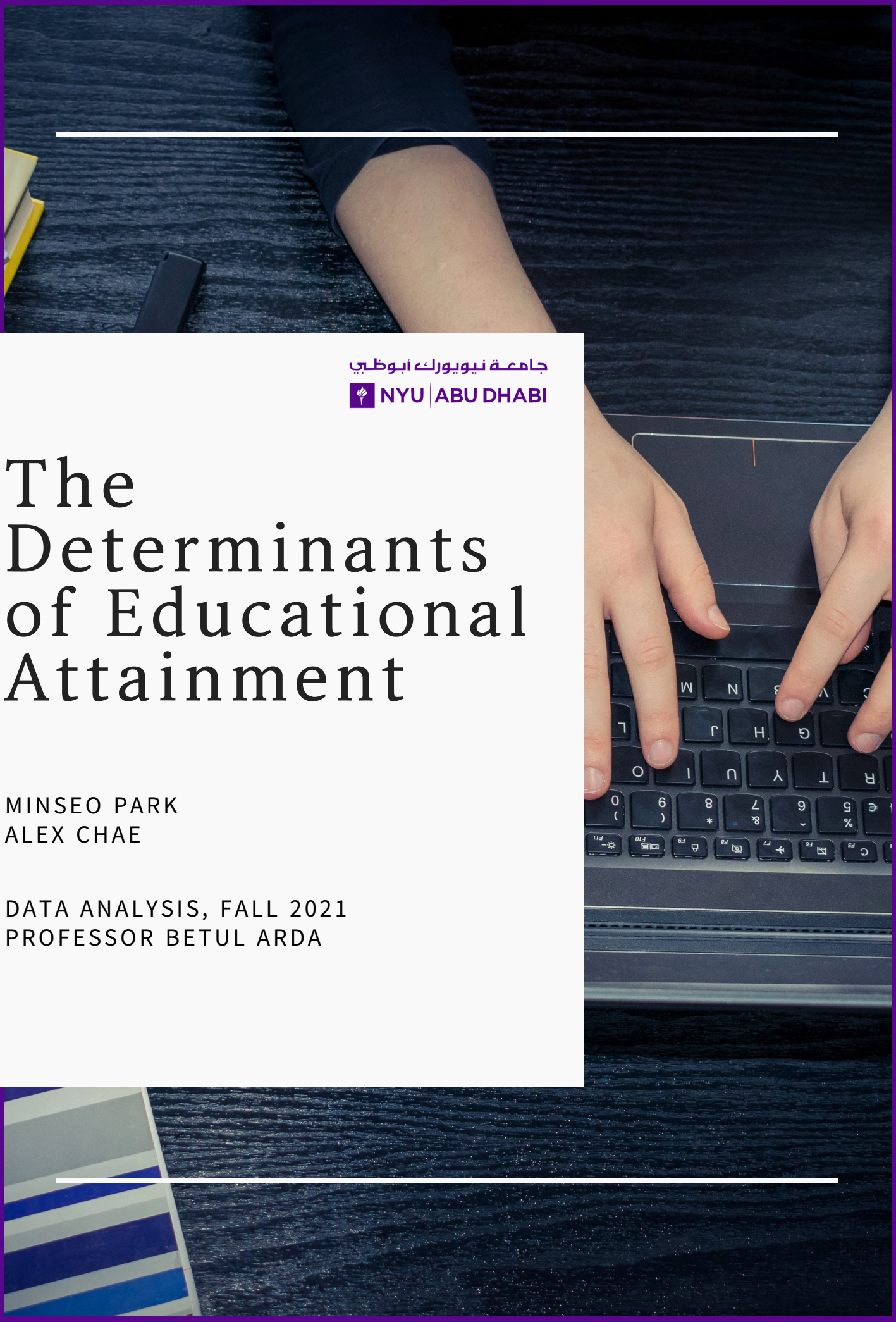

جامعة نيويورك أبوظبي
NYU ABU DHABI

The Determinants of Educational Attainment

MINSEO PARK
ALEX CHAE

DATA ANALYSIS, FALL 2021
PROFESSOR BETUL ARDA

Introduction

This research aims to answer the question "What is the relationship between educational attainment of high school students and GDP per capita, happiness level, money spent in education, inequality and teaching hours per year in OECD countries?" To measure each variable, various socioeconomic indicators will be used as proxies.

The results of the PISA test were used to measure educational attainment. PISA test is an abbreviation for OECD's Programme for International Student Assessment, which measures "15-year-olds' ability to use their reading, math and science skills" (OECD). Theoretically, there is no minimum or maximum score in PISA; rather, the results are scaled to fit approximately normal distributions. The mean PISA score is around 500 points. We decided to use the PISA results in 2015 and 2018, as they are the most recent exam data currently available.

The GINI index was used as a proxy for inequality. GINI index measures the extent to which the distribution of income among individuals within an economy deviates from a perfectly equal distribution. Thus, the GINI index of 0 represents perfect equality, while an index of 100 implies perfect inequality (United States Census Bureau).

Happiness level was measured with the Happiness Index provided by the World Happiness Report. The World Happiness Report is an annual report published by the United Nations Sustainable Development Solutions Network since 2012 providing happiness rankings of multiple countries. The survey results used in the report are primarily from Gallup World Poll, which is a firm collecting a wide range of data to assist policy making and human development (Gallup). The index ranges from 0 to 10, a higher number indicating a higher happiness level.

This research holds much importance, because education of the future generation is one of the most significant concerns of a country. The research question concerns the relationship between a standardized test score and various social factors in a country. Thus, the results of this paper may provide insight into the aspects of society that governments should focus on for a more effective education for high school students and aid in policy making processes. Moreover, on a personal level, as university students who have first-hand experienced high school education in multiple countries including South Korea, the United Arab Emirates and Brazil, we were curious to find how the different aspects of each country affects its education attainment.

Literature Review

There is various literature conducted on the factors that influence educational attainment. Woessmann (2001) analyzed the results of the Third International Math and Science Study (TIMSS) exam in 39 countries, some of which are OECD countries. He concluded that there is no significant relationship between the overall spending per student and the exam scores, but that the number of books at home was positively correlated with exam scores. The independent variable in this research is the amount of money spent in education, not the entire spending of a student. Hence, we expect it to be positively correlated with PISA test score as well. Furthermore, the study mentions that the number of books represents the educational and social background of families. Although not directly, the number of books at home may be a proxy for the money spent on education, as they both demonstrate a family's interest and level of investment in education. Hence, our hypothesis is that the number of books will have a positive correlation with PISA scores.

Another research found a positive correlation between PISA math scores in 2012 and GDP per capita in 27 countries (DiCorrado et al, 2015). However, the research did acknowledge that the sample size was small and the R-squared value was small. Contrastingly, OECD (2012) reported that a country's wealth is not necessarily associated with better education. It claims that USD 35,000 is the threshold up until which the country's GDP increases PISA scores, but after that there is no significant correlation. Rather, the level of investment towards teachers and the level of interest towards students with lower academic grades were found to be more important for a higher PISA score. We expect a similar result in our analysis that the PISA score will be positively correlated with GDP per capita, but at a diminishing rate, as other factors become more important when a certain level of investment in education is attained.

Philip (2018) found out using PISA data for a selected set of countries that there is a negative relationship between average academic excellence and inequality. The data examined OECD countries of 2000 with the exception of Mexico and Turkey. Philip argues that educational policy thus must aim to maximize educational excellence and reduce inequity. We expect a similar negative relationship between inequality and the PISA test.

Moreover, Duckworth (2007) found out that Happiness and academic achievement are positively correlated. Children higher in subjective well-being earn higher grades, even when controlling for intelligence and past academic performance. Thus, children who perform well in school may do so in part because they are happy and performing well academically may make children happier. We expect a similar result in our analysis that the happiness level will be positively correlated with PISA results.

Overall, our research replicates the previous literature conducted in the field, but is differentiated as we combine the independent variables into one model. Therefore, our model

may demonstrate the joint effect of multiple variables and may reveal further reasons as to why certain independent variables have an insignificant effect on standardized scores. Moreover, we analyze data of OECD countries only, so the results may indicate an implication of the mission of OECD in the educational attainment of high school students.

Methodology

The data used in this research are collected from a variety of trustworthy sources including the World Happiness Report, OECD and the World Bank. The analysis will start with a multi linear regression model with cross sectional data with the theoretical model as the following:

$$pisa = \beta_1 gdp + \beta_2 hapind + \beta_3 moneduc + \beta_4 ineq + \beta_5 teachhr + \epsilon$$

Equation 1. Initial theoretical model

pisa: PISA average score in 2018
gdp: GDP per capita in USD
hapind: Happiness Index in 2018
moneduc: Money Spent in Education in USD in 2018
ineq: Inequality measured in Gini Coefficient in 2018
teachhr: Number of teaching hours per year in 2018

Equation 1 shows the initial theoretical model of the study. As the analysis was conducted, the model was also modified to find the best fit.

Cross Sectional Data Analysis

Descriptive Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
pisa	38	486.229	28.923	405.3	525.3
gdp	38	46055.238	19010.58	15256.201	116480.8
hapind	38	6.621	.696	5.358	7.632
moneduc	37	16669.112	8653.473	2862.727	47693.711
ineq	38	32.447	6.47	24.4	48.9
teachhr	31	655.696	154.307	471.91	1267.2

Table 1. Summary statistics of the cross sectional data

Table 1 shows the summary statistics of the variables in the year of 2018. There are 38 OECD countries observed in total. The list of countries considered is in the Appendix. However, for some of the variables, there were some missing points. The variable for money spent on education does not have values for Switzerland and the variable of number of teaching hours has missing values for Belgium, Chile, Estonia, Luxembourg, Sweden, United Kingdom and the United States, as the values were unavailable on the OECD database.

The sample consists of OECD countries, which are countries demonstrating "a "readiness" and a "commitment" to adhere to essentially two fundamental requirements: (i) democratic societies committed to rule of law and protection of human rights; and (ii) open, transparent and free-market economies" (TUAC, 2018, p. 2). Hence, they are countries with similar characteristics, including being developed countries. Hence, although the happiness index ranges from 0 to 10, the range in our sample is from 5.358 to 7.632, as the countries are likely to be relatively happier.

The dataset was used to conduct a multiple linear regression, whose results are shown below in Table 2.

MLR 1

pisa	Coef.	St.Err.	t-value	p-value	[95% Conf	Interval]	Sig
gdp	0	0	0.92	.367	0	.001	
hapind	-13.868	6.529	-2.12	.044	-27.342	-.393	**
moneduc	.003	.001	2.71	.012	.001	.005	**
ineq	-1.57	.931	-1.69	.105	-3.492	.352	
teachhr	-.021	.03	-0.69	.497	-.082	.041	
Constant	584.658	39.334	14.86	0	503.476	665.84	***

Mean dependent var	484.643	SD dependent var	29.928
R-squared	0.750	Number of obs	30
F-test	14.383	Prob > F	0.000
Akaike crit. (AIC)	258.486	Bayesian crit. (BIC)	266.893

*** $p<.01$, ** $p<.05$, * $p<.1$

Table 2. Multi-linear regression results

The regression results show that the coefficient of *gdp* is too small to be indicated in the table. Hence, we decided to change the units of the variable from 1 USD to 1000 USD. Another multiple linear regression was performed.

MLR 2

pisa	Coef.	St.Err.	t-value	p-value	[95% Conf	Interval]	Sig
gdp2	.333	.363	0.92	.367	-.415	1.082	
hapind	-13.868	6.529	-2.12	.044	-27.342	-.393	**
moneduc	.003	.001	2.71	.012	.001	.005	**
ineq	-1.57	.931	-1.69	.105	-3.492	.352	
teachhr	-.021	.03	-0.69	.497	-.082	.041	
Constant	584.658	39.334	14.86	0	503.476	665.84	***

Mean dependent var	484.643	SD dependent var	29.928
R-squared	0.750	Number of obs	30
F-test	14.383	Prob > F	0.000
Akaike crit. (AIC)	258.486	Bayesian crit. (BIC)	266.893

*** $p<.01$, ** $p<.05$, * $p<.1$

Table 3. Multi-linear regression results with GDP per capita in $1000

Table 3 shows that the p-values for all coefficients except that of GDP per capita, number of teaching hours and inequality are significant. Happiness index is negatively correlated with PISA scores, indicating that on average, ceteris paribus, the less happier the students are the higher their exam scores. This may be due to the stress that high school students get when they spend hours studying for exams while losing sleep and social life. Money spent on education is positively correlated with exam scores, aligning with our hypothesis.

As three of the five independent variables were insignificant, we checked the model validity in other measures.

White's test for Ho: homoskedasticity
against Ha: unrestricted heteroskedasticity
chi2(20) = 21.37
Prob > chi2 = 0.3756
Cameron & Trivedi's decomposition of IM-test

Source	chi2	df	p
Heteroskedasticity	21.370	20	0.376
Skewness	7.890	5	0.163
Kurtosis	0.750	1	0.387
Total	30.000	26	0.268

Table 4. White's test for homoskedasticity for the cross sectional data

The White's test shows a p-value bigger than 0.1, so we cannot reject the null hypothesis of homoskedasticity. Hence, the model is homoskedastic and does not violate the Gauss-Markov assumption of homoskedasticity.

VIF	1/VIF
3.970	0.252
3.760	0.266
3.160	0.316
2.450	0.409
2.340	0.427
3.140	

Table 5. VIF for the cross sectional data

The variance inflation factor is 3.14 overall and all of the individual variance inflation factors are below 5. Hence, multicollinearity is not presenting a problem in our dataset either. As there is no imminent violation of Gauss-Markov assumptions or multicollinearity problem, we decided to observe further into why the three independent variables were insignificant.

Joint hypothesis gdp ineq teachhr
(1) gdp = 0
(2) ineq = 0
(3) teachhr = 0
F(3, 24) = 4.28
Prob > F = 0.0149

Table 6. Joint hypothesis test of GDP per capita, inequality and number of teaching hours

Since the three variables were insignificant separately, we were curious to see if they had statistical significance if considered jointly. We ran an F-test to see if they jointly have an effect. The p-value was 0.0149, which is statistically significant at 5% level. Hence, there is enough evidence that *gdp, teachhr,* and *ineq* are jointly significant. We plan to conduct further research into the relationship between the three variables to decide whether the three variables should

jointly be included in the model. Moreover, consideration of the small size of the observations is crucial, as the size may have increased the standard errors of the multi-linear regression and caused the statistical insignificance.

As we continued our analysis, we realized that our number of observations was too small to obtain meaningful results. Hence, with the need for a wider database, we decided to expand the data to a panel dataset. We gathered data on the variables in 2015, which is another year when the PISA test was held.

Panel Data Analysis

We gathered data on the same variables in the year of 2015 and created a panel dataset. The following table reports the data on the summary statistics of the entire dataset.

Summary Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
Pisa	76	486.766	29.807	405.3	528.7
GDP	74	43171.373	18463.252	13330.6	116481
happi	75	6.598	.769	4.8	7.632
MonEdu	75	13193.757	8236.909	2159.7	47693.699
Ineq	73	33.714	7.104	24.4	51
TeachHr	68	693.491	177.535	471.91	1267.2
Year	76	2016.5	1.51	2015	2018

Table 7. Summary statistics of the panel data

The same 38 OECD countries were considered in the model. With the new panel dataset, we conducted a linear regression analysis.

Base year 2015

Pisa	Coef.	St.Err.	t-value	p-value	[95% Conf	Interval]	Sig
GDP	0	0	-0.28	.78	-.001	0	
happi	2.14	4.547	0.47	.64	-6.972	11.252	
MonEdu	.001	.001	1.65	.105	0	.003	
Ineq	-1.929	.589	-3.28	.002	-3.109	-.75	***
TeachHr	-.055	.025	-2.25	.029	-.105	-.006	**
2015b	0	.	.	.	.	.	
2018	-16.343	6.319	-2.59	.012	-29.007	-3.679	**
Constant	572.077	30.021	19.06	0	511.913	632.241	***

Mean dependent var	485.982	SD dependent var	31.304
R-squared	0.584	Number of obs	62
F-test	12.869	Prob > F	0.000
Akaike crit. (AIC)	561.587	Bayesian crit. (BIC)	576.477

*** p<.01, ** p<.05, * p<.1

Table 8. Panel data regression with base year as 2015

Table 8 shows the results of a regression analysis with the base year as 2015. As in the previous analysis, the unit of GDP was too big, hence the coefficient of *GDP* was 0. Hence, we decided to create a new variable *GDP2* by changing the unit of GDP per capita to 1000 USD. Then we ran the same analysis again.

Base year 2015 with GDP2

Pisa	Coef.	St.Err.	t-value	p-value	[95% Conf	Interval]	Sig
GDP2	-.071	.253	-0.28	.78	-.579	.437	
happi	2.14	4.547	0.47	.64	-6.972	11.252	
MonEdu	.001	.001	1.65	.105	0	.003	
Ineq	-1.929	.589	-3.28	.002	-3.109	-.75	***
TeachHr	-.055	.025	-2.25	.029	-.105	-.006	**
2015b	0	.	.	.	.	.	
2018	-16.343	6.319	-2.59	.012	-29.007	-3.679	**
Constant	572.077	30.021	19.06	0	511.913	632.241	***

Mean dependent var	485.982	SD dependent var	31.304
R-squared	0.584	Number of obs	62
F-test	12.869	Prob > F	0.000
Akaike crit. (AIC)	561.587	Bayesian crit. (BIC)	576.477

*** *p<.01,* ** *p<.05,* * *p<.1*

Table 9. Panel data regression with base year as 2015 with GDP per capita in $1000.

In this result, the p-value for *GDP, happi,* and *MonEdu* are greater than the significance level of 10% and thus are statistically insignificant. As for the other two variables, *ineq* and *teachhr,* we can conclude that on average ceteris paribus, compared to 2015, the GINI coefficient in 2018 is likely to be 1.929 lower and the number of teaching hours is likely to be 0.055 hours less.

Base country Australia

Pisa	Coef.	St.Err.	t-value	p-value	[95% Conf	Interval]	Sig
GDP2	-.184	.254	-0.72	.476	-.709	.341	
happi	-.53	6.112	-0.09	.932	-13.175	12.114	
MonEdu	0	0	0.07	.947	-.001	.001	
Ineq	.269	.503	0.54	.598	-.771	1.31	
TeachHr	.01	.02	0.52	.61	-.031	.051	
: base Australia	0	.	.	.	.	.	
Austria	-5.397	7.887	-0.68	.501	-21.713	10.919	
Canada	19.981	7.759	2.58	.017	3.93	36.033	**
Chile	-72.318	12.008	-6.02	0	-97.159	-47.478	***
Colombia	-106.488	13.476	-7.90	0	-134.364	-78.611	***
Costa Rica	-100.128	14.02	-7.14	0	-129.131	-71.125	***
Czech Republic	-3.238	10.077	-0.32	.751	-24.083	17.607	
Denmark	5.606	7.989	0.70	.49	-10.921	22.133	
Estonia	21.712	12.267	1.77	.09	-3.664	47.088	*
Finland	22.366	8.761	2.55	.018	4.243	40.488	**
France	-5.725	8.353	-0.69	.5	-23.005	11.555	
Germany	4.839	7.482	0.65	.524	-10.638	20.317	
Greece	-48.241	14.32	-3.37	.003	-77.864	-18.618	***
Hungary	-26.056	13.918	-1.87	.074	-54.848	2.736	*
Iceland	-12.14	10.856	-1.12	.275	-34.598	10.318	
Ireland	12.447	10.209	1.22	.235	-8.673	33.566	
Israel	-34.064	9.371	-3.64	.001	-53.449	-14.679	***
Italy	-20.461	10.462	-1.96	.063	-42.102	1.181	*
Japan	22.283	11.578	1.92	.067	-1.667	46.234	*
Lativa	-17.068	11.904	-1.43	.165	-41.693	7.557	
Lithuania	-26.383	11.115	-2.37	.026	-49.375	-3.391	**
Luxembourg	-9.55	15.174	-0.63	.535	-40.94	21.84	
Mexico	-97.82	11.617	-8.42	0	-121.85	-73.789	***
Netherlands	4.263	8.005	0.53	.599	-12.296	20.822	
New Zealand	1.331	9.011	0.15	.884	-17.309	19.971	
Norway	6.492	8.511	0.76	.453	-11.116	24.099	
Poland	6.734	11.742	0.57	.572	-17.556	31.024	
Portugal	-8.698	13.627	-0.64	.53	-36.888	19.492	
Slovak Republic	-34.883	10.581	-3.30	.003	-56.773	-12.994	***
Slovenia	6.44	11.132	0.58	.569	-16.589	29.469	
South Korea	19.538	10.828	1.80	.084	-2.861	41.936	*
Spain	-16.852	10.895	-1.55	.136	-39.39	5.687	
Switzerland	9.033	9.167	0.99	.335	-9.93	27.996	
Turkey	-61.511	14.602	-4.21	0	-91.719	-31.304	***
Constant	496.454	44.55	11.14	0	404.296	588.612	***

Mean dependent var	485.982	SD dependent var	31.304
R-squared	0.981	Number of obs	62
F-test	31.815	Prob > F	0.000
Akaike crit. (AIC)	433.151	Bayesian crit. (BIC)	516.109

*** $p<.01$, ** $p<.05$, * $p<.1$

Table 10. Panel data regression with base country as Australia

Then we conducted another regression, setting the base country as Australia. Table 10 shows the results compared to the base country of Australia. However, the p-value of all of the

independent variables are greater than the significance level of 10%. Hence none of the coefficients are statistically significant. The high p-values may be due to the small number of observations, as we only considered 38 OECD countries in our model. Also, there may be omitted variables that have not been controlled for but are confounding variables. This called for further examination of the dataset as no conclusion can be drawn.

Panel Data Analysis without GDP per Capita

Upon further examination of the independent variables used in the model, we noticed that GDP per capita may be highly correlated with money spent in education.

Correlation between MonEdu GDP

Variables	(1)	(2)
(1) MonEdu	1.000	
(2) GDP	0.755	1.000

Table 11. Correlation between money spent in education and GDP per capita

Table 11 shows the correlation between money spent in education and GDP per capita. The correlation coefficient is 0.755, signifying a high positive correlation. Therefore, including both variables in the model may create collinearity and may undermine the validity of the model. As money spent one education is more directly relevant to finding the determinants of educational attainment, we decided to proceed with the analysis without GDP per capita as an independent variable.

White's Test
White's test
H0: Homoskedasticity
Ha: Unrestricted heteroskedasticity
chi2(62) = 63.00
Prob > chi2 = 0.4407
Cameron & Trivedi's decomposition of IM-test

Source	chi2	df	p
Heteroskedasticity	63.000	62	0.441
Skewness	2.790	37	1.000
Kurtosis	2.490	1	0.115
Total	68.280	100	0.994

Table 12. White's test for homoskedasticity

To confirm that the model is valid, we conducted the White's test. The p-value of 0.4407 was greater than the significance level of 10%, so we failed to reject the null hypothesis of homoskedasticity. Hence, our dataset is homoskedastic.

VIF of MLR		
Variable	VIF	1/VIF
happi	**1.69**	**0.592268**
MonEdu	**1.86**	**0.538937**
Ineq	**2.45**	**0.407470**
TeachHr	**2.42**	**0.413455**
Year		
2018	**1.23**	**0.816094**
Mean VIF	**1.93**	

Table 13. VIF to check multicollinearity

Moreover, the variance inflation factor is 1.93 overall and all of the individual variance inflation factors are below 5. Hence, the model does not suffer from multicollinearity either. Therefore, there is no imminent violation of Gauss-Markov assumptions or a multicollinearity problem.

Base year 2015

Pisa	Coef.	St.Err.	t-value	p-value	[95% Conf	Interval]	Sig
happi	1.828	4.35	0.42	.676	-6.883	10.54	
MonEdu	.001	0	2.11	.039	0	.002	**
Ineq	-1.945	.579	-3.36	.001	-3.104	-.787	***
TeachHr	-.058	.023	-2.53	.014	-.104	-.012	**
2015b	0	.	.	.	.	.	
2018	-15.461	5.857	-2.64	.011	-27.189	-3.733	**
Constant	574.451	29.341	19.58	0	515.697	633.204	***

Mean dependent var	484.867	SD dependent var	32.289
R-squared	0.612	Number of obs	63
F-test	17.988	Prob > F	0.000
Akaike crit. (AIC)	567.934	Bayesian crit. (BIC)	580.792

*** $p<.01$, ** $p<.05$, * $p<.1$

Table 14. Panel data regression without GDP per capita with base year as 2015

Then we performed a regression analysis with the new panel dataset without GDP per capita with the base year as 2025. However, the coefficient of *MonEdu* was too small to appear in the table. Hence, we changed the unit from 1 USD per student to 1000 USD per student.

Base year 2015 with MonEdu2

Pisa	Coef.	St.Err.	t-value	p-value	[95% Conf	Interval]	Sig
happi	1.828	4.35	0.42	.676	-6.883	10.54	
MonEdu2	1.02	.483	2.11	.039	.053	1.988	**
Ineq	-1.945	.579	-3.36	.001	-3.104	-.787	***
TeachHr	-.058	.023	-2.53	.014	-.104	-.012	**
2015b	0	.	.	.	.	.	
2018	-15.461	5.857	-2.64	.011	-27.189	-3.733	**
Constant	574.451	29.341	19.58	0	515.697	633.204	***

Mean dependent var	484.867	SD dependent var	32.289
R-squared	0.612	Number of obs	63
F-test	17.988	Prob > F	0.000
Akaike crit. (AIC)	567.934	Bayesian crit. (BIC)	580.792

*** *p*<.01, ** *p*<.05, * *p*<.1

Table 15. Panel data regression without GDP per capita with base year as 2015 with money spent in education in $1000 per student

Then we ran a regression analysis again, setting the base year as 2015. The results show that the p-value for *MonEdu2, Ineq, TeachHr* are less than the significance level of 10% and thus are statistically significant. Money spent in education is positively correlated with PISA scores, while inequality and number of teaching hours are negatively correlated with PISA scores.

Base country Australia with MonEdu2

Pisa	Coef.	St.Err.	t-value	p-value	[95% Conf	Interval]	Sig
happi	-1.529	5.732	-0.27	.792	-13.336	10.277	
MonEdu2	-.111	.229	-0.49	.631	-.582	.36	
Ineq	.207	.479	0.43	.67	-.781	1.194	
TeachHr	.012	.019	0.66	.514	-.026	.051	
: base Australia	0	.	.	.	.	.	
Austria	-5.659	7.645	-0.74	.466	-21.404	10.086	
Canada	21.863	7.091	3.08	.005	7.259	36.466	***
Chile	-68.796	10.665	-6.45	0	-90.761	-46.832	***
Colombia	-101.504	11.257	-9.02	0	-124.688	-78.32	***
Costa Rica	-95.279	11.331	-8.41	0	-118.615	-71.942	***
Czech Republic	-1.768	9.581	-0.18	.855	-21.499	17.964	
Denmark	5.839	7.744	0.75	.458	-10.109	21.788	
Estonia	23.493	11.658	2.02	.055	-.517	47.503	*
Finland	23.689	8.308	2.85	.009	6.579	40.799	***
France	-5.171	8.068	-0.64	.527	-21.788	11.446	
Germany	4.427	7.232	0.61	.546	-10.467	19.322	
Greece	-47.005	13.776	-3.41	.002	-75.377	-18.633	***
Hungary	-24.566	13.334	-1.84	.077	-52.028	2.896	*
Iceland	-12.911	10.484	-1.23	.23	-34.503	8.681	
Ireland	7.291	7.074	1.03	.313	-7.278	21.861	
Israel	-32.957	8.956	-3.68	.001	-51.402	-14.512	***
Italy	-19.96	10.12	-1.97	.06	-40.802	.882	*
Japan	23.429	11.129	2.11	.045	.507	46.35	**
Lativa	-14.892	11.183	-1.33	.195	-37.923	8.139	
Lithuania	-24.674	10.542	-2.34	.028	-46.385	-2.964	**
Luxembourg	-18.368	8.766	-2.10	.046	-36.422	-.313	**
Mexico	-93.155	9.376	-9.94	0	-112.466	-73.844	***
Netherlands	6.541	7.143	0.92	.369	-8.171	21.252	
New Zealand	3.206	8.371	0.38	.705	-14.034	20.447	
Norway	5.181	8.073	0.64	.527	-11.445	21.808	
Poland	9.344	10.855	0.86	.398	-13.011	31.699	
Portugal	-7.468	13.102	-0.57	.574	-34.453	19.516	
Slovak Republic	-33.07	9.976	-3.31	.003	-53.616	-12.524	***
Slovenia	7.528	10.689	0.70	.488	-14.486	29.542	
South Korea	20.095	10.467	1.92	.066	-1.462	41.653	*
Spain	-16.458	10.543	-1.56	.131	-38.171	5.255	
Switzerland	7.958	8.779	0.91	.373	-10.123	26.04	
Turkey	-58.783	13.698	-4.29	0	-86.995	-30.57	***
Constant	496.48	43.009	11.54	0	407.9	585.06	***

Mean dependent var	484.867	SD dependent var	32.289
R-squared	0.982	Number of obs	63
F-test	37.569	Prob > F	0.000
Akaike crit. (AIC)	437.321	Bayesian crit. (BIC)	518.760

*** $p<.01$, ** $p<.05$, * $p<.1$

Table 16. Panel data regression without GDP per capita with base country as Australia

We conducted a regression analysis with the base country as Australia. Despite the small number of observations, 18 of the coefficients were still significant. Some of the results suggest further investigation and present sources for deeper discussion.

The regression result for Canada was statistically significant at 1% level. The PISA test of Canada is higher than that of Australia by 21.863 points on average, ceteris paribus. The result was unexpected, because Canada and Australia share multiple common characteristics. Both are English-speaking countries, developed countries, have a thinly dispersed population with large immigrant communities. The reason lies in equity in education. In Canada, more than 90% of the students attend public schools, while in Australia only 66% of the students attend public schools (Greenwell, 2020). Most of the Canadian schools have a good curriculum overall, while in Australia quality education is provided mostly in private schools, creating a stratified schooling system. The entry barrier for Canadian schools is low, as residency is often the only criteria considered, whereas Australian private schools are highly selective (Perry, 2015). Another evident difference is that Canada had less socioeconomic divide in education, while in Australia the difference between the highest and the lowest quartile was 3 years of schooling. This means that students of lower socioeconomic scores were better compared to other countries in Canada, while they did very badly in Australia.

The regression result for Finland was also statistically significant at 1% level which was expected as Finland is famous for its education. Finland also prioritizes equity as it tries to improve all schools. Students receive less work outside of school claiming that students do not need tutoring therefore economic status does not directly determine the amount of education received. Finland emphasizes cooperation more than competition along with qualified teachers (Colagrossi, 2018).

The regression result for Estonia was also higher at 10% level. This is due to free compulsory education. Schools are not selective thus claiming that anyone can receive quality education. Estonia also emphasizes individual success in students rather than collective success and especially ensures equal access to technology. The PISA test is conducted through the Internet meaning students are being exposed to technology more than other countries (Education Estonia, 2020). Therefore, this may have helped Estonia to receive high PISA test results .

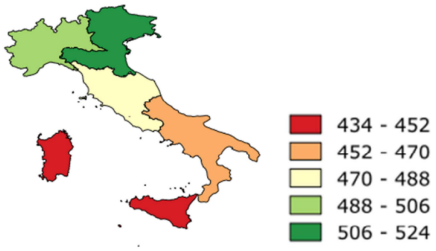

Figure 1. Map of Italy with PISA test results in 2018 (Hippe et al., 2018)

The regression results for Italy were examined as well. Italy had a significantly lower result also related to equity at 10% level. Figure 1 shows 2015 PISA results. The average PISA

test result is 485.3 however there is clear evidence of North-South divide. The Northern part of Italy did much better than the Southern part. The difference between the North and the South was almost 2 years of schooling and the immigrant population performed worse. Italian students spend their time learning outside of school such as tutoring (Hippe et al, 2018). Therefore, socioeconomic status does determine the amount of education received. Thus, educational equity is associated directly with PISA test results.

The regression results for South Korea was also higher at 1% level thus PISA test results were higher than that of Australia. The PISA report also claims that South Korea was a nation with the smallest variable. This shows that all the schools were well educating their students. South Korea also displayed above average scores in reading literacy and had low impact on social background on test performance. The underlying reason for this is because South Korea's education is more equitable than other countries due to its homogenous population. South Korea is a largely ethnically homogenous country with 95% of the population being Korean (Lyne, 2001). The immigration to South Korea is significantly lower than that of Italy due to its restrictive immigration policies. For instance, foreigners are mainly allowed to enter to fulfill low wage jobs and are excluded from getting social services such as health care services. This results in small variation in test scores. Homogenous society therefore indicate that educational equity as racial segregation is not a big part of the society.

Equity as a Variable

As we noticed that equity is a factor not included in the model but heavily influences PISa scores, we decided to test its relationship with PISA scores. We found an index created by the OECD that measures the effect of socioeconomic background on the probability of being a top performer in the PISA reading test in 2018. The higher the value, the more students of higher socioeconomic status there are in the top performers. We called this variable *equity.* A higher value of *equity* would signify lower equity in education.

Summary Statistics Equity

Variable	Obs	Mean	Std. Dev.	Min	Max
Equity	33	.407	.069	.29	.56

Table 17. Summary statistics of the OECD index for equity in education

Table 17 shows the summary statistics for *equity.* There was no data published for 5 countries, which are Chile, Colombia, Costa Rica, Mexico and Spain. Then we conducted a simple linear regression with PISA 2018 score as the dependent variable and equity in education as the independent variable.

Equity

Pisa	Coef.	St.Err.	t-value	p-value	[95% Conf	Interval]	Sig
Equity	-141.795	37.632	-3.77	.001	-218.546	-65.045	***
Constant	552.355	15.536	35.55	0	520.669	584.042	***

Mean dependent var	494.606	SD dependent var	17.371
R-squared	0.314	Number of obs	33
F-test	14.198	Prob > F	0.001
Akaike crit. (AIC)	272.607	Bayesian crit. (BIC)	275.600

*** *p*<.01, ** *p*<.05, * *p*<.1

Table 18. Linear regression with PISA scores as the dependent variable and equity as the independent variable

Table 18 shows the results of the linear regression. The p-value of 0.001 is significant at 1% significance level. Hence, the variable *equity* is negatively correlated with PISA scores and on average, ceteris paribus, an increase in equity in education is associated with an increase in PISA scores. Hence, there is enough evidence that socioeconomic status influences educational attainment.

This result points to the importance of educational policies that focus on increasing equity in education. The example of the United States demonstrates the importance of equity in education. In 2015, the United States had various structural barriers such as inequitable funding systems in education. A Department of Education in the United States claimed that 45 per cent of high poverty schools received less state and local funding in comparison to other schools. As a result, lack of equitable policies results in minories, immigrants and traditionally underserved students to attend and complete education at far lower rates than normal people. In 2017, the US government recognized these disparities and committed to advancing equity in education. The 2017 US government budget was spent on education opportunities for students in high quality learning, stronger and diverse schools and increased data for students. The United States government gave $15.4billion grants to local agencies to make sure students graduate high school and look for job opportunities. The US government also funded community driven strategies to increase socioeconomic diversity in school. Equitable funding system in education resulted in PISA test results to increase from 487 to 485 (U.S. Department of Education).

Limitations and Conclusion

This study provided various statistical analyses to find the determinants of educational attainment. We found that the model has a better fit without GDP per capita as an independent variable to reduce collinearity and that equity in education is a variable that was not included in the initial model but does have a significant influence on educational attainment. Although the research drew out meaningful conclusions, there were some limitations.

First, the study examines only 38 OECD countries. Hence, the number of observations was small. This resulted in high p-values and statistically insignificant results in the analyses. Moreover, as all of the countries examined are OECD member countries that satisfy the same aforementioned criteria, it was difficult to examine other factors that may influence PISA scores. After defining equity in education as a meaningful indicator of PISA scores, we wanted to examine the effect of wealth on PISA scores as well. Had the relationship between wealth and PISA scores been statistically insignificant or negatively correlated, then it would mean that the proportion of the wealth invested in education and the level of equity in education is more important than the wealth per se. Thus we gathered World Bank data on GNI per capita, as the World Bank classifies countries into high, middle and low-income countries with GNI per capita. Table 19 presents the summary statistics of GNI per capita of the 38 countries.

Summary Statistics GNI

Variable	Obs	Mean	Std. Dev.	Min	Max
GNI	38	44099.474	15164.527	14480	74310

Table 19. Summary statistics of GNI per capita

Table 19 shows that the minimum value of GNI per capita observed is $14480, which is higher than the criteria to be considered a high-income country, which is $12696. Hence, all of the observed countries are high-income countries. Noticing that the observed countries have an important common characteristic of being relatively richer countries, we performed a linear regression analysis of GNI per capita and PISA scores.

GNI

Pisa	Coef.	St.Err.	t-value	p-value	[95% Conf	Interval]	Sig
GNI2	1.11	.259	4.29	0	.586	1.634	***
Constant	437.287	12.039	36.32	0	412.87	461.703	***

Mean dependent var	486.229	SD dependent var	28.923
R-squared	0.339	Number of obs	38
F-test	18.429	Prob > F	0.000
Akaike crit. (AIC)	350.830	Bayesian crit. (BIC)	354.105

*** $p<.01$, ** $p<.05$, * $p<.1$

Table 20. Linear regression with PISA scores as the dependent variable and GNI per capita as the independent variable

Table 20 reports the results of the regression analysis. The p-value is statistically significant at 1% significance level and the coefficient is positive. Hence, there is enough evidence that on average, a higher GNI per capita is associated with a higher PISA score. This result was what was expected: wealthier countries have higher scores in PISA. Hence, it was difficult for us to draw out further discussion regarding the importance of equity in education. However, considering that the analysis only concerned high-income countries, the results may be different if the analysis considers all countries. Hence, it is suggested for further research to include countries of varying wealth status in the analysis to find the relationship between wealth and PISA score.

Another limitation is that the PISA score, which was used as a proxy of educational attainment, has several limitations itself. The test is conducted among 15-year old students only and only once every three years. Hence, there are not enough results to create a panel dataset with multiple observations. Furthermore, it is conducted online, hence students of countries such as Estonia who are exposed to technology in schools have an advantage over students from countries such as South Korea who are used to paper-based tests. Therefore, as the model uses PISA test scores as the dependent variable, the model did not have a large array of results and may have been biased towards certain countries.

References

Colagrossi, M. (2018, September 10). *10 Reasons Why Finland's Education System is the Best in the World.* World Economic Forum. https://www.weforum.org/agenda/2018/09/10-reasons-why-finlands-education-system-is-the-best-in-the-world.

DiCorrado, E., Kelly, K., Wright, M. (2015). The Relationship Between Mathematical Performance and GDP per Capita. *Georgia Tech.*

Duckworth, A. (2007). Happiness and Academic Achievement: Evidence for Reciprocal Causality.

Education Estonia. (2020, September 11). *Equity Made Estonia an Educational Front Runner.* https://www.educationestonia.org/equity-made-estonia-an-educational-front-runner/.

Gallup. (n.d.). *What we do.* Gallup. www.gallup.com/corporate/212381/who-we-are.aspx/

Greenwell, T. (2020, April 9). *Why do Canada's Schools Outperform Australia's?* Inside Story. https://insidestory.org.au/why-do-canadas-schools-outperform-australias/.

Hippie, R., Jakubowski, M., Araujo, L. (2018). Regional Inequalities in PISA: the Case of Italy and Spain. *JRC Technical Reports.*

Lyne, J. (2001, December 10). *Who's No. 1? Finland, Japan, and Korea, Says OECD Education Study.* Site Selection. https://siteselection.com/ssinsider/snapshot/sf011210.htm.

OECD (n.d.). *PISA*. OECD. https://www.oecd.org/pisa/.

Perry, L., (2015, February 19). *The Lesson from Canada: Why Australia Should Have Fewer Schools*. The Conversation. https://theconversation.com/the-lesson-from-canada-why-australia-should-have-fewer-selective-schools-35534.

Philip, P., Herbert M., John J., (2018). Inequity and Excellence in Academic Performance: Evidence From 27 Countries*, Review of Research in Education.* vol 55 (4).

TUAC (2018). OECD Membership and the Values of the Organisation.

United States Census Bureau. (2021). Gini Index. https://www.census.gov/topics/income-poverty/income-inequality/about/metrics/gini-index.html/.

U.S. Department of Education. Equity of Opportunity. https://www.ed.gov/equity.

Woessmann, L. (2001). Why Students in Some Countries Do Better, *Education Matters*. vol 1 (2), pp. 67-74.

Data

Helliwell, J., Layard, R., Sachs, J. (2018). *World Happiness Report* [Data set].

OECD (2019). *PISA 2018 Results Combined Executive Summaries* [Data set].

OECD. (2019). *Socio-economic Background Affects Probability of Being a Top Performer* [Infographic]. OECD. https://www.oecd.org/coronavirus/en/education-equity.

OECD. (2021). *Education Spending* [Data set].

OECD. (2021). *Gross Domestic Product (GDP)* [Data set].

OECD. (2021). *Teaching Hours* [Data set].

World Bank. (2021). *GINI Index (World Bank Estimate)* [Data set].

World Bank. (2021) *GNI per Capita, PPP* [Data set].

Appendix

The following list shows the 38 countries observed in the study.

Australia	Finland	South Korea	Slovak Republic
Austria	France	Latvia	Slovenia
Belgium	Germany	Lithuania	Spain
Canada	Greece	Luxembourg	Sweden
Chile	Hungary	Mexico	Switzerland
Colombia	Iceland	Netherlands	Turkey
Costa Rica	Ireland	New Zealand	United Kingdom
Czech Republic	Israel	Norway	United States
Denmark	Italy	Poland	
Estonia	Japan	Portugal	