599 | 47452 | 0.00001 |
| SOCBEL(online_4) | 281.47 | 0.00368 | 76403 | 0.00001 |
| PARSUPP(Encourage)_1 | -1099 | 0.00684 | -16066 | 0.00000 |
| numbook_5 | 568.33 | 0.01406 | 40418 | 0.00002 |
| PARSUPP(facing diff)_1 | 519.23 | 0.00685 | 75780 | 0.00001 |
| numebook_3 | -243.65 | 0.01181 | -20631 | 0.00003 |
| dad_3 | 124.31 | 0.00188 | 66168 | 0.00001 |
| SOCBEL(making friends)_4 | -177.36 | 0.00259 | -68542 | 0.00001 |
| PSDiff(to many words)_1 | 68.83 | 0.00207 | 33336 | 0.00002 |
| PSDiff(too many words)_4 | -255.09 | 0.00304 | -83854 | 0.00001 |
| dad_1 | 12.95 | 0.00063 | 20569 | 0.00003 |
| mom_3 | 14.20 | 0.00282 | 5042 | 0.00013 |
| PSDiff(many texts)_4 | 19.18 | 0.01024 | 1872 | 0.00034 |
| SOCBEL(making friends)_1 | -1.29 | 0.00317 | -409 | 0.00156 |

3.2 Mathematics

Multiple linear regression is used on the set of independent variables retained after performing principal component analysis with the total score in mathematics in the region as the dependent variable. The analysis of variance (F-test) showed that the model is significant with a p-value of 0.00000. The coefficient of determination (R^2) is 0.999999.

The significant correlates that contributed to the increase in the score in mathematics are female, 3 or more cellphones at home, those who agree that they feel a sense of belongingness, those who are in grade 9, those who mostly speak their heritage language, and those with higher class period. For the decrease in the score in mathematics, the significant correlates are no internet access at school, those who agree and strongly agree that they get support in educational efforts, those who disagree and strongly disagree that they feel a sense of belongingness, 2 computers at home, and those who strongly disagree on the perception that PISA is difficult (too many words) and agree that it is difficult to navigate through the pages. In this study, having no internet at school significantly affected the score of the students, however in Lara-Porrae and colleagues' 2019 article, their results showed that the variable internet access at home showed significant effects in the scores of the students. Table 3 shows the summary of the results.

For diagnostic checking, the assumptions for this MLRM were checked and found to be satisfied.

Table 3. Regression Summary for Dependent Variable: Total Score in Math

| | b | Std.Err. of b | t(2) | p-value |
|-------------------------|---------|------------------|---------|---------|
| Intercept | -5245.5 | 1162.56 | -4.5120 | 0.04577 |
| gen_Female | 154.77 | 1.803 | 85.8627 | 0.00013 |
| numcp_4 | 194.92 | 1.940 | 100.481 | 0.00009 |
| SOCBEL(Belongness...)_2 | 478.41 | 2.957 | 161.776 | 0.00003 |

| | | | | |
|------------------------------|---------|--------|---------|---------|
| grl_9 | 200.05 | 0.896 | 223.310 | 0.00002 |
| wifiacc_2 | -159.08 | 2.344 | -67.875 | 0.00021 |
| AVEMIN | 129.52 | 16.671 | 7.7691 | 0.01616 |
| PARSUPP(Ed ucefforts)_3 | -33.61 | 3.780 | -8.8919 | 0.01241 |
| numcomp_3 | -17.62 | 7.856 | -2.2427 | 0.15413 |
| PSDiff(Too manywords)1 | -176.96 | 2.330 | -75.95 | 0.00017 |
| langschl_1 | 135.60 | 2.108 | 64.3262 | 0.00024 |
| PARSUPP(Ed uc effort)_4 | -105.37 | 1.883 | -55.96 | 0.00031 |
| SOCBEL(Belo ngness...)_4 | -246.41 | 5.949 | -41.418 | 0.00058 |
| SOCBEL (Belongness3 | -172.14 | 5.456 | -31.550 | 0.00100 |
| PSDiff(Naviga ting page)3 | -13.36 | 0.997 | -13.402 | 0.00552 |

3.3 Reading

Principal components analysis was used to find the set of predictors. This is done to avoid redundancy. After applying principal component analysis, multiple linear regression is used to find the significant correlates. The analysis of variance (F-test) showed that the model is significant with a p-value of 0.00001. The coefficient of determination (R^2) is 1.000000.

The significant correlates that contributed to the increase in scores are female, 3 or more cell phones, 3 or more e-book readers, 101-200 books at home, mothers finished junior high school, those who strongly agree that PISA has too many words that can't be understood, and those who agree that they have difficulty in navigating through the pages, those who disagree that they find that there are too many texts, and those who use their heritage language. For the decrease in scores, the significant correlates are mothers finished primary education, 201-500 books at home, and those who agree, strongly agree, and strongly disagree that their parents

support them in any difficulty. These results are also supported by looking at Vasquez-Cano's 2020 article, where they found that both parents' interest and support in their child's academic capabilities could significantly affect the child's reading performance. Table 4 shows the summary of the results. For diagnostic checking, the assumptions for this MLRM were checked and found to be satisfied.

