Unraveling a secret: Vietnam's outstanding performance on the PISA tests

Suhas D. Parandekar*
Elisabeth K. Sedmik

Global Practice for Education, The World Bank

Date of this draft: Saturday 7th November, 2015

Abstract

This paper presents an analyis of the factors that explain Vietnam's outstanding performance on the PISA assessment in 2012. The paper presents a comparative analytical perspective between Vietnam and Colombia, using an Oaxaca-Blinder decomposition of a test score production function. The findings reveal that a) b) and c).

Keywords: PISA;Vietnam;Colombia;Oaxaca-Blinder Decomposition; Economics of Education.

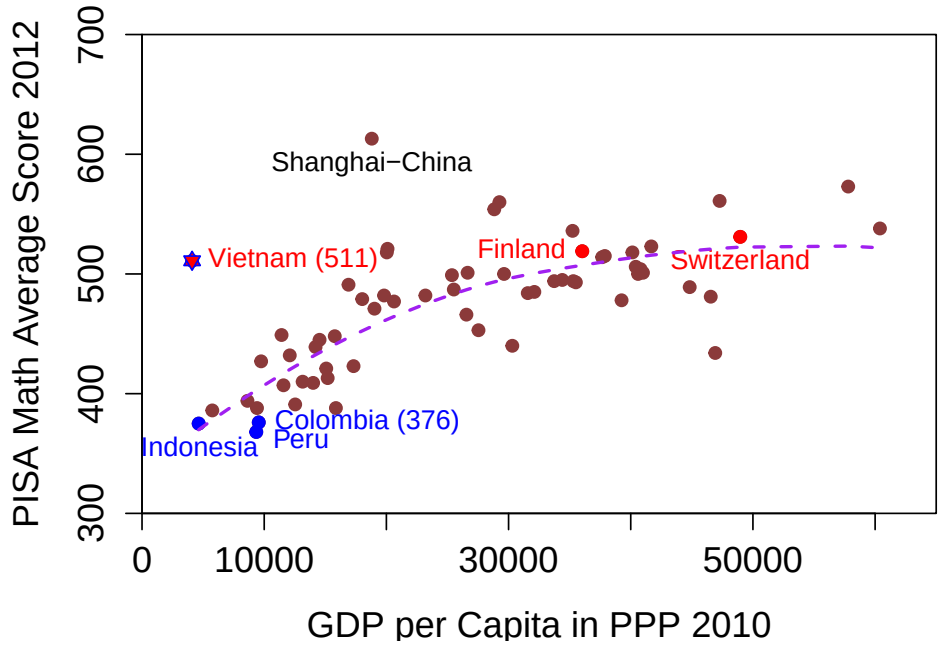
JEL Classification Numbers: I21 (Analysis of Education); I28(Government Policy); Z18(Public Policy).

*e-mail for corresponding author: *sparandekar@worldbank.org* This paper has been written using open source software: R for the econometric analysis and graphics and LaTeX for typesetting. Thanks to all who make free software possible and to OECD for making the PISA data freely and easily available to anyone. The code used in writing this paper is freely available for download at http://economist-at-work-and-play.blogspot.com/2015/02/pisa20121a.html

1 Introduction

Vietnam participated in PISA for the first time in 2012 and its performance has been much higher than other developing countries that take part in this OECD led initiative. PISA scores are calibrated to an OECD mean of 500 and standard deviation of 100 points. Only a few developing countries take part in PISA, perhaps because most of them have results much lower than the OECD countries. As can be seen in Figure 1, there is a positive, albeit non-linear correlation between GDP per capita and PISA test scores that can be seen by the dashed line representing a loess regression. The figure shows that Vietnam's performance in PISA (mathematics mean score of 511) is closer to that of Finland and Switzerland rather than of Peru and Colombia. Vietnam, represented by a red star in Figure 1, lies much above the cluster of developing countries in the lower left hand corner of Figure 1.

Figure 1: PISA 2012 results compared with GDP per capita

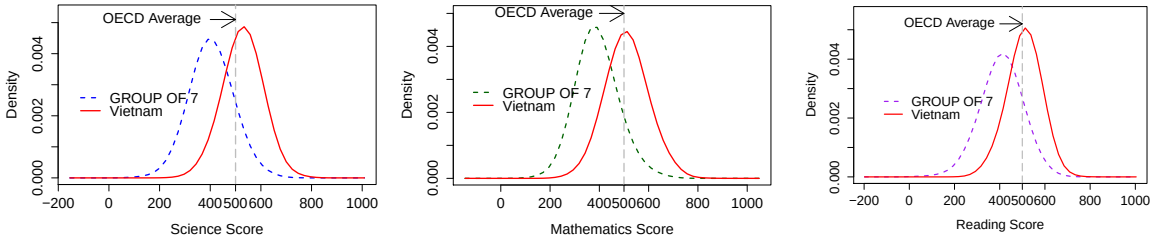


Source:OECD-PISA database

In the OECD-PISA database, there are seven countries other than Vietnam with a per capita GDP (in PPP dollars) below US$ 10,000 - Albania, Colombia, Indonesia, Jordan, Peru, Thailand and Tunisia. Their collective weighted average performance in mathematics was a mean score of 383. It is helpful to understand the significance of the 128 point difference with Vietnam. According to a recent OECD publication ([OECD (2013a)]) *"An entire proficiency level in mathematics spans about 70 score points –a large difference in the skills and knowledge students at that level possess. Such a gap represents the equivalent of*

about two years of schooling in the typical OECD country." Applying this heuristic would imply a nearly 3 year difference in attainment between Vietnam and the group of 7 developing countries in the PISA database. It should be noted at the outset that cross-section data from one instalment of PISA does not permit causal inference, but correlations can still provide useful insights. The difference is not only for mathematics and not just in the mean score, but spanning the entire test distribution, as can be seen in Figure 2.

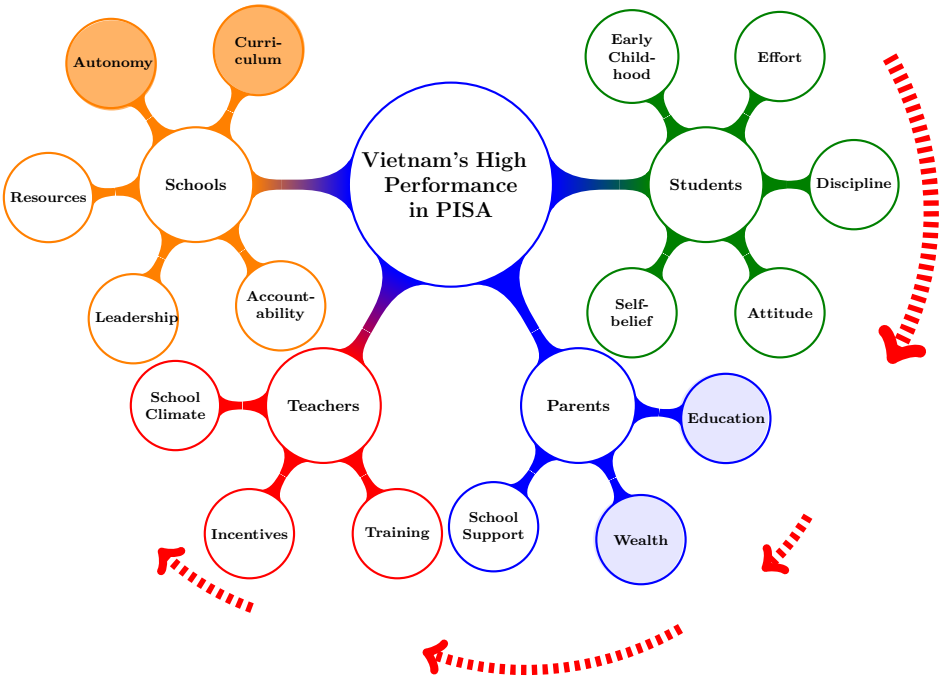
Figure 2: Kernel Density comparison between Vietnam and other Developing Countries



(a) Science **(b)** Mathematics **(c)** Reading

A range of alternative classifications are possible to organize the possible explanatory factors available in the OECD-PISA database. Figure 3 presents four sets of factors, starting clockwise from the right. This is admittedly an arbitrary classification, utilized merely for expository purposes as we consider each of the constituent variables in turn.

Figure 3: Conceptual Scheme based on available comparative variables



The approach of this paper is as follows. We begin in Section 2 by examining closely the differences in endowments between Vietnam and the collective group of 7 developing countries, termed as Dev7 for this paper (not to be confused with the G-7 of wealthy countries). The word 'endowments' is used in a general sense, referring not merely to wealth related variables such as school resources, but to the range of policy, attitudinal and reported practices in the PISA dataset. The variables can also be considered as exogenous variables for the purpose of this analysis - in the first section we examine the mean differences in the levels of these variables. Comparing means in this context is a first pass at understanding the performance anomaly of Vietnam on empirical grounds. Without any priors, we want to look at seemingly obvious possible explanation - that Vietnamese 15 year olds somehow enjoy better endowments - economic, social, cultural and so on. This first pass can be quite revealing. For instance, even though we know that Vietnam's GDP per capita is the lowest in this G8 group of countries, perhaps the Vietnamese government and people have invested heavily in basic education, and schools in Vietnam enjoy much higher quality of facilities and equipment. Or the teachers have much higher level of education, and so on. Alternatively, one could find empirically that the schools are worse equipped and the teachers poorly trained. An examination of mean differences will provide us with a first set of tentative hypotheses.

We make a second closer approach in Section 3 by adopting the regression methodology used by Fryer and Levitt in understanding differences in test score results of black children in the United States [Fryer and Levitt (2004)]. Differences in mean values of endowments leave some questions unanswered, which further analytical approaches can seek to resolve. In a multi-variate setting, we can try to understand how much of the variation can be explained by differences in observable endowments, and how much can be attributed to effects other than the observable differences. Fryer and Levitt have quite successfully followed the approach to explain what they see as evidence of a systemic difference in schools attended by black children. We adopt a similar approach to find out how much of the 'Vietnam gap' could reasonably be attributed to the earlier differences in mean endowments. As a matter of fact, we show in this section that the large differences in mean endowments is important, but it explains at best only half of the test score differences.

The Fryer and Levitt method deepens the understanding from mean comparisons, but what it does not reveal may be as interesting as what it does reveal. Our Fryer-Levitt adaption is based on a pooled regression of eight developing countries, where we follow the fate of the magnitude of the coefficient of the dummy variable representing the Vietnamese student in the sample. But we also need to investigate structural differences in the effects of endowments between Vietnam and Dev7 countries. In Section 4, we adopt an approach first

used to explain variation in PISA performance between Germany and Finland by Andreas Ammermueller [Ammermueller 2007]. This is an adaptation of the popular Oaxaca-Blinder decomposition of wage earnings equation to uncover evidence of discrimination on the basis of gender. [Blinder 1973] and [Oaxaca 1973] In this section, we can examine closely the structural differences between Vietnam and the Dev7 countries, including the contribution of differences in endowments and the coefficients to the gap in test scores.

Even a multi-variate regression approach only proves correlation with nothing more than a hint regarding causation, and so far we have only one year (2012) of PISA data for Vietnam. PISA 2015 data will be available in 2016 and analytical approaches closer to establishing causality can be attempted. Even with the caveat regarding causality, there are useful policy related conclusions that we can derive from the analysis presented in this paper - learning for Vietnam, with regard to areas of strengths and weaknesses; and learning for other countries that may wish to emulate aspects of Vietnam's performance. There is a veritable industry of papers regarding Finland's PISA performance, directed mostly towards other OECD countries with lower scores, for instance United States. Vietnam's superlative performance points to a similar future stream of research, with the added advantage of relevance for developing countries. Section 5 provides concluding ideas that might be amongst the first of many more such ideas from future investigations of Vietnam's performance.

2 Endowment Differences

2.1 Student Characteristics

Table 1 begins an exploration of differences in mean values between Vietnam and Dev7 students. This exploration only provides us with a beginning in our task of unravelling the mystery of Vietnam's performance. Neither mean differences nor the regressions that follow can claim to uncover causality. However, they do provide what in a court of law could be called 'circumstantial evidence'. Table 1 presents student characteristics, and it should be noted that the absence of differences is sometimes as important as the presence of differences. Table 1 indicates no differences by age or gender of student. The PRESCHOOL variable shows the first instance of a statistically siginificant difference. While 78.88% of Dev7 students reported attending preschool, the number of students attending preschool from the Vietnam sample was 91.20% - a sizeable difference both statistically and economically significant. The relationship between preschool and later outcomes has been studied very closely over the years. Longitudinal impact evaluation studies regarding the Perry Preschool

project and Head Start in the US are amongst the most cited studies in the economics literature[1]. We can also see from the numbers for REPEAT in Table 1 that PISA takers in Vietnam were three times less likely to have repeated a grade in the past (6.79% compared to 19.15%).

Table 1: Student characteristics and family background

		Dev7 countries		Vietnam	
Variable	Description	MS	Valid N	MS	Valid N
Fixed characteristics					
FEMALE	Sex of student	0.5265 (0.4993)	41394	0.5336 (0.4989)	4882
AGE	Age of student	15.8211 (0.2895)	41394	15.7692 (0.2885)	4853
Student's prior history					
PRESCHOOL	Attended Preschool	0.7888 (0.4082)	40114	0.912 (0.2833)	4866
REPEAT	Grade repeating	0.1915 (0.3935)	40343	0.0679 (0.2516)	4860
Truancy from School					
ST08Q01	Times late for school	1.5131 (0.7648)	40663	1.1872 (0.4685)	4873
ST09Q01	Days unexcused absence	1.2192 (0.5276)	40650	1.0999 (0.3527)	4875
ST115Q01	Times skipped classes	1.2585 (0.545)	40632	1.0764 (0.3216)	4880
Parental background and family wealth					
HISEI	Highest parental occupational status	40.4196 (22.5168)	32814	26.6023 (19.855)	4860
MISCED	Educational level of mother (ISCED)	3.1193 (1.9853)	40486	2.1744 (1.6059)	4844
WEALTH	Family wealth possessions	-1.4606 (1.2267)	40821	-2.1343 (1.1656)	4881
CULTPOS	Cultural possessions	-0.1424 (0.9678)	39905	-0.2361 (1.0173)	4809
HEDRES	Home educational resources	-0.7427 (1.1473)	40579	-1.0743 (0.9364)	4874
BOOK_N	Number of books in family home	53.6393 (94.5556)	39631	50.786 (75.4031)	4841

Notes: The variables relate to the questionnaires administered to students in the general (non-rotated) booklet. For a more detailed description of variables, please see Table xx. Items marked with *(r)* are taken from the rotated student questionnaire. The variable means of Dev7 and Vietnam are statistically different at the 95% significance level, except FEMALE.

The finding regarding PRESCHOOL and REPEAT indicates the possible importance of the trajectory of the student prior to High School. Repetition rates are difficult as comparative indicators of system quality because of the variations across countries in curriculum and standards, but REPEAT is another interesting variable to keep in mind as a possible clue to the mystery of Vietnam's PISA performance. As in some other East Asian cultures,

[1]For detailed meta-analysis, see [Barnett 1995] and [Schweinhart et al, 2005]

Vietnamese children are expected by their parents to study hard. Though Mark Twain in Vietnamese translation is quite a best seller for young readers in Vietnam, truancy from school is not perceived benovelently by parents.[2] Table 1 indicates a consistently lower truancy rate for the three variables used - the question refers to the past week **(to confirm)** and Vietnamese students are less likely to be late for school, have fewer days of unexcused absence and skip fewer classes.[3]

The final set of variables in Table 1 concern parental background and wealth at the students home, including cultural resources and books at home which may work to stimulate cognitive development. The PISA database includes a number of indices to measure aspects such as weath and possessions. These indices are based on underlying data regarding occupations and possessions. The scaling of raw data to indices is described in detail in the PISA technical report [OECD (2014a)]. For HISEI, the parents occupation status, the OECD mean is 50 and the OECD standard deviation is 15. Table 1 shows that the HISEI for Dev7 parents at 40.42 was much higher than 26.60 for Vietnamese parents. MISCED refers to the International Standard Classification of Education (ISCED) developed by UNESCO. Table 1 shows that the average level of mother's education for Dev7 was just over 3, meaning Upper Secondary education, while for Vietnam the mean was just over 2, meaning Lower Secondary education. The WEALTH index is set for an OECD mean of zero and standard deviation of 1. Dev 7 countries wealth level was -1.5 and Vietnam's was -2.1, consistent with the data regarding occupational classification and mother's education. These findings indicate the close correlation of these variables with GDP per capita. The more interesting finding concerns the indices CULTPOS and HEDRES, which have OECD mean 0 and s.d. 1, and the number of books. CULTPOS includes classical literature, books of poetry and works of art. HEDRES includes reference books and books to help with school work as well as a study desk and 'a quiet place to study'. These three variables are also in line with per capita income - with the Dev7 mean being lower than the OECD mean, and Vietname being lower than Dev7. So one explanation regarding Vietnam's PISA performance can probably be ruled out - it does not seem likely that Vietnamese households spend disproportionately higher amount of their incomes on acquiring possessions such as books and other objects that would give their children an edge in life.

[2]A cultural explanation is possibly quite important in explaining Vietnam's anomalous PISA results, though the PISA dataset may only be able to measure the possible effects of culture rather than measuring cultural differences. Literature from the World Values Survey, that does seek to measure cultural differences, indicates that Vietnam is a positive outlier on discipline and authority orientation[Dalton and Ong (2005)].

[3]In the student's questionnaire, there is a telling question - student's have to agree or disagree on a four point Likert scale to the statement "If I had different teachers, I would try harder at school.". Converted into an index, the mean for Vietnam at 0.363 is lower than that for Dev7 at 0.525. This suggests a tendency in Vietnamese students for greater self-responsibility.

2.2 Student Effort

The phenomenon of primary and high school children taking extra classes to supplement in school instruction in Vietnam is well known [Ha and Harpham, 2005] and [Dang 2007]. Table 2 indicates that while Dev7 students spent roughly 4.7 hours in such classes (total of OUTMATH, OUTLANG and OUTSCIE), the Vietnamese student spent nearly 2 hours more for a total of 6.6 hours, with the difference being highest for OUTMATH. Vietnamese students also spent about 1 additional hour per week doing homework (total of ST57Q01 and ST57Q02) compared to Dev7 students. The highest difference in this set of variables concerns the variable ST57Q04, which relates to extra classes taught by a commercial company. While most of the schools in Vietnam are public or government schools, it is interesting to note that students report nearly 5 hours of commercially provided extra lessons, while the total for Dev7 countries is only about 2 hours. Collectively, these variables indicate that Vietnamese students spent about 16 hours per week studying outside of school, compared to 13 hours per week for Dev7 students.

Table 2: Student studying time out of school

		Dev7 countries		Vietnam	
Variable	Description	MS	Valid N	MS	Valid N
Weekly out-of-school hours per subject					
OUTMATH *(r)*	weekly out-of-school lessons in math	1.828 (2.1539)	23603	3.1305 (2.3133)	3227
OUTREAD *(r)*	weekly out-of-school lessons in 'test language'	1.2882 (1.9623)	23531	1.4483 (1.8837)	3223
OUTSCIE*(r)*	weekly out-of-school lessons in science	1.5609 (2.0456)	23298	2.0927 (2.1776)	3205
Weekly out-of-school hours approach					
ST57Q01 *(r)*	Out-of-school time homework	5.0953 (5.0319)	23696	5.8145 (5.7196)	3164
ST57Q02 *(r)*	Out-of-school time guided homework	2.551 (2.9296)	19355	2.8814 (3.2384)	2285
ST57Q03 *(r)*	Out-of-school time personal tutor	1.7276 (2.7884)	20367	1.5749 (2.938)	3049
ST57Q04 *(r)*	Out-of-school time classes by company	1.892 (3.3487)	19517	4.878 (4.8058)	3091
ST57Q05 *(r)*	Out-of-school time parent/family member	2.1354 (3.055)	21542	1.7646 (3.2442)	3092
ST57Q06 *(r)*	Out-of-school time learn on computer	2.588 (3.5519)	21338	1.8029 (3.0496)	3079

Notes: The variables relate to the questionnaires administered to students in the rotated booklet. Items marked with *(r)* are taken from the rotated student questionnaire. The variable means of Dev7 and Vietnam are statistically different at the 95% significance level.

2.3 Student Attitudes

PISA applications in each round have a focus on one of the subjects and in PISA 2012 the focus subject was mathematics. Mathematics happens to be the subject where the score difference is highest between Vietnam and Dev7 countries. The PISA questionnaire for students includes a very interesting series of questions regarding student's perceptions of their abilities, their effort and their reported practices. The details of these questions can be found in the PISA Technical Manual. Typically, each question includes a set of likert scaled items to which the student provides a discrete response on a four point agree-disagree scale. These responses then are combined under specified algorithms to provide an index value. For instance, there is a question to which the response is meant to measure a student's MATWKETH or 'mathematics work ethic'. Students either agree or disagree with a set of 9 items on a 4 point likert scale - strongly disagree, disagree, agree and strongly disagree. The items include items such as "I work hard on my mathematics homework", and "I listen in mathematics class", "I keep my mathematics work well organized". In the case of this example, when a student agrees or strongly agrees with a positive statement, or disagrees/strongly disagrees with a negative statement, he or she would tend to be deemed to have a stronger work ethic towards mathematics. The raw data from the likert scale is converted into an index using IRT scaling procedures, so that the mean for OECD countries is 0 and standard deviation is 1. Table 3 indicates a most interesting finding regarding a range of such indices from the PISA database.

Table 3: Student self-perception regarding mathematical ability and student effort

		Dev7 countries		Vietnam	
Variable	Description	MS	Valid N	MS	Valid N
Indices susceptible to 'bragging' tag					
MATWKETH *(r)*	Mathematics work ethic	0.4514 (0.9782)	26140	-0.0014 (0.6915)	3217
SUBNORM *(r)*	Subjective norms in mathematics	0.716 (1.165)	26509	-0.0923 (0.8395)	3220
OPENPS *(r)*	Openness to problem solving	0.1949 (0.9787)	25612	-0.6125 (0.8708)	3207
SCMAT *(r)*	Self-concept of own math skills	0.1673 (0.8101)	26222	-0.1896 (0.5903)	3249
Indices less related to being modest/boastful					
PERSEV *(r)*	Perseverance in problem solving	0.3387 (0.9605)	25710	0.4475 (0.8767)	3211
ANXMAT *(r)*	Mathematics Anxiety	0.3995 (0.7724)	26275	0.2115 (0.6354)	3248
MATINTFC *(r)*	Mathematics intentions	0.092 (0.9837)	24827	0.3285 (1.0964)	3181

Notes: For the full set of results, please consult the github repository for this paper

The upper panel in Table 3 indicates a set of indices where the scores for Vietnamese students are lower than the scores for Dev7 students. For example, the score for MATWKETH is 0.45 for Dev7 and 0 for Vietnam. The variable SUBNORM is supposed to measure subjective norms regarding mathematics. This construct relates to a student's perceptions regarding how other people in the student's life value mathematics. It includes items such as "my friends enjoy taking mathematics tests" and "my parents believe it's important for me to study mathematics." Presumably, when this measure is high, the student has a high subjective norm for mathematics. Table 3 shows the resulting mean for Dev7 countries is 0.72 and the value for Vietnam is -0.09. The index SCMAT includes items such as "I learn mathematics quickly" and "I have always believed that mathematics is one of my best subjects". Vietnamese students, who scored more than 1 standard deviation above the Dev7 students on the PISA math test, scored half a standard deviation lower on SCMAT. What is going on here ?

This mini-mystery within the overall mystery can possibly be resolved by looking at all the indices. There was a set of indices with low performance difference between Dev7 and Vietnam and they do not shed much light on the seemingly contradictory result reported in the upper panel on Table 3. The lower panel of Table 3 reports on indices where the balance tips on the other side - these are indices where Vietnamese students performed better than the Dev7 students. The three indices here bear close examination. PERSEV consists of items that purport to capture perseverance with a task or a problem to resolve; ANXMAT is a negative index (less is better) that deals with mathematics anxiety (for example an item is "I get very nervous doing mathematics problems"); MATINTFC relates to future mathematics intention, including items such as "I am planning on majoring in a subject in college that requires lot of mathematics".

One possible explanation, as indicated in the heading of the Table 3 panels, is that Vietnamese students are bought up in a culture that stresses the importance of modesty and humility as a pathway to learning. They may find it difficult to say great things about themselves, because of cultural norms against bragging or boasting. The lower panel in Table 3, on the other hand includes items that are less prone for an immodest interpretation. To say that you are not afraid of mathematics may not be perceived as bragging. And in this context, the Vietnamese students are less anxious and more confident about the the future of mathematics in their life.[4]

[4]It will be straightforward to examine this hypothesis more closely by performing an IRT scaling of the underlying items for the indices. We can then test for differences between Vietnam and the Dev7 countries in values of the location parameters linking the items to the index. Systematic differences will tend to support the hypothesis laid out here.

2.4 Mathematics Curriculum

In addition to beliefs and perceptions of students regarding mathematics in general, PISA also seeks to investigate more closely the issue of content of mathematics instruction. PISA incorporates a very interesting approach to avoid or minimize the bragging or over-claiming problem referred to in the previous sub-section. The index FAMCON is constructed out of a response to a question about mathematical concepts for which students are asked "How familiar are you with the following items?" The list of items includes items such as 'Linear Equation', 'Quadratic Function' and 'Cosine.' The list of items also includes three nonsensical items or pseudo-concepts that sound fancy: 'Proper Number','Subjunctive Scaling' and Declarative Fraction'. These items are termed as FOIL, and are used as trick items to calibrate the response for over-claiming on part of the students. The index without correction is presented as FAMCON, and the index with correction is presented as FAMCONC. It is quite fascinating that with FAMCON, the uncorrected version, Dev7 students come out apparently better than Vietnam students mean of 0.26 as compared to 0.12. Unfortunately, it appears that this also included familiarity with non-existent items like 'subjunctive scaling' - or bragging. With the corrected version, FAMCONC, the Vietnamese students turn out to do much better, with a mean value of 0.43 as compared to Dev7 mean of -0.54.

Table 4: Student reported experience in mathematics

		Dev7 countries		Vietnam	
Variable	Description	MS	Valid N	MS	Valid N
FAMCON *(r)*	Familiarity with math concepts	0.2559 (1.1654)	26164	0.1225 (0.6935)	3243
FAMCONC *(r)*	FAMCON corrected with FOIL	-0.5441 (0.8768)	25832	0.4297 (0.9057)	3231
EXAPPLM *(r)*	Experience with applied math tasks	0.1111 (1.06)	26133	-0.2418 (0.7624)	3243
EXPUREM *(r)*	Experience with pure math tasks	-0.1384 (0.9809)	25973	0.1587 (0.8076)	3244

Notes: The variables relate to the questionnaires administered to students in the rotated booklet. Items marked with *(r)* are taken from the rotated student questionnaire. The variable means of Dev7 and Vietnam are statistically different at the 5% significance level.

The index EXAPPLM asks students about their experience during school work with examples of applied mathematics problems. Similarily, the index EXPUEM refers to experience with examples of pure mathematics. Not surprisingly, Vietnam students indicate a lower performance on EXAPPLM and a higher performance on EXPUREM.[5] To the ex-

[5]It has been a long standing issue that Vietnamese students are expected to learn a curriculum that is more crammed than the international norm, but has more of theory and abstract mathematics rather than applied mathematics.See [Danh Nam Nguyen and Trung Tran, 2013] and [Tuan Anh Le, 2007].

tent that the PISA mathematics test almost by definition tends towards pure rather than applications in real life, the figures help to understand Vietnam's performance.

2.5 Parental Support at School

The publication of the bestseller book [Chua 2011] "Battle Hymn of the Tiger Mother" in 2011 ignited a firestorm of controversy. The book gave prominence in the popular culture to a vast academic literature regarding parenting styles and about the perceived higher performance of children from Asian immigrant families in the US and other western countries.One of the ways that parents influence children's education outcome is through the interaction that parents have with their child's teachers and others at school. The PISA data includes a question that asks about parental expectations.This question (SC24) includes a statement "There is *constant pressure* from many parents, who expect our school to set very high academic standards and to have our students achieve them."[6] Table 4 indicates a higher level of PARPRESSURE for Vietnam as compared to Dev7. Another question (SC25) asks school principals about the proportion of parents that take part in a set of 12 activities. While the question does not specify which parent (or both) may be involved, the variables have been named after the mother for ease of exposition.

Table 5: Parental Support at School

		Dev7 countries		Vietnam	
Variable	Description	MS	Valid N	MS	Valid N
PARPRESSURE	Parental achievement pressure	0.2665 (0.4421)	40372	0.3837 (0.4863)	4866
TIGERMOM	Parent initiates - progress discussion	52.4472 (38.097)	41394	62.4183 (41.3743)	4882
DUTYMOM	Teacher initiates - progress discussion	66.9737 (36.727)	41394	68.5543 (37.4796)	4882
VOLUMOM	Parent Participation - Volunteering	35.2134 (38.8428)	41394	38.3623 (39.9773)	4882
TEACHMOM	Parent Participation - Teaching Assistance	12.1764 (23.4241)	41394	38.2821 (41.5357)	4882
FUNDMOM	Parent Participation - Fundraising	23.0784 (35.2134)	41394	59.6022 (44.0376)	4882
COUNCILMOM	Parent Participation - School government	36.4546 (37.2252)	41394	23.1174 (36.4406)	4882

Notes: The variables relate to the questionnaires administered to students in the rotated booklet. Items marked with *(r)* are taken from the rotated student questionnaire.The variable means of Dev7 and Vietnam are statistically different at the 5% significance level.

[6] [Hsin and Xie, 2014] investigate in detail the data from a set of longitudinal surveys that cover thousands of children over a long period starting from their early childhood through high school. As part of the explanation of the superior performance of Asian immigrant children, the authors report that "Asian students report greater parental expectations of academic success."

TIGERMOM refers to the reported proportion of parents who discussed their child's behavior or the child's progress "on their own initiative", to differentiate from cases where parents might have done so on the initiative of the teacher, termed as DUTYMOM. Table 4 shows a slightly higher number on DUTYMOM for Vietnamese parents compared to Dev7, but a greater difference, more than ten percentage points for TIGERMOM. VOLUMOM refers to parents volunteering in various non-academic activities, such as field trips or carpentry and yard work. Vietnamese parents appear to have slight advantage with regard to VOLUMOM, buta much higher advantage when considering TEACHMOM, which refers to parents volunteering as assistants to the teacher - 12.18% compared to 38.28%. Vietnamese parents also appear to be much more active in fund raising, from the values for FUNDMOM, though they may have less formal influence through school committees.

2.6 Teacher Characteristics

Conventional measures regarding student-teacher ratio and teacher certification show some advantage for Vietnam over Dev7 as shown in Table 6.

Table 6: Teacher characteristics and management

		Dev7 countries		Vietnam	
Variable	Description	MS	Valid N	MS	Valid N
Teacher numbers and teacher management					
PROPCERT	Proportion of certified teacher	0.6757 (0.4042)	35130	0.7961 (0.3978)	4586
SMRATIO	Mathematics teacher-student ratio	188.1791 (158.6256)	33985	120.9773 (43.6092)	4777
SC35Q02	Professional development in math in last 3 months	40.5068 (40.8546)	39550	49.0086 (45.1706)	4762
	teaching style	(0.9613)		(0.774)	
STUDREL *(r)*	Teacher student relations	0.3794 (1.0178)	25870	0.0186 (0.8883)	3253
TCH_INCENTV	Teacher appraisal linked to incentives	-0.0317 (1.0301)	41394	0.2687 (0.6336)	4882
Quality assurance of mathematics teachers through ...					
TCH_MENT	Teacher mentoring as quality assurance	0.8566 (0.3505)	40734	0.9859 (0.1181)	4882
TCM_PEER	Teacher peer review of lectures, methods etc	0.7916 (0.4061)	41095	0.8382 (0.3683)	4882
TCM_OBSER	Principal or senior staff observations	0.8015 (0.3989)	41170	0.9785 (0.1451)	4882
TCM_INSPE	Observation of classes external inspector	0.5882 (0.4922)	41020	0.8664 (0.3402)	4882

Notes: The variables relate to the questionnaires administered to students in the rotated booklet. Items marked with *(r)* are taken from the rotated student questionnaire. The variable means of Dev7 and Vietnam are statistically different at the 5% significance level.

The student-teacher ratio overall is not much different at about 20 students per teacher,

but there are more specialized mathematics teachers in Vietnam, as shown by the values for SMRATIO (121 in Vietnam compared to 188 for Dev7). There is a higher percentage of certified teachers in Vietnam, and higher reported professional development in mathematics (SC35Q02). A very interesting variable from a policy point of view regards incentives for teachers. School principals were questioned to what extent performance appraisal or other forms of feedback are related to incentives for teachers in seven different forms, from salary and bonus to public recognition and greater job responsibilities. The answers were on a 4 point scale: 'No change', 'A small change', 'A moderate change' and 'A large change'. We converted the rating into a Rasch index, scaled to an OECD mean of 0 and standard deviation of 1. The mean for Dev7 for this index, TCH_INCENTV was -0.03 for Dev7 and 0.27 for Vietnam, indicating greater presence of teacher incentives in Vietnam. The final set of variables in Table 6 deal with the way that quality assurance regarding teacher performance is carried out, with help of a mentor, peer, supervisor and external inspector. These variables indicate a higher prevalence of oversight for teachers in Vietnam, with the difference being greatest for external inspections (86.64% in Vietnam compared to 58.82% in Dev7 countries).

2.7 Pedagogical Practices

Pedaogical practices are an outcome of a complex interaction between curriculum and other policy, economic possibilities and the cultural and historical context. It is difficult to trace differences in these practices in a quantitative survey.[7] Table 7 presents a few variables that seek to capture variation in pedagogical practices. Table 7 indicates the higher prevalence of national policies in Vietnam regarding the use of computers in the classroom and the use of a standardized curriculum that specifies what has to be taught each month. There is no difference with regard to the use of a single textbook. There is some difference in the use of formative student assessment, with slightly higher percentage of use of assessments to monitor teachers and schools. COGACT is an OECD-PISA index variable based on response to student reports regarding classroom practices such as teacher requiring students to reflect on a problem or develop procedures rather than practice given procedures. This variables show much lower level of cognitive activation in Vietnam (-0.33) compared to 0.30 for Dev7. In the final set of classroom management variables, an interesting variation can be seen in DISCLIMA, an index variable that measures disciplinary climate in class, and is higher for Vietnam (0.38) as compared to Dev7 (-0.02).

[7]For an interesting recent qualitative study that seeks to emulate the TIMSS video study for Vietnam, see [Vu Dinh Phuong, 2014].

Table 7: Pedagogical practices

		Dev7 countries		Vietnam	
Variable	Description	MS	Valid N	MS	Valid N
Policies applied					
COMP_USE	Math policy - use of computers in class	0.4345 (0.4957)	40800	0.6447 (0.4787)	4815
TXT_BOOK	Math policy - same textbook	0.7905 (0.4069)	40557	0.7855 (0.4105)	4882
STD_CUR	Maths policy - standardized curriculum	0.8705 (0.3358)	40595	0.949 (0.22)	4882
Fromative assessment used to ...					
ASS_SCH	monitor the schools yearly progress	0.9111 (0.2846)	40555	0.9799 (0.1403)	4882
ASS_TCH	make judgements on teachers' effectiveness	0.7764 (0.4166)	40400	0.9912 (0.0934)	4882
Cognitive Activation in Mathematics					
COGACT *(r)*	Cognitive activation in mathematics lessons	0.2998 (0.975)	26217	-0.3278 (0.6647)	3249
Classroom Management					
STU_FEEDB	Seeking written feedback from students	0.7105 (0.4536)	40788	0.8419 (0.3649)	4882
CLSMAN *(r)*	Teacher classroom management (in math)	0.2394 (0.905)	25753	0.2163 (0.7761)	3252
DISCLIMA *(r)*	Disciplinary climate in class (in math)	-0.0243 (0.9055)	26242	0.3747 (0.6926)	3254

Notes: The variables relate to the questionnaires administered to students in the rotated booklet. Items marked with *(r)* are taken from the rotated student questionnaire. The variable means of Dev7 and Vietnam are statistically different at the 5% significance level.

2.8 School Characteristics

Table 8 indicates interesting basic differences between Vietnam and Dev7 school characteristics. Vietnamese schools are about half as likely to be private schools (8% compared to 17%) and less dependent on funding from student fees (in Vietnam, student fees account ofr 17% of the school's financing as compared to 26% for Dev7. One very useful comparison comes from a question regarding the geographic location of high schools. The percentage of schools reported in a VILLAGE (defined in PISA as population center below 3,000 inhabitants, was 46% in Vietnam as compared to 14% for Dev7. With CITY, defined as pouplation above 100,000 inhabitants, we find only 23% Vietnamese schools as compared to 41% for Dev7.

Table 8: School characteristics

		Dev7 countries		Vietnam	
Variable	Description	MS	Valid N	MS	Valid N
PRIVATESCL	Private school dummy variable	0.1714 (0.3768)	41182	0.0832 (0.2762)	4882
SC02Q02	Funding for school from student fees	25.7233 (36.0117)	34621	16.6104 (26.3564)	4848
VILLAGE	School located in a village	0.1403 (0.3473)	41347	0.4584 (0.4983)	4882
TOWN	School located in a town	0.4508 (0.4976)	41347	0.3101 (0.4626)	4882
CITY	School located in a city	0.4089 (0.4916)	41347	0.2315 (0.4218)	4882
CLSIZE	Average class size	35.013 (9.764)	40771	42.5043 (8.7236)	4882
SCHSIZE	Number of enrolled students at school	1057.0332 (924.2422)	35062	1302.9009 (648.6821)	4882
PCGIRLS	Proportion of girls at school	0.4900 (0.2597)	36342	0.5282 (0.0801)	4882

Notes: The variables relate to the questionnaires administered to students in the rotated booklet. Items marked with *(r)* are taken from the rotated student questionnaire. The variable means of Dev7 and Vietnam are statistically different at the 5% significance level.

The average class size in Vietnam is higher, with 43 students compared to 35 students in Dev7 countries, and the schools in Vietnam are bigger, with average enrollment of 1,303 students as compared to 1,057 in Dev7. There is also a slightly higher percentage of girls in the schools in Vietnam.

2.9 School Resources

The comparison of Vietnam and Dev7 regarding school resources shows a mixed picture (Table 9). The schools in Vietnam have lower number of computers per student (0.22) compared to 0.39 for Dev7. However, the ratio of computers connected to the internet is slightly higher in Vietnam (78% compared to 76%). Indices on quality of school educational resources (SCMATEDU) show Vietnam with -0.4941 and Dev7 with 0.8145, and similar difference for quality of physical infrastructure at the school (SCMATBUI). There is also a higher proportion of schools that offer additional math classes. These differences indicate that Vietnam has put in a priority for investment in Basic Education that compensates to some extent for its income disadvantage compared to the Dev7. With regard to extra-curricular activities, there is likewise a mixed picture. Not all extra-curricular activities are shows in Table 9, but some indicate lower prevalence in Vietnam as compared to Dev7 - for instance school band and math club (not shown, with similar pattern are chess club, IT club, art club). Some activities have higher prevalence in Vietnam - school play/musical, mathematics competition, and sports (not shown). It would appear that even for extra-

curricular activities, the prevalence of more serious activities or activities that require greater effort or competition are more prevalent in Vietnam as compared to Dev7.

Table 9: School resources and Management

		Dev7 countries		Vietnam	
Variable	Description	MS	Valid N	MS	Valid N
Resource quantity and quality					
RATCMP15	Available computers for 15-year-olds	0.3909 (0.5476)	39490	0.2216 (0.3411)	4875
COMPWEB	Ratio of computers connected to internet	0.7556 (0.3578)	37446	0.7795 (0.3109)	3634
SCMATEDU	Quality of school educational resources	-0.8145 (1.1538)	41373	-0.4941 (0.9718)	4882
SCMATBUI	Quality of physical infrastructure	-0.6322 (1.1113)	41221	-0.3988 (1.0161)	4882
SCL_EXTR_CL	School offers additional math classes	0.6538 (0.4757)	40869	0.9584 (0.1997)	4882
Extra-curriculars					
EXC1_BAND	School offers Band, orchestra or choir	0.4710 (0.4992)	40044	0.1678 (0.3737)	4882
EXC2_PLAY	School offers schoo play/musical	0.5928 (0.4913)	40122	0.8509 (0.3562)	4882
EXC5_MCLUB	School offers mathematics club	0.453 (0.4978)	40154	0.2687 (0.4434)	4882
EXC6_MATHCOMP	School offers Mathematics competition	0.6268 (0.4837)	40215	0.8032 (0.3977)	4882
EXC10_SPORT	School offers sporting activities	0.9321 (0.2516)	40581	0.992 (0.089)	4882
Leadership accountability and autonomy					
SCORE_PUBLIC	Achievement data posted publicly	0.345 (0.4754)	40965	0.7567 (0.4291)	4882
SCORE_AUTHRITS	Achievement data tracked by authority	0.8003 (0.3998)	41139	0.8282 (0.3773)	4778
SCHAUTON	School Autonomy in admin. decisions	-0.2542 (1.1328)	41394	-1.0419 (0.9378)	4882
TCHPARTI	Teacher participation in admin. decisions	-0.2169 (1.4457)	41394	-1.6445 (0.5188)	4882
LEADCOM	Communicating and acting on defined school goals	0.2387 (1.1105)	41252	0.0894 (0.6744)	4882
STUDCLIM	Student-related aspects of school climate	0.0485 (1.1642)	40973	0.0418 (0.6849)	4874
TEACCLIM	Teacher-related aspects of school climate	-0.1997 (1.1474)	40973	-0.0873 (0.7125)	4874

Notes: The variables relate to the questionnaires administered to students in the rotated booklet. Items marked with *(r)* are taken from the rotated student questionnaire. The variable means of Dev7 and Vietnam are statistically different at the 5% significance level.

With regard to school leadership and autonomy, there appears to be less autonomy and more accountability in Vietnam. The index variable SCHAUTON indicates a Dev7 mean of -0.2542 that is higher than the Vietnam mean of -1.0419 (recall that indices are set to OECD mean of zero). Teachers in Vietnam have lower chances to particpate in school management - TCHPARTI indicates Dev7 mean of -0.2169 compared to 1.6445 for Vietnam.

Principals are more likely in Dev7 to say that they communicate and act on information (LEADCOM), but there is much higher prevalence of public posting of school achievement data (SCORE_PUBLIC). Interestingly, even Dev7 countries have high level of tracking of achievement data by authorities (80% of schools report this). Finally, with regard to the school climate, indices described further in the PISA documentation, STUDCLIM (student climate) is roughly even between Vietnam and Dev7, but TEACCLIM (teacher climate), that incluses variables such as teacher absenteeism and teacher expectations of students is higher for Vietnam.

2.10 Preliminary conclusions from comparison of endowments

In summary of the mean comparisons between Vietnam and Dev7 students, we do find a number of potentially insightful results. Consider the four-fold classification of factors presented in the conceptual diagram of Figure 3 - students, teachers, the school and the parents.

Students: Students in Vietnam are more likely to have attended Pre-school, and less likely to have repeated grades in the past. They are likely to have been more disciplined at school, skip fewer classes, and assume greater responsibility for their own learning. The Vietnamese students are less likely to brag about their abilities and experience and yet work harder, especially out of school, in extra classes. Vietnamese students tend to have lower anxiety about mathematics and higher confidence about the usefulness of mathematics in their future.

Parents: Parents in Vietnam are likely to be more involved in the school life of their children than the parents in Dev7 countries. Though time spent on homework help is similar in both groups, Vietnamese parents are more likely to volunteer at school, take part in fundrasing for the school, and help the teachers as classroom assistants. Vietnamese parents are also more likely to seek to meet the teacher to discuss their child's progress or the child's behavior on the parent's own initiative. Principals in Vietnam report higher level of parental pressure.

Teachers: Teachers have similar level of formal education in both groups, but Vietnamese teachers may have had more recent professional development activity. There are more specialist mathematics teachers at high schools in Vietnam, and teachers overall also are more likely to be certified. The performance of teachers is more likely to be monitored in Vietnam, with higher emphasis on student achievement and on making public the information about that achievement. Teachers also tend to have lower autonomy, more likely

to be subject to centralized policies and work in an environment with higher prevalence of incentives for performance. Principals report fewer problems with regard to problems like teacher absenteeism, which squares with a cultural explanation about a Confucian heritage culture.

Schools: Vietnam has a much lower level of economic development as compared to the Dev7 countries, which is reflected in lower levels of education of parents and lower level of home possessions, including so called cultural possessions such as artwork and books. Also, many more Vietnamese students go to school in villages and small towns, reflecting the national population distribution. Yet, two things are striking about schools - though schools have fewer computers as compared to Dev7 countries, those computers are as likely as Dev7 countries to be connected to the internet. Also, indices regarding quality of school infrastructure and school educational resources are less deficient in Vietnam compared to Dev7, indicative of substantive investments in schools in the past few decades.

Overall, across these four domains of information, it seems likely that the PISA dataset is able to detect significant cultural differences between the context of Vietnam and Dev7 countries. There appears to be some influence of policy, with student achievement assessment and teacher incentives, and higher levels of centralized controls, but the effectiveness of such policies is also likely tied to cultural factors. Unlike the 'World Values Survey' the set of PISA instruments is not suited to clearly identify cultural differences, for instance through responses regarding beliefs, attitudes and practices defined so as to discriminate between cultures. While mean differences provide interesting hints, they are essentially bi-variate correlations. In order to tell us more about the correlations, which ones are more important than others, and whether indeed some unobservable 'Vietnamese culture' variable may be a plausible explanation, we need to unravel the mystery further through a study of multi-variate correlations. We do this first using the Fryer-Levitt approach.

3 Regression Approach I: Fryer-Levitt

We are now ready to continue unravelling the secret a bit further by deepening our analytical approach beyond a mere comparison of means. We adopt a simple methodology that is easy to understand and interpret. Our approach closely follows [Fryer and Levitt (2004)] who sought to explain the achievement gap for Black children in the US. For the results presented in this section, we pool the student level data from Vietnam and Dev7 countries. Recall that Dev7 stands for the 7 developing countries in the PISA dataset for 2012 with a per capita income below the cut-off of US$10,000. The reason for focussing on developing

countries is because we want to have a common support with regard to a country's wealth. If a rich country shows outstanding results, perhaps it may be of interest to other rich countries which do not do so well, but it is hardly of great interest to a poor country. But if a poor country does very well, and stands out from the pack of poor countries that mostly do poorly in PISA, readers from poor countries want to know what can explain such a phenomenon, since it clearly cannot be the wealth of the country (as captured albeit imperfectly by per capita GDP). We start by looking at Mathematics, with the identical approach being used for the three PISA subjects.

3.1 Mathematics

We estimate a weighted least squares regression of student level test scores as follows[8]:-

$$TESTSCORE_i = VIETNAM_i'\gamma + X_i'\Theta + \epsilon_i \tag{1}$$

A key estimate of interest is γ, the coefficient on $VIETNAM$, a 0 or 1 dummy variable. Regressions are run in a sequence, starting from one without any covariates in X, and then adding variables in groups to expand X in consecutive columns in Table 10. Column (1) in Table 10 shows that the Vietnam dummy had a coefficient of 128.05. By construction, this is the absolute difference in means between Vietnam and the Dev7 countries. Now we want to see the extent to which observable variables included in the PISA dataset can help to explain this large gap of 128.05. The first set of variables included in the regression reported in column (2) concern the students themselves. The student characteristics were - whether they went to Pre-School, repeated a grade in the past, and how often they are late for school (ST08Q01) or skipped classes(ST115Q01). With these variables included, the coefficient on the dummy or the Vietnamese advantage or gap, comes down by nearly 20 points, or roughly 0.2 standard deviation units, to 108.91. In other words, one key reason that the Vietnam gap is so high is because of these student related variables - this result was hinted at in the endowment comparison presented earlier in Section 2. Note that of the four student variables used in column (2), only two are statistically significant. The method used for variable selection depends on whether or not, within each group of variables (students, parents etc.) the inclusion of the variable leads to a reduction of the coefficient on the Vietnam dummy.

[8]This is a simplification for presentation of the main idea. In PISA, the test score is not provided as a single value but as a set of five plausible values for each student, and complex algorithms have to be used for weighting based on a method called Fay's variant of Balanced Repeated Replication (BRR). The details are provided in the PISA technical manual. In this paper, we utilize the R *intsvy* package for implementation.

Table 10: The estimated impact of 'Vietnam' on Mathematics PISA test scores

	Mathematics									
Variables	(1)		(2)		(3)		(4)		(5)	
VIETNAM	128.05	(5.65)	108.91	(5.32)	97.46	(5.48)	95.13	(5.87)	77.26	(7.84)
PRESCHOOL	-	-	45.86	(3.92)	40.54	(3.95)	39.21	(4.09)	24.90	(3.80)
REPEAT	-	-	-50.57	(2.59)	-47.55	(2.56)	-45.05	(3.19)	-36.96	(3.00)
ST08Q01	-	-	-8.59	(1.20)	-8.41	(1.18)	-8.38	(1.33)	-7.84	(1.32)
ST115Q01	-	-	-4.94	(1.70)	-4.57	(1.73)	-6.10	(1.80)	-5.40	(1.86)
BOOK_N	-	-	-	-	0.09	(0.01)	0.08	(0.01)	0.07	(0.01)
PARPRESSURE	-	-	-	-	10.73	(5.01)	12.51	(4.78)	10.02	(4.40)
FUNDMOM	-	-	-	-	0.27	(0.06)	0.24	(0.06)	0.19	(0.07)
COUNCILMOM	-	-	-	-	-0.14	(0.06)	-0.18	(0.06)	-0.10	(0.07)
DUTYMOM	-	-	-	-	-0.07	(0.06)	-0.12	(0.07)	-0.10	(0.07)
PROPCERT	-	-	-	-	-	-	16.08	(5.50)	16.32	(6.87)
SMRATIO	-	-	-	-	-	-	-0.01	(0.01)	-0.03	(0.01)
TCSHORT	-	-	-	-	-	-	-1.91	(1.97)	2.24	(1.87)
TCFOCST	-	-	-	-	-	-	0.30	(2.19)	-1.45	(1.88)
TCM_STUASS	-	-	-	-	-	-	10.85	(7.45)	-0.18	(7.85)
TCM_PEER	-	-	-	-	-	-	-1.53	(6.59)	-5.61	(5.65)
TCH_INCENTV	-	-	-	-	-	-	-0.92	(2.73)	-2.75	(2.72)
ASS_PROG	-	-	-	-	-	-	-0.51	(15.51)	-22.58	(8.04)
ASS_PROM	-	-	-	-	-	-	7.60	(6.11)	14.09	(5.80)
ASS_SCH	-	-	-	-	-	-	5.51	(6.51)	0.51	(7.31)
STU_FEEDB	-	-	-	-	-	-	3.66	(5.27)	2.20	(5.07)
PCGIRLS	-	-	-	-	-	-	-	-	14.59	(13.65)
COMP_USE	-	-	-	-	-	-	-	-	-1.57	(5.30)
TXT_BOOK	-	-	-	-	-	-	-	-	-9.51	(7.05)
TOWN	-	-	-	-	-	-	-	-	-9.53	(3.76)
CLSIZE	-	-	-	-	-	-	-	-	0.81	(0.23)
COMPWEB	-	-	-	-	-	-	-	-	15.28	(6.31)
SCMATEDU	-	-	-	-	-	-	-	-	5.58	(2.94)
SCMATBUI	-	-	-	-	-	-	-	-	3.46	(2.45)
EXC2_PLAY	-	-	-	-	-	-	-	-	8.69	(3.96)
EXC6_MATHCOMP	-	-	-	-	-	-	-	-	-1.70	(5.37)
EXC10_SPORT	-	-	-	-	-	-	-	-	-5.65	(9.15)
EXC11_UNICORN	-	-	-	-	-	-	-	-	6.81	(5.59)
SCL_EXTR_CL	-	-	-	-	-	-	-	-	10.90	(5.08)
SCORE_PUBLIC	-	-	-	-	-	-	-	-	10.10	(4.75)
QUAL_RECORD	-	-	-	-	-	-	-	-	6.99	(6.77)
SCHSEL	-	-	-	-	-	-	-	-	1.57	(3.29)
R2	27.21		37.24		40.03		43.15		43.93	
N	48483		46267		44046		30051		25612	

Figures in parentheses are t-values for null hypothesis of coefficent being statistically significantly different from zero

The second set of variables, included cumulatively while retaining the earlier variables, relates to the home background and parents of the student. In the earlier section on endowments, it was explained that Vietnam has the lowest per capita income of the eight countries included in the sample and it is natural that PISA indices on family wealth or parental education are much lower for Vietnam as compared to the other 7 countries. Inclusion of these variables increase the coefficient on the Vietnam dummy and would take us away from our objective. So the parent variables that are retained and presented in column (3) are only those variables that reduce the Vietnam dummy. The reduction is only 11 points as

compared to the nearly 20 point reduction for student variables. Only the variable PARPRESSURE is statistically significant in this group and indeed has a sizeable impact on the score - this variable reports on principals claiming that "there is *constant pressure* from many parents, who expect our school to set very high academic standards and to have our students achieve them."

The third set of variables, related to teachers, is presented in column (4) and result in a reduction of the Vietnam dummy only by 3 points to a level of 95.13. This does not mean that teachers are unimportant as a reason for Vietnam's superior performance in mathematics, only that the observed teacher related variables in the PISA dataset do not collectively help as an explanation. In the regression itself, one can see that PROPCERT, proportion of certified teachers affects mathematics scores positively. The same goes for a slew of student assessment related variables, including TCM_STUASS and STU_FEEDB, that relate to the use of student assessment and student written feedback for assessing teachers.

The final set of variables, related to schools is presented in column (5), that results in a further reduction in the dummy coefficient by 12 points, to a level of 77.26. Interestingly, all the school variables with just one exception are statistically significant. There are interesting insights from some of the school variables. COMP_USE does not have a positive effect on the mathematics score, but COMPWEB does - recall that Vietnam does relatively better on internet connectivity as compared to mere availability of computers. The presence of SCMATEDU and SCMATBUI - quality of school educational resources and quality of physical infrastructure - helps explain the gap, recall from Table 9 the superior endowments in Vietnam compared to the Dev7 countries. Table 10 shows that Extra classes organized by the school (SCL_EXTR_CL), and the "systematic recording of data including teacher and student attendance and graduation rates, test results and professional development of teachers" (QUAL_RECORD) are also part of the story. Public dissemination by the school of testing results (SCORE_PUBLIC) has a positive coefficient, which is also one of the variable where Vietnam appears to have twice as many schools following the practice.

Note that as we successively add variables in the regression equation, the number of observations in the regressions drops because of missing values. Comparing available observable values on other variables (not reported) for the dropped observations does not seem to indicate a systematic bias in this attrition. The 2012 round of the PISA included a so called 'rotated' module for the student's questionnaire. The idea of the rotated module was to ask additional sets of questions in a systematic division that increases the overall data available without burdening the respondents with excessively long questionnaires. Appendix Table 1 reports the regression results that includes the 'rotated' variables related to different factors

such as parents, teachers and the school. It can be seen from Appendi Tablex 1 that the lowest possible alue for the Vietnam dummy was 46.2. This means that after accounting for a range of variables, there is still a substantive (half a standard deviation) score gap that is left uneplained from the PISA data. It is quite likely that this unexplained factor is a combination of cultural effects beyond the discipline and effort related variables covered by PISA. We now turn to an examination of the science and reading results.

3.2 Reading

Table 11: The estimated impact of 'Vietnam' on Reading PISA test scores

	Reading									
Variables	(1)		(2)		(3)		(4)		(5)	
VIETNAM	105.16	(5.03)	85.03	(4.33)	75.49	(4.69)	74.13	(4.59)	60.31	(6.06)
FEMALE	-	-	25.84	(1.69)	25.24	(1.74)	26.65	(1.87)	23.02	(1.60)
PRESCHOOL	-	-	41.47	(3.67)	37.82	(3.78)	34.86	(3.83)	23.68	(3.40)
REPEAT	-	-	-58.46	(2.79)	-55.65	(3.03)	-52.78	(3.50)	-43.16	(3.24)
ST08Q01	-	-	-8.6	(1.29)	-8.31	(1.25)	-8.03	(1.38)	-7.67	(1.43)
ST115Q01	-	-	-7.9	(1.66)	-7.82	(1.68)	-9.63	(1.78)	-8.37	(1.80)
BOOK_N	-	-	-	-	0.08	(0.01)	0.06	(0.01)	0.05	(0.01)
PARPRESSURE	-	-	-	-	10.72	(4.32)	11.21	(4.24)	3.69	(4.18)
VOLUMOM	-	-	-	-	-0.05	(0.06)	-0.06	(0.05)	-0.03	(0.06)
FUNDMOM	-	-	-	-	0.2	(0.06)	0.17	(0.05)	0.1	(0.06)
COUNCILMOM	-	-	-	-	-0.12	(0.06)	-0.16	(0.06)	-0.09	(0.07)
DUTYMOM	-	-	-	-	-0.05	(0.06)	-0.08	(0.06)	-0.1	(0.07)
PROPCERT	-	-	-	-	-	-	9.14	(4.84)	3.14	(5.59)
TCSHORT	-	-	-	-	-	-	-3.93	(1.84)	0.47	(1.91)
TCM_STUASS	-	-	-	-	-	-	15.91	(7.67)	7.35	(8.22)
ASS_PROG	-	-	-	-	-	-	4.31	(16.68)	-15.85	(10.44)
ASS_PROM	-	-	-	-	-	-	6.25	(5.45)	13.94	(5.92)
ASS_NAT	-	-	-	-	-	-	5.71	(5.41)	1.44	(5.91)
ASS_CUR	-	-	-	-	-	-	-2.15	(7.50)	-3.81	(9.14)
STU_FEEDB	-	-	-	-	-	-	4.95	(4.67)	6.12	(4.28)
PCGIRLS	-	-	-	-	-	-	-	-	23.11	(10.88)
TOWN	-	-	-	-	-	-	-	-	-6.88	(3.71)
CLSIZE	-	-	-	-	-	-	-	-	0.95	(0.23)
COMPWEB	-	-	-	-	-	-	-	-	16.08	(5.56)
SCMATEDU	-	-	-	-	-	-	-	-	5.16	(2.52)
SCMATBUI	-	-	-	-	-	-	-	-	0.98	(2.35)
EXC2_PLAY	-	-	-	-	-	-	-	-	13.05	(3.92)
EXC6_MATHCOMP	-	-	-	-	-	-	-	-	7.65	(5.02)
EXC10_SPORT	-	-	-	-	-	-	-	-	-4.43	(10.86)
EXC11_UNICORN	-	-	-	-	-	-	-	-	10.61	(5.11)
SCORE_PUBLIC	-	-	-	-	-	-	-	-	5.24	(4.22)
LEADINST	-	-	-	-	-	-	-	-	1.54	(2.03)
QUAL_RECORD	-	-	-	-	-	-	-	-	-6.06	(7.10)
SCHSEL	-	-	-	-	-	-	-	-	1.39	(3.05)
TEACCLIM	-	-	-	-	-	-	-	-	-1.3	(2.63)
R2	19.61		34.5		36.45		39.63		41.84	
N	48483		46267		44046		35442		27331	

For scores in reading (Table 11), the gap is lower to begin with: 105.16 compared to 128.05 for mathematics. The pattern revealed by the regessions is quite similar: the student specific variables brings down the Vietnam dummy coeffcient to 85; the parent related variables accounts for a further 9 points, with only PARPRESSURE being statistically significant.

3.3 Science

Table 12: The estimated impact of 'Vietnam' on Science PISA test scores

	Science									
Variables	(1)		(2)		(3)		(4)		(5)	
VIETNAM	134.56	(4.91)	115.45	(4.37)	103.85	(4.56)	101.83	(4.63)	88.54	(5.99)
FEMALE	-	-	-2.21	(1.61)	-3.08	(1.58)	-1.95	(1.78)	-4.93	(1.46)
PRESCHOOL	-	-	43.06	(3.42)	37.94	(3.60)	36.09	(3.65)	24.36	(3.57)
REPEAT	-	-	-53.8	(2.63)	-50.74	(2.86)	-48.77	(3.26)	-40.6	(3.24)
ST08Q01	-	-	-9.28	(1.11)	-9.02	(1.16)	-8.76	(1.28)	-8.45	(1.26)
ST115Q01	-	-	-5.95	(1.68)	-5.55	(1.72)	-6.88	(1.81)	-6.18	(1.89)
BOOK_N	-	-	-	-	0.08	(0.01)	0.06	(0.01)	0.06	(0.01)
PARPRESSURE	-	-	-	-	5.87	(3.94)	6.42	(3.96)	-0.27	(3.76)
FUNDMOM	-	-	-	-	0.25	(0.05)	0.22	(0.05)	0.17	(0.06)
COUNCILMOM	-	-	-	-	-0.18	(0.05)	-0.2	(0.05)	-0.12	(0.06)
DUTYMOM	-	-	-	-	-0.02	(0.05)	-0.04	(0.06)	-0.06	(0.07)
PROPCERT	-	-	-	-	-	-	11.67	(4.91)	5.35	(5.47)
TCSHORT	-	-	-	-	-	-	-0.68	(1.82)	2.7	(1.83)
TCM_STUASS	-	-	-	-	-	-	16.04	(6.86)	9.43	(7.39)
TCM_PEER	-	-	-	-	-	-	-2.65	(6.05)	-5.68	(5.79)
ASS_PROG	-	-	-	-	-	-	-5.42	(16.65)	-23.01	(9.17)
ASS_PROM	-	-	-	-	-	-	4.23	(5.79)	12.41	(6.68)
ASS_NAT	-	-	-	-	-	-	7.33	(4.71)	1.49	(4.93)
ASS_SCH	-	-	-	-	-	-	2.63	(6.01)	0.66	(6.72)
ASS_CUR	-	-	-	-	-	-	-6.33	(7.54)	-4.87	(8.73)
STU_FEEDB	-	-	-	-	-	-	4.59	(4.48)	5.90	(4.68)
PCGIRLS	-	-	-	-	-	-	-	-	18.92	(11.11)
PRIVATESCL	-	-	-	-	-	-	-	-	-1.03	(5.34)
TOWN	-	-	-	-	-	-	-	-	-8.25	(3.16)
CLSIZE	-	-	-	-	-	-	-	-	0.83	(0.20)
COMPWEB	-	-	-	-	-	-	-	-	17.22	(5.61)
SCMATEDU	-	-	-	-	-	-	-	-	5.38	(1.75)
EXC2_PLAY	-	-	-	-	-	-	-	-	8.52	(3.34)
EXC6_MATHCOMP	-	-	-	-	-	-	-	-	1.5	(4.53)
EXC10_SPORT	-	-	-	-	-	-	-	-	-1.73	(9.13)
EXC11_UNICORN	-	-	-	-	-	-	-	-	8.56	(5.04)
SCORE_PUBLIC	-	-	-	-	-	-	-	-	11.28	(3.75)
LEADINST	-	-	-	-	-	-	-	-	1.57	(1.95)
QUAL_RECORD	-	-	-	-	-	-	-	-	-0.92	(5.55)
SCHSEL	-	-	-	-	-	-	-	-	1.52	(3.11)
TEACCLIM	-	-	-	-	-	-	-	-	-1.82	(2.46)
R2	30.75		41.14		43.35		46.11		47.35	
N	48483		46267		44046		35302		27224	

Teacher related variables only account for a further decline of 2 points, and school variables reduced the dummy by 14 points to a level of 60.31 reported in column (5). Appendix

Table 2 reports the regressions with rotated variables included that brings down the dummy coefficient to a lowest level of 52.04. Finally we look at the Science results.

The gap between Vietnam and Dev7 is largest for science scores, with column (1) in Table 12 indicating the dummy coefficient at 134.56. The pattern of reduction in the dummy again is similar to the previous two subjects - with the student specific variables accounting for the biggest decline (19 points) , followed by school (13 points), parents (11.5 points) and teacher related variables (2 points). The final specification shows the dummy come down only to 88.54. Together with rotated variables, presented in Appendix Table 3, the lowest value for Vietnam dummy is 80.09, a total reduction of 55 points explained, but still leaving a large portion unattributable through the Fryer-Levitt method.

3.4 Summarized insights from Fryer-Levitt

We can see that across the three test subjects, the Vietnam dummy comes down roughly by nearly 50 points, or half a standard deviation. We find that student, parent and school related variables appear to explain part of Vietnam's superior performance in PISA.

Students: The student related variables reflect two policy elements that could be useful for other countries that seek to learn from Vietnam. The investment made by Vietnam in Preschool appears to have long lasting effects, and indeed in Vietnam the government continues to invest deeply not only for universal Preschool, but also for early childhood care services even prior to Preschool. A policy lesson also can be derived from the effect of repetition - cause and effect is difficult to extract in the case of repetition and test score performance, but one can see that repetition is much lower in Vietnam (Table 1) and the regression coefficient on a student being a repeater has a large negative value, even in the final specification with all other variables included. The other student related variables regarding being late for school and skipping classes perhaps do not have clear policy implications for other countries but help us understand a cultural effect regarding Vietnam.

Parents: As noted in the text, the houshold wealth/possessions and parents education levels and socio-economic index reflects Vietnam's per capita GDP and act against explanations of the test score gap. Following the trend, Vietnam would have benefited from a much higher gap if it had been a wealthier country. There is an advantage that Vietnamese children have in having more demanding parents (though perhaps Vietnamese teenagers may not always see it that way). The parents are demanding not only of their children, but apparently also from the schools, and the parents appear to back up their demands by contributing on their own as volunteers. Interestingly, even though the individual coeffi-

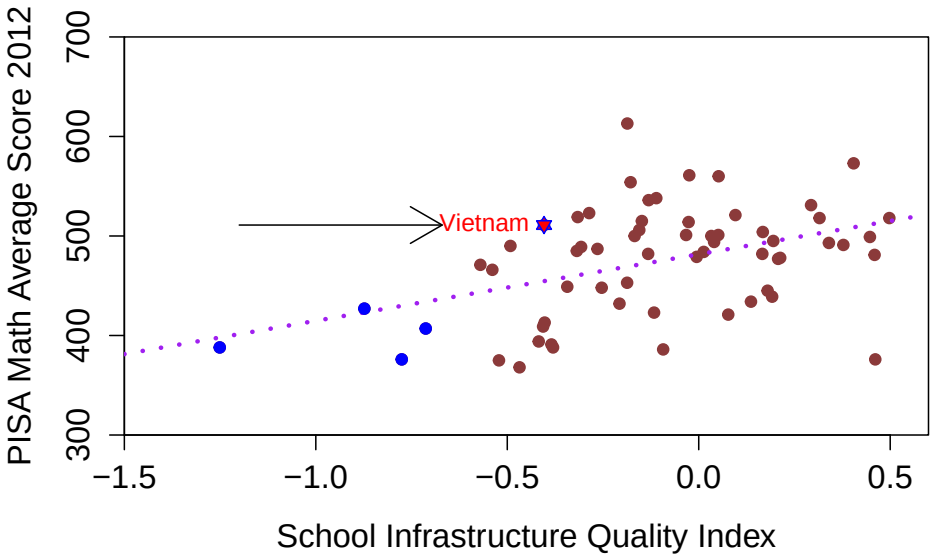
cients of parent related variables are not statistically significant except for one variable, the variables appear to collectively ifluence the dummy coefficient upto one-tenth of a standard deviation of test scores. As Amy Chua attested, parental attitudes and behaviors are deeply influenced by cultural norms. However, there is a policy lesson here too, concerning the freedom of access provided to parents to take part in the school life. Sometimes schools tend to be insular places without much scope for parents to contribute, but measures to harness parents contributions in their time as well as in cash and kind may yield positive results.

Teachers: Teachers are widely recognized to be the most important factor in many studies of student achievement. Yet, in this case, the inclusion of a number of teacher related variables does not appear to be useful in explaining Vietnam's achievement gap. It is interesting to note that in the regressions the variables individually tend to have statistically significant coefficients, but do not affect the dummy coefficient. One teacher variable, the proportion of certified teachers, is clearly economically and statistically significant, and it is one where Vietnam has an advantage (80% vs. 68% certified). Variables that relate to the use of student assessment in providing feedback to teachers on their performance are seen to be important. The presence of other assessment and feedback related variables are also in line with intuition. It is possible that the advantages which Vietnam enjoys with regard to teachers are swamped by the effects of variables for which Vietnam does not have an advantage, so the net result is that the gap is not explained by PISA related teachers variables. Also, it is possible that the effect of teachers is particularly context specific, so we can see the weakness of the pooled regression approach of Fryer-Levitt. This last explanation is further investigated in the next section of the paper.

School School resources matter with regard to PISA results in the international perspective - as the scatter plot in Figure 1 motivating this paper clearly shows. Developing countries with the notable exception of Vietnam are clustered at the bottom left hand side. And there is a positive slope with high scoring countries tending to be on the higher income side. In this section, we see that the effort made by the Vietnamese government to invest in education plays an important part in explaining the achievement gap. Even though Vietnam may be poor with regard to per capita income, it is not as poor with regard to the quality of educational resources and the quality of physical infrastructure. This can be seen in Figure 4, where Vietnam moves to the right as compared to Figure 1, where it was the country with the lowest GDP per capita in the entire dataset. With regard to the SCMATBUI (quality of school infrastructure) jumps ahead 11 places, with a similar story (not shown) regarding SCMATEDU (quality of educational materials). A key reason is the investments in schools in smaller towns and rural areas as classified by PISA - the dispersion of school infrastructure

is lower in Vietnam as compared to other countries.

Figure 4: PISA 2012 results compared with School Infrastructure Quality



Source:OECD-PISA database

4 Regression Approach II: Oaxaca-Blinder Decomposition

A key weakness of the Fryer-Levitt approach is that the pooled regression does not allow for regression coefficients to be different across the different countries. In this section, we set aside the Fryer-Levitt method to look at the data from a different analytical perspective. We use the Oaxaca-Blinder decomposition (OB) method, that has recently become quite popular for PISA analysis after the initial work by [Ammermueller 2007] comparing the PISA results of Finland and Germany.

The objective of OB is to decompose the mean differences, in this case the mean difference in Vietnam's PISA performance with each of the DEV7 countries. Extensions of OB allow for decompositions to be made throughout the distribution rather than only at the mean values, but we leave such an extension for further research regarding Vietnam's PISA performance. OB is based on a simple algebraic rearrangement of terms of the OLS regression of test scores. The mean outcome difference to be explained ($\Delta\bar{\boldsymbol{Y}}$) is simply the difference of the mean outcomes for Vietnam and the comparison country. let us denote the scores as $\bar{\boldsymbol{Y}}_V$ and $\bar{\boldsymbol{Y}}_O$, respectively:

$$\Delta\bar{\boldsymbol{Y}} = \bar{\boldsymbol{Y}}_V - \bar{\boldsymbol{Y}}_O \tag{2}$$

Now, as the OLS error terms are of mean zero by construction, (2) can be represented by

$$\Delta\bar{\boldsymbol{Y}} = \bar{\boldsymbol{X}}'_V\hat{\boldsymbol{\beta}}_V - \bar{\boldsymbol{X}}'_O\hat{\boldsymbol{\beta}}_O \tag{3}$$

In the twofold version of OB that we use in this paper, (3) can be represented either as

$$\Delta\bar{\boldsymbol{Y}} = \underbrace{(\bar{\boldsymbol{X}}_V - \bar{\boldsymbol{X}}_O)'\hat{\boldsymbol{\beta}}_V}_{\text{endowments}} + \underbrace{\bar{\boldsymbol{X}}'_O(\hat{\boldsymbol{\beta}}_V - \hat{\boldsymbol{\beta}}_O)}_{\text{coefficients}} \tag{4}$$

or as

$$\Delta\bar{\boldsymbol{Y}} = \underbrace{(\bar{\boldsymbol{X}}_V - \bar{\boldsymbol{X}}_O)'\hat{\boldsymbol{\beta}}_O}_{\text{endowments}} + \underbrace{\bar{\boldsymbol{X}}'_V(\hat{\boldsymbol{\beta}}_O - \hat{\boldsymbol{\beta}}_V)}_{\text{coefficients}} \tag{5}$$

depending on which country is used as a reference country. It is not obvious which of the specifications (4) or (5) is better, and OB variations include further combinations of the two. We focus on this paper on the approach of (4), using Vietnam as the reference, and leave (5) and additional OB variations to subsequent research.

We base the choice of regression specification on the findings so far. The previous analysis has shown that there in line with per capita GDP of Vietnam compared with the Dev7 variables, there is a series of income level or wealth related variables for which Vietnam does poorly in comparison with Dev7.

As Equation 4 shows, the decomposition divides the difference in mean outcomes into a portion that is explained by cross-group differences in the explanatory variables, and a part that remains unexplained by these differences.

The unexplained portion of the mean outcome gap has often been attributed to discrimination, but may also result from the influence of unobserved variables. It can be further decomposed into two sub-components, labeled "unexplained A" and "unexplained B" above. If one interprets the reference coefficient