VNU Journal of Science, Vol. 32, No. 1S (2016) 138-148

What Explains Vietnam's Exceptional Performance Relative to other Countries, and What Explains Gaps within Vietnam, on the 2012 PISA Assessment?

Paul Glewwe*

Department of Applied Economics, University of Minnesota, USA

Received 06 October 2016
Revised 18 October 2016; Accepted 28 November 2016

Abstract: Vietnam's performance on the 2012 PISA assessment has attracted the interest both within Vietnam and across the world. Internationally, many countries want to understand why Vietnam's education system performs so well for a lower middle income country, and what Vietnam can show them to improve their own education systems. Within Vietnam, satisfaction with this high average performance is tempered by the knowledge of gaps within Vietnam by geography (urban/rural, eight regions), income level, and ethnicity. This paper will use the Oaxaca-Blinder decomposition method to investigate possible explanations for both Vietnam's high performance on the PISA data relative to the other 64 PISA countries and for variation in student performance within Vietnam.

Keywords: Exceptional performance, gaps, pisa assessment, Vietnam.

1. Introduction

Vietnam's achievements in terms of economic growth in the last 30 years have resulted in its transformation from one of the poorest countries in the world to a middle income country [1]. While these economic achievements have attracted much attention, in more recent years Vietnam's accomplishments in education have also generated a great deal of international attention.

Vietnam's high performance in the "quantity" of education is exemplified by its high primary completion rate of 97%, and its high lower secondary enrollment rate of 92%. More striking still, is the 2012 PISA assessment: Vietnam's performance ranked 17th in math and 19th in reading out of 65 countries, ahead of both the US and the UK and much higher than that of any other developing country. Its 2012 PISA mathematics and readings scores (at 511 and 508), for example, were more than one standard deviation higher than those of Indonesia (375 and 396).

Vietnam's achievements in education are particularly notable given that it is a lower middle income country. This is shown in figures 1 and 2, which plot PISA scores in math and reading by the log of per capita GDP for all 63 countries (excluding Shanghai and "Perm", both of which are not countries). In both figures, Vietnam is in the upper left of the figure, much higher above the line that shows the expected test score given per capita GDP.

This paper uses the PISA data to understand this unusually high performance. More

Email: pglewwe@umn.edu

specifically, it does three things. First, it compares the characteristics of the students in the PISA data with the characteristics of students enrolled in school in 2012 of the same age as the PISA students, to investigate whether the PISA students are representative of 15-year-old students in 2012. Second, it uses regression methods to investigate what family or school characteristics in the PISA data can "explain" the high performance of Vietnamese students. Third, it applies an Oaxaca-Blinder decomposition to better understand the difference in average test scores between Vietnamese students and students in the other countries that participated in the 2012 PISA assessment.

This paper, while still preliminary, tentatively draws the following conclusions. First, it appears that the sample of students born in 1996, and thus about 15 years old in 2012, in the PISA sample are more urban and also of higher socio-economic status than 15 year old students in the 2012 Vietnam Household Living Standards Survey (VHLSS). Second, adding household level variables in the PISA data does little to explain Vietnam's higher performance on the 2012 PISA relative to its income level, explaining only about 9% of the gap between its actual (high) test scores and the scores predicted by its income level. Adding school level variables explains only about 20% of the gap. Third, the Blinder-Oxaca decompositions indicate that the gap in average test scores between Vietnam and the other 62 countries primarily reflects greater "productivity" of household and school characteristics in Vietnam relative to the "productivity" in other countries, as opposed to higher amounts of those household and school characteristics.

2. Are the 15-year-olds in the PISA Data Representative of Vietnam's 15-year-olds?

Some observers, both Vietnamese and international, of Vietnam's high performance on the 2012 PISA have expressed surprise that Vietnam could perform so well. This raises the question of whether the 15-year-old Vietnamese students who participated in the 2012 PISA assessment are representative of Vietnamese 15-year-old students. In each country, the students who participated in the PISA should be a random sample of children born in 1996 (and thus were 15 years old at the start of 2012) who were enrolled in school in 2012. The question for Vietnam then becomes, are the Vietnamese students who participated in the 2012 PISA assessment representative of children born in Vietnam in 1996 who were students in 2012?

This can be assessed by using data from the 2012 Vietnam Household Living Standards Survey (VHLSS). Vietnam's General Statistical Office conducts the VHLSS every two years on a random sample of Vietnamese households. This data set can be used to compare the characteristics of the Vietnamese students who participated in the 2012 PISA with a general sample of children born in 1996 who were still students in 2012.

Table 1 uses data from the 2012 PISA assessment and the 2012 VHLSS to assess the representativeness of the Vietnamese students who participated in the 2012 PISA. There do seem to be some discrepancies between the two data sources. Assuming that the VHLSS data are accurate, the students who participated in the 2012 PISA are more likely to be from urban areas (50% vs. 26%), are more likely to be in grade 10, have somewhat more educated mothers, and are more likely to live in homes with air conditioners, cars and computers. The findings in Table 1 suggest that the PISA students come from better off (and more urban) families than the typical 15-year-old student in Vietnam. This could explain part of the unusually high performance of Vietnamese students on the 2012 PISA assessment, but it is unlikely to explain all of it In fact, more thorough checking needs to be done to determine whether it really is the case that the students who participated in the 2012 PISA are "above average" students in Vietnam. Thus these findings should be treated as preliminary.

Table 1. Characteristics of Students in 2012 Who Were Born in 1996: PISA vs. VHLSS

Variable	PISA	VHLSS (PISA-eligible only)
Rural	50.0%	73.8%
Male	46.6%	48.3%
Current grade: 10th grade	85.3%	56.4%
Current grade: 9th grade	8.0%	33.5%
Current grade: 10th grade (control for interview month)	85.3%	39.1%
Current grade: 9th grade (control for interview month)	8.0%	47.2%
Father's education: above middle school	33.4%	28.0%
Mother's education: above middle school	27.5%	18.3%
Air-conditioner	15.7%	7.0%
Motorbike	92.6%	90.0%
Car	7.3%	0.7%
Computer	38.8%	24.7%
TV	97.6%	94.0%

3. What Observed Variables in PISA Explain the Gaps Conditional on Income?

Recall figures 1 and 2. Presumably there is some reason why Vietnamese students perform better than students in other countries after conditioning on (controlling for) per capita GDP. More specifically, those two figures are based on the following simple linear regression equation:

$$\text{Test Score} = \beta_0 + \beta_{gdp} \times \text{Log(GDP per capita)} + u \quad (1)$$

where β_0 is a constant term (the "intercept") and β_{gdp} is the slope coefficient for the GDP per capita variable.

In figures 1 and 2, the distance between any particular country and its performance on the test is given by u in equation (1). In particular, the value of u for Vietnam is very high. The simple regressions that generated Figures 1 and 2 is shown in Table 2. These regress the student level data in the 2012 PISA data on a constant term and the log of per capita GDP. As expected, the predictive power of GDP per capita is positive: on average, countries with a higher GDP have higher test scores. However, Vietnam's test scores in the 2012 PISA are much higher than those indicated by this regression equation. In particular, for the math regression Vietnam's average value of u is 135.8, and for the reading regression it is 119.0. These are the highest values in figures 1 and 2. This raises the question of why u is so high for Vietnam. More specifically, would adding more variables to the regression equation result in a "better fit" in which the average residual (value of u) for Vietnam would not be so high. This question is addressed in the rest of this section, first adding household and student level characteristics, and then adding school characteristics, using data from the 2012 PISA data set, which not only administered tests but also collected data from students, parents and schools.

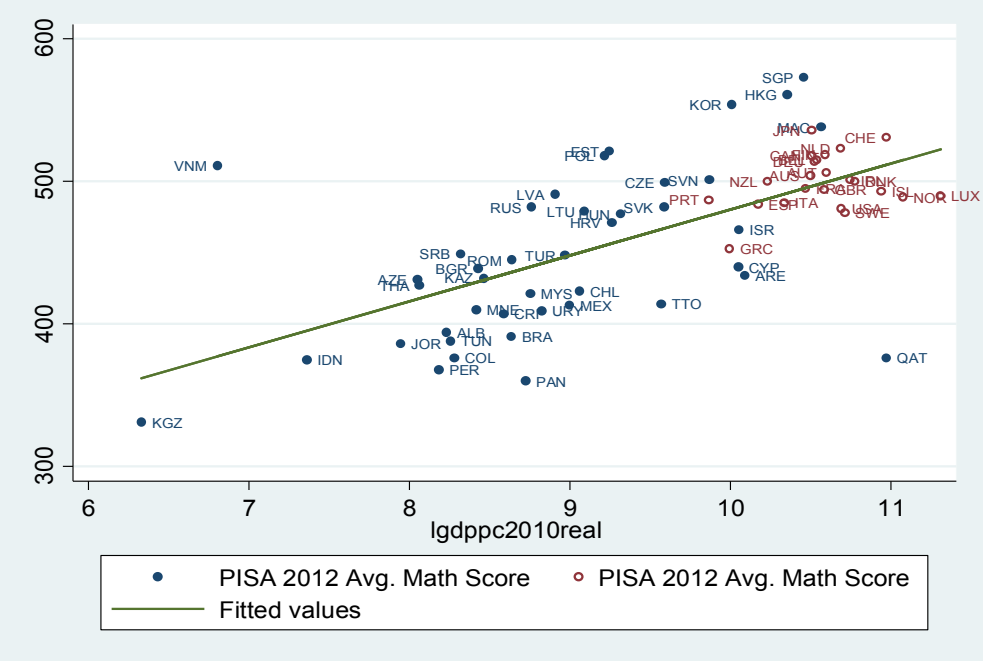

Figure 1. Mean Age 15 Math Scores in 2012 (PISA), by 2010 Log Real GDP/capita.

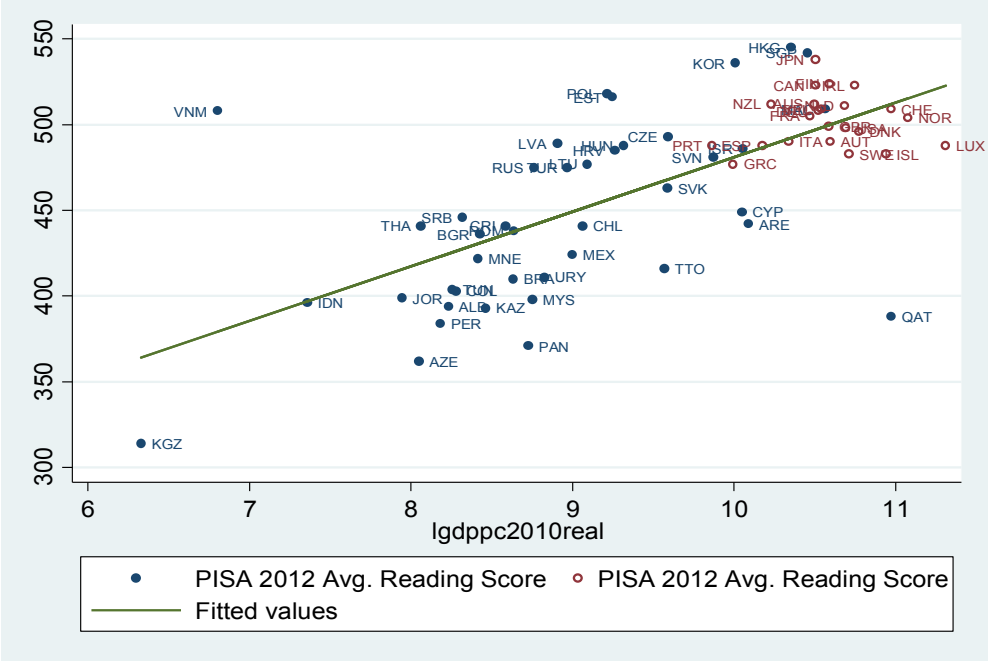

Figure 2. Mean Age 15 Reading Scores in 2012 PISA, by 2010 Log Real GDP/capita.

Table 2. Regressions of Test Scores on Log of GDP/capita: Student Level Data

VARIABLES	(1) PV1MATH	(2) PV1READ
Lpcgdp	34.14***	31.53***
	(0.136)	(0.135)
Constant	126.1***	159.5***
	(1.319)	(1.310)
Vietnam residual (average)	**135.8**	**119.0**
Observations	473,236	473,236
R-squared	0.117	0.103

Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table 3 shows regression equations similar to that in Table 2, except that the last two columns adds four household characteristics that may explain students' test score performance: an index of the number of siblings in the home (0 = none, 1 = brothers but no sisters, or sisters but no brothers, and 2 = sisters and brothers); mother's years of schooling, father's years of schooling and a wealth index (applying principal components to ownership of major durable goods). Each of these household variables sometimes has missing values. This was particularly common for the sibling index. To avoid losing many observations due to the sibling variable being missing, missing values were assigned the average value and an additional variable was created that indicates that the sibling variable was missing. A smaller percentage of observations was missing for the other variables, and so no "missing indicator' was created for those variables. This results in a decrease in the sample size from 473,236 observations to 401,489 observations.

The key question for Table 3 is whether adding these household level variables "explains" the gap in test scores between Vietnam's average value and the value predicted by the regression equations in the last 2 columns of Table 3. To see how much the average residual decreases, it is important to use the same sample for the "simple" regression (where the only explanatory variable is log(GDP/capita)) and the regression with the household characteristics added. This is done in the third and fourth columns of Table 3, which drop all observations that are missing from the last two columns.

The average Vietnam residuals (average of u) after adding the additional variables to the regression equation does not decrease by very much. For the math test, using regressions with the same sample size, the average residual drops from 129.3 to 118.2, which is a decline of only 9%.. For the reading test, the average residual for Vietnam drops from 112.5 to 102.0, which is also a drop of about 9%. Thus the household level variables in the PISA data do little to explain Vietnam's strong performance in the 2012 PISA.

Table 4 shows regression equations similar to those in Table 3, except that the last two columns adds not only household variables but also school variables. The key question for this table is whether adding the school characteristic variables "explains" more of the gap in test scores between Vietnam's average test scores and the test score than was predicted using only household level variables, as was seen in the last 2 columns of Table 3.

The average Vietnam residuals (average of u) after adding the school level variables to the household level variables in the regression equation reduces the gap, but again not by very much. For the math test, using regressions with the same sample size, the average residual drops from 119.2 to 96.6, which is a decline of 19%. For the reading test, the average residual for Vietnam drops from 103.0 to 82.1, which is a drop of about 20%. Thus combining the school variables with the household level variables in the PISA data explains only about one fifth of Vietnam's strong performance in the 2012 PISA relative to its income level.

Table 3. Regressions of Test Scores on Log(GDP/capita) and Student and Household Variables

VARIABLES	(1) PV1MATH	(2) PV1READ	(3) PV1MATH	(4) PV1READ	(5) PV1MATH	(6) PV1READ
Log(gdp/capita)	34.14***	31.53***	34.41***	32.16***	13.19***	12.71***
	(0.136)	(0.135)	(0.144)	(0.141)	(0.184)	(0.182)
Sibling index					- 3.276***	- 2.924***
					(0.227)	(0.225)
Sib. index missing					- 22.24***	- 16.76***
					(0.334)	(0.331)
Mom years school					3.035***	2.289***
					(0.0542)	(0.0537)
Dad years school					4.503***	3.804***
					(0.0535)	(0.0530)
Wealth index					10.05***	10.13***
					(0.116)	(0.115)
Constant	126.1***	159.5***	130.6***	161.5***	261.8***	289.6***
	(1.319)	(1.310)	(1.399)	(1.366)	(1.826)	(1.809)
Observations	473236	473236	401489	401489	401489	401489
R-squared	0.117	0.103	0.124	0.115	0.228	0.197

Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

4. What Can Be Learned from Oaxaca-Blinder Decompositions?

The analysis thus far assumes that the impacts of each of the variables on test scores are the same for all 63 countries in the analysis. But perhaps Vietnam's exceptional performance is partly due to it being "more effective" in using various "inputs". For example, maybe Vietnamese parents' years of schooling represent a higher level of cognitive skills.

To examine this possibility, consider the standard Oaxaca-Blinder decomposition, applied to differences in test scores between Vietnam and all other countries. The scores on the tests, denoted by S, are assumed to be linear functions of the variables used in the regression in Table 4, which are denoted by the vector **x**. The impacts of these variables on test scores, denoted by the vector **β**, are allowed to be different in Vietnam than in the other countries that participated in the PISA assessment. This yields the following two equations:

$$S_{VN} = \boldsymbol{\beta}_{VN}'\mathbf{x}_{VN} + u_{VN} \quad \text{(Vietnam)} \qquad (2)$$

$$S_{O} = \boldsymbol{\beta}_{O}'\mathbf{x}_{O} + u_{O} \quad \text{(Other countries)} \qquad (3)$$

where the error terms are denoted by u.

Table 4. Regressions Test Scores on Log(GDP/capita), Household & School Variables

VARIABLES	PV1MATH	PV1READ	PV1MATH	PV1READ
Log(gdp/capita)	32.25***	30.13***	14.69***	13.56***
	(0.155)	(0.150)	(0.219)	(0.215)
Sibling index			- 1.890***	- 2.334***
			(0.235)	(0.231)
Sib. index missing			- 20.00***	- 13.93***
			(0.346)	(0.340)
Mom years school			2.276***	1.278***
			(0.0571)	(0.0561)
Dad years school			2.905***	1.986***
			(0.0567)	(0.0557)

Wealth index			5.908***	5.794***
			(0.124)	(0.122)
Educational input index			12.28***	10.10***
			(0.110)	(0.111)
Number books in home				0.0737***
				(0.000862)
Class size			0.760***	0.905***
			(0.0146)	(0.0143)
Ratio qualified teachers			42.18***	31.05***
			(0.571)	(0.562)
Qual. tchr. ratio missing			- 30.19***	- 22.93***
			(0.398)	(0.391)
Log(computers/pupil)			1.533***	1.475***
			(0.169)	(0.166)
Stud. perf. to assess tchrs			0.0458	0.568
			(0.361)	(0.354)
Teacher absenteeism			- 9.215***	- 7.824***
			(0.196)	(0.192)
Parents pressure teachers			12.57***	12.67***
			(0.209)	(0.205)
Principal observes tchrs			- 4.721***	- 1.363***
			(0.409)	(0.401)
Inspector observes tchrs			- 2.126***	- 4.755***
			(0.318)	(0.311)
Tchr pay linked stud perf			1.362***	- 1.382***
			(0.175)	(0.172)
Teacher mentoring index			8.093***	7.007***
			(0.335)	(0.329)
Constant	156.4***	185.8***	208.1***	244.0***
	(1.499)	(1.451)	(2.495)	(2.451)
Vietnam residual	**119.2**	**103.0**	**96.6 (81%)**	**82.1 (80%)**
Observations	340964	340964	343750	340964
R-squared	0.113	0.106	0.289	0.270

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

The constant term in each of these two regression equations can be normalized so that the mean of the error term equals 0. Then taking the mean (average) of both sides of each regression equation gives the following expressions for the average test scores in Vietnam, denoted by $\bar{S}_{VN}$, and in the other 62 PISA countries, denoted by $\bar{S}_{O}$:

$$\bar{S}_{VN} = \boldsymbol{\beta}_{VN}'\bar{\mathbf{x}}_{VN} \quad (4)$$

$$\bar{S}_{O} = \boldsymbol{\beta}_{O}'\bar{\mathbf{x}}_{O} \quad (5)$$

The Oaxaca-Blinder decomposition uses equations (4) and (5) to express the difference in the mean test scores between Vietnam and the 62 other countries in the PISA data as follows:

$$\begin{aligned} \bar{S}_{VN} - \bar{S}_{O} &= \boldsymbol{\beta}_{VN}'\bar{\mathbf{x}}_{VN} - \boldsymbol{\beta}_{O}'\bar{\mathbf{x}}_{O} \quad (6) \\ &= \boldsymbol{\beta}_{VN}'\bar{\mathbf{x}}_{VN} - \boldsymbol{\beta}_{O}'\bar{\mathbf{x}}_{O} + \boldsymbol{\beta}_{O}'\bar{\mathbf{x}}_{VN} - \boldsymbol{\beta}_{O}'\bar{\mathbf{x}}_{VN} \\ &= \boldsymbol{\beta}_{O}'(\bar{\mathbf{x}}_{VN} - \bar{\mathbf{x}}_{O}) + (\boldsymbol{\beta}_{VN} - \boldsymbol{\beta}_{O})'\bar{\mathbf{x}}_{VN} \end{aligned}$$

Thus the difference in the average test scores in Vietnam and the average test scores in the other 62 countries consists of two components. The first component is the difference in the mean values of the **x** variables between Vietnam and the other countries, multiplied by the **β** coefficient for the other countries (denoted by $\boldsymbol{\beta}_{O}$). The second is the difference in the "effectiveness" of the **x** variables between Vietnam and the other countries, that is $\boldsymbol{\beta}_{VN} - \boldsymbol{\beta}_{O}$, multiplied by the

mean value of the **x** variables for Vietnam (denoted by $\bar{\mathbf{x}}_{VN}$).

Table 5 shows the mean values of the **x** variables separately for Vietnam and for the other PISA countries. At the bottom of the table, it also shows the mean math test score for Vietnam, 519.1, which is denoted by $\bar{S}_{VN}$, and the mean math test score for the other 62 countries, 473.7, which is denoted by $\bar{S}_{O}$. The gap between the two mean math scores is 55.0, and the gap between the two mean reading scores is 41.0. These gaps are smaller than the gaps shown at the bottom of Tables 2, 3 and 4, that is the average of the residuals for Vietnam, for two reasons:

First, and most importantly, the gaps based on the test scores in Table 5 do not account for the difference in mean incomes between Vietnam and the other 62 countries. As seen in Table 5, the mean of the wealth index variable is much lower in Vietnam: -1.837 for Vietnam and 0.132 for the other 62 countries. This will be discussed further below. Second, the regressions in Tables 2, 3 and 4 included both the GDP per capita for each country, which does not vary within countries, and the wealth index, which does vary within countries. In contrast, the Oaxaca decomposition can be done only for variables that vary within countries, more specifically that vary within Vietnam, since this is the only way to calculate the β_{VN} coefficient that corresponds to each variable. The regression results in Tables 6 and 7 give somewhat different results than those in Tables 2, 3 and 4, because Tables 6 and 7 do not include GDP per capita as a regressor.

Returning to Table 5, the **x** variables for which the mean is higher in Vietnam than in the other 62 countries, and for which the corresponding **β** coefficients are positive, can explain part of the gap between the mean test scores in Vietnam and the other 62 countries. That is, the contribution of such variables to the $\boldsymbol{\beta}_O'(\bar{\mathbf{x}}_{VN} - \bar{\mathbf{x}}_O)$ component in equation (6) above is positive. The contribution is also positive when the mean for Vietnam is lower than for the other 62 countries and the corresponding **β** coefficient is negative. An example of the former is the variable on whether teachers are mentored. This is higher in Vietnam than in other countries, and one may expect that teachers who are mentored would be better teachers and thus would increase their students' test scores.

Table 5. Means of Regression Variables, for Vietnam and for Other Countries

Variable (x)	Vietnam	Other PISA Countries
Sibling index	1.048	1.085
Sibling index missing	0.1520	0.2379
Mom years schooling	8.392	11.04
Dad years schooling	8.954	11.14
Wealth index	- 1.837	0.1323
Education inputs index (desk, books)	- 0.2899	0.1637
Books in home	57.99	115.0
Class size	45.23	32.51
Proportion of teachers who are qualified	0.8019	0.8369
Proportion qualified teacher missing	0.07011	0.1867
log(computers/pupil)	- 1.879	- 1.168
Stud. perf. used to assess tchrs: 1=yes 2=no	1.008	1.294
Teacher absenteeism	1.688	1.775
Parents pressure teachers	2.327	1.965
Principal observes teachers	0.9647	0.8006
Outside Inspector observes teachers	0.8476	0.4056
Teacher pay linked to student perform.	2.489	1.701
Teachers are mentored	0.8457	0.6822

In contrast, if the mean is higher in Vietnam but the corresponding **β** coefficient is negative, or the mean is lower in Vietnam and the corresponding **β** coefficient is positive, this widens the gap and in that sense makes the gap even harder to explain. For example, the mean years of schooling of the mother and of the father is lower in Vietnam than in the other 62 countries, and since one would expect that the corresponding **β** coefficients would be positive (more educated parents increase a child's test score), the parent education variables do not explain why Vietnamese students' scores are higher than those of students in the other countries, and in fact these variables "increase the burden" on other variables to explain that gap.

Briefly examining the variables in Table 5, the sibling index is similar in both columns and so is unlikely to be able to explain why Vietnamese students do better. In terms of equation (6), $\mathbf{x}_{VN} - \mathbf{x}_{O}$ is close to 0 for this variable and thus it has little chance to explain the gap. As already mentioned, since parental education is lower in Vietnam those two variables are unlike to explain the gap, and the same holds for the wealth index (accounting for the index, that is conditioning on wealth, increases the gap). The next four variables in Table 4 that one would expect to increase student learning (education input index, number of books in the home, class size, and proportion of teachers who are qualified), are all lower (or in the case of class size, higher) and so are unlikely to be able to explain the gap in average test scores between Vietnam and the other 62 countries in the PISA assessment.

There are a few variables in Table 5 that may be able to explain the gap. First, the fact the students' academic performance is used to assess teachers is more common in Vietnam may explain higher test scores in that country if this gives teachers a greater incentive to increase their students' learning. Similarly, teacher pay in Vietnam is more likely to be related to student performance. Second, the fact that teacher absenteeism is somewhat less of a problem, and that parents are more likely to pressure teachers in Vietnam, are also reasons why Vietnamese students may learn more. Third, observations of teachers by school principals and inspectors from the Ministry of Education are more common in Vietnam than elsewhere. Finally, as mentioned above teachers in Vietnam are more likely to be mentored.

Table 6. Mathematics Decomposition (difference = 519.1 – 464.1 = 55)

Variable	β_{vn}	$\mathbf{X}_{vn}$	$\beta_{vn}'\mathbf{X}_{vn}$	β_{o}	$\mathbf{X}_{o}$	$\beta_{o}'\mathbf{X}_{o}$	$\beta_{o}'(\mathbf{X}_{vn}-\mathbf{X}_{o})$	$(\beta_{vn}-\beta_{o})'\mathbf{X}_{vn}$
sibling index	4.959	1.048	5.20	-2.32	1.085	-2.52	0.09	7.63
sibling index missing	- 0.3057	0.152	-0.05	-18.38	0.2379	-4.37	1.58	2.75
Mom years schooling	1.635	8.392	13.72	2.36	11.04	26.05	-6.25	-6.08
Dad years schooling	1.988	8.954	17.80	2.746	11.14	30.59	-6.00	-6.79
wealth index	9.419	-1.837	-17.30	10.38	0.1323	1.37	-20.44	1.77
educ inputs index	7.73	-0.2899	-2.24	8.54	0.1637	1.40	-3.87	0.23
books in home	0.0142	57.99	0.82	0.0939	115	10.80	-5.35	-4.62
class size	0.6681	45.23	30.22	0.3213	32.51	10.45	4.09	15.69
ratio qualified tchrs	10.57	0.8019	8.48	46.45	0.8369	38.87	-1.63	-28.77
ratio qual tchr missing	- 12.1	0.0701	-0.85	-26.3	0.1867	-4.91	3.07	1.00
log(computers/pupil)	- 18.26	-1.879	34.31	4.454	-1.168	-5.20	-3.17	42.68
stud perf assess tchrs	- 24.76	1.008	-24.96	4.721	1.294	6.11	-1.35	-29.72
teacher absenteeism	8.539	1.688	14.41	-8.101	1.775	-14.38	0.70	28.09
parents pressure tchr	21.31	2.327	49.59	7.915	1.965	15.55	2.87	31.17
principal observe tchr	13.46	0.9647	12.98	-5.675	0.8006	-4.54	-0.93	18.46
inspect. observe tchr	- 13.85	0.8476	-11.74	-10.15	0.4056	-4.12	-4.49	-3.14
tchr pay link stud perf	4.956	2.489	12.34	-2.896	1.701	-4.93	-2.28	19.54
teachers are mentored	10.34	0.8457	8.74	6.958	0.6822	4.75	1.14	2.86
Constant	367.6	1	367.60	363.2	1	363.20	0.00	4.40
			519.08			464.18	-42.24	97.14

Table 6 presents the information needed to implement the Oaxaca-Blinder decomposition for the 2012 PISA mathematics test. As mentioned above, the overall gap to explain is 55points. In fact, differences in the **x** variables, which are expressed as the $\boldsymbol{\beta}_O'(\bar{\mathbf{x}}_{VN} - \bar{\mathbf{x}}_O)$ component of the decomposition, do little to explain the gap. Indeed, summing over all of the **x** variables shows that the values of the **x** variables lead one to expect an even bigger gap, with the overall contribution of -42.24 (see the bottom of the second to last column in Table 6). Instead, the main explanation is that the β coefficients for Vietnam reveal that Vietnam is "more efficient" in "converting" **x** variables into higher test scores; this is seen in the last column in Table 6. This is particularly true for the phenomenon of teacher absenteeism and parents pressuring teachers. Table 7 yields similar results. The differences in the **x** variables explain little, and in fact widen the gap to be explained, while the "greater efficiency" of the **x** variables explains the gap. This "greater efficiency" effect is most apparent in teacher absenteeism, principals observing teachers, and teacher pay being linked to student performance.

Table 7. Reading Decomposition (difference = 514.7 – 473.7 = 41)

Variable	$\boldsymbol{\beta}_{vn}$	$\mathbf{X}_{vn}$	$\boldsymbol{\beta}_{vn}'\mathbf{X}_{vn}$	$\boldsymbol{\beta}_o$	$\mathbf{X}_o$	$\boldsymbol{\beta}_o'\mathbf{X}_o$	$\boldsymbol{\beta}_o'(\mathbf{X}_{vn}-\mathbf{X}_o)$	$(\boldsymbol{\beta}_{vn}-\boldsymbol{\beta}_o)'\mathbf{X}_{vn}$
sibling index	5.337	1.048	5.59	-2.346	1.085	-2.55	0.09	8.05
sibling index missing	0.0101	0.152	0.00	-12.53	0.2379	-2.98	1.08	1.91
Mom years schooling	1.259	8.392	10.57	1.602	11.04	17.69	-4.24	-2.88
Dad years schooling	1.037	8.954	9.29	2.221	11.14	24.74	-4.86	-10.60
wealth index	7.096	-1.837	-13.04	10.26	0.1323	1.36	-20.21	5.81
educ inputs index	7.69	-0.2899	-2.23	9.103	0.1637	1.49	-4.13	0.41
books in home	0.00312	57.99	0.18	0.0800	115	9.20	-4.56	-4.46
class size	0.8689	45.23	39.30	0.518	32.51	16.84	6.59	15.87
ratio qualified tchrs	8.313	0.8019	6.67	36.77	0.8369	30.77	-1.29	-22.82
ratio qual tchr missing	-11.07	0.07011	-0.78	-21.46	0.1867	-4.01	2.50	0.73
log(computers/pupil)	-17.78	-1.879	33.41	4.096	-1.168	-4.78	-2.91	41.11
stud perf assess tchrs	-5.571	1.008	-5.62	5.188	1.294	6.71	-1.48	-10.85
teacher absenteeism	7.961	1.688	13.44	-7.515	1.775	-13.34	0.65	26.12
parents pressure tchr	14.56	2.327	33.88	9.708	1.965	19.08	3.51	11.29
principal observe tchr	33.63	0.9647	32.44	-2.879	0.8006	-2.30	-0.47	35.22
inspect. observe tchr	-13.23	0.8476	-11.21	-11.82	0.4056	-4.79	-5.22	-1.20
tchr pay link stud perf	6.136	2.489	15.27	-5.519	1.701	-9.39	-4.35	29.01
teachers are mentored	14.04	0.8457	11.87	6.269	0.6822	4.28	1.02	6.57
constant	335.7	1	335.70	385.7	1	385.70	0.00	-50.00
			514.74			473.72	-38.28	79.30

5. Conclusion

Vietnam's performance in education in the past 25 years has been exceptional in many respects. Perhaps the most impressive aspect of Vietnam's educational performance is the very high scores that it obtained on the 2012 PISA assessment. This is particularly impressive given Vietnam's relatively low per capita income. This paper attempts to explain this performance, using the 2012 PISA data. The following tentative conclusions can be drawn, but further analysis is warranted before final conclusions can be made.

First, the sample of children in the PISA data may not be representative of all children in Vietnam who were born in 1996 and who were still enrolled in school in 2012, as seen in Table

1. One possible explanation for this discrepancy is the timing of the administration of the PISA assessment. If the PISA tests were administered in Vietnam in the last few months of 2012, then many children born in 1996 who finished lower secondary in June of 2012 would not be included in the PISA sample if they did not continue on to upper secondary in September of 2012, and such students would be "below average" lower secondary students. This could well explain this discrepancy, but the PISA data do not include the date when the test was administered. It would be very useful to obtain from the relevant officials in Vietnam the dates of the testing for the 2012 PISA in Vietnam.

Second, regression analysis that assumes that the impacts of child, household and school variables on test scores are the same in Vietnam and in the other countries that participated in the PISA assessment provide little explanation of the reasons for Vietnam's impressive performance. The differences in those variables between Vietnam and the other participating countries explain at most only about 20% of Vietnam's exception performance (20% of the residual in the initial regression model).

Third, and consistent with the second point, Oaxaca-Blinder decompositions indicate that the explanation for Vietnam's exception performance is not that Vietnamese children, households and schools have "better" characteristics (have higher values for those characteristics) than those of other countries. Instead, the "productivity" of those child, household and school characteristics is, on average, higher in Vietnam than in other countries. Further research is needed as to why this happens, and whether most of the difference is coming from higher productivity of child and household characteristics or from higher productivity of school and teacher characteristics.

Clearly, there is much more to be learned to understand Vietnam's exceptional performance, and research on this should be given very high priority.

References

[1] World Bank, 2013, Skilling up Vietnam: Preparing the workforce for a modern market economy, Vietnam Development Report 2014. World Bank, Washington.DC.

Điều gì có thể giải thích cho thành công của nền giáo dục Việt Nam so với các quốc gia khác trên bảng đánh giá PISA năm 2012, và điều gì giải thích khoảng cách trong Việt Nam?

Paul Glewwe
Khoa Kinh tế Ứng dụng, Đại học Minnesota, Hoa Kỳ

Tóm tắt: Kết quả của Việt Nam trên bảng đánh giá PISA năm 2012 đã thu hút được đông đảo sự quan tâm từ cả 2 phía trong và ngoài Việt Nam. Trên trường quốc tế, nhiều quốc gia muốn biết được tại sao hệ thống giáo dục của Việt Nam lại có thể hoạt động tốt như vậy đối với một quốc gia có nguồn thu dưới trung bình, và những điều Việt Nam có thể chỉ ra để các quốc gia này cải thiện chất lượng giáo dục. Trong nước, những kết quả tốt được thống kê bắt nguồn từ việc ý thức được sự khác biệt về đại lí giữa các vùng miền của Việt Nam (thành thị/nông thôn, 8 vùng miền), mức thu nhập, và văn hóa dân tộc. Nghiên cứu này sẽ sử dụng phương pháp phân tích Oaxaca-Blinder để đưa ra lời giải thích cho kết quả rất tốt của Việt Nam trên bảng đánh giá PISA so với 64 nước thành viên *và* sự biến chuyển về khả năng của học sinh ở Việt Nam.

Từ khóa: Thành công vượt trội; Việt Nam.