Table 4. Regression Summary for Dependent Variable: Total Score in Reading

| | b | Std.Err. of b | t(2) | p-value |
|----------------------------|----------|------------------|---------|---------|
| Intercept | 1200.73 | 0.49337 | 2433.7 | 0.00026 |
| gen_Female | 152.975 | 0.01207 | 12670.6 | 0.00005 |
| numcp_4 | 166.700 | 0.00678 | 24595.5 | 0.00003 |
| PARSUPP (Facingdiff)1 | -549.170 | 0.01653 | -33214 | 0.00002 |
| PSDiff(Many texts...)_2 | 333.972 | 0.00960 | 34784.5 | 0.00002 |
| PSDiff(Nav pages...)_3 | 248.854 | 0.01117 | 22275.5 | 0.00003 |
| mom_1 | -158.710 | 0.01074 | -14773 | 0.00004 |
| numebook_4 | 331.184 | 0.02669 | 12407.1 | 0.00005 |
| mom_2 | 83.167 | 0.00729 | 11401.5 | 0.00005 |
| numbook_5 | -379.458 | 0.06177 | -6142.7 | 0.00010 |
| langschl_1 | 31.948 | 0.00488 | 6547.0 | 0.0001 |
| PARSUPP (Facing diff)4 | -20.041 | 0.0099 | -2025.1 | 0.00031 |
| numbook_4 | 96.511 | 0.03561 | 2709.6 | 0.00023 |
| PSDiff(Too manywords)4 | 55.531 | 0.03701 | 1500.4 | 0.00042 |
| PARSUPP(Fa cing diff)_3 | -3.796 | 0.00646 | -587.4 | 0.00108 |

4. CONCLUSIONS

In this study, the significant correlates that contribute to the increase or decrease in PISA scores were obtained using multiple linear regression models. The significant correlates resulting from this statistical analysis showed that there are several socio-economic factors and academic factors that affected the scores in the different subjects. These factors may be further studied and analyzed so as to gain understanding and to make strategies on how to help students get better scores.

5. ACKNOWLEDGMENTS

The researchers would like to thank the advisers, Ms. Shirlee Ocampo and Ms. Regina Tresvalles, for their continued support throughout the process.

6. REFERENCES

- Araújo, L., Costa, P., & Crato, N. (2020). Assessment Background: What PISA Measures and How. Improving a Country's Education, 249-263. https://doi.org/10.1007/978-3-030-59031-4_12
- Beers, B. (2021, August 27). *What regression measures*. Investopedia. <https://www.investopedia.com/terms/r/regression.asp>
- Besa, F. (2019). *Philippines - country note - PISA 2018 results*. Organisation for Economic Co-operation and Development. https://www.oecd.org/pisa/publications/PISA2018_CN_PHL.pdf
- Brems, M. (2017, April 18). A One-Stop Shop for Principal Component Analysis. towards data science. Retrieved from <https://towardsdatascience.com/a-one-stop-shop-for-principal-component-analysis-5582fb7e0a9c>
- Chlebounová, D., & Odei, S. A. (2019). Impact of national culture on pupils' pisa results: The case of European countries. *Studia Universitatis Babeş-Bolyai Negotia*, 64(1), 135–153.
- Department of Education. (2019, December). *PISA 2018 philippine national report*. <https://www.deped.gov.ph/wp-content/uploads/2019/12/PISA-2018-Philippine-National-Report.pdf>
- Gamazo, A & Martines-Abas, F. (2020). An Exploration of Factors Linked to Academic Performance in PISA 2018 Through Data Mining Techniques. *Frontiers in Psychology*, 11. 1-17. <https://doi.org/10.3389/fpsyg.2020.575167>
- Lara-Porras, A. M., Rueda-García, M. D. M., & Molina-Muñoz, D. (2019). Identifying the factors influencing mathematical literacy in several Spanish regions. *South African Journal of Education*, 39(S2), 1–13. <https://doi.org/10.15700/saje.v39ns2a1630>
- Lockheed, M., Prokic-Bruer, T., & Shadrova, A. (2015). The Experience of Middle Income Countries Participating in PISA 2000-2015, PISA, World Bank, Washington, D.C./OECD Publishing, Paris, 80-82. <https://doi.org/10.1787/9789264246195-en>
- Organisation for Economic Co-Operation and Development. (n.d.). *About- PISA*. <https://www.oecd.org/pisa/aboutpisa/pisa-based-test-for-schools-faq.htm>
- People for Education. (2020). *What's in a test? Making sense of the latest PISA scores*. Retrieved from <https://peopleforeducation.ca/our-work/whats-in-a-testmaking-sense-of-the-latest-pisa-scores/>
- Vázquez-Cano, E., De la Calle-Cabrera, A. M., Hervás-Gómez, C., & López-Meneses, E. (2020). Socio-Family context and its influence on students' PISA reading performance scores: Evidence from three countries in three continents. *Educational Sciences: Theory & Practice*, 20(2), 50–62. <https://doi.org/10.12738/jestp.2020.2.004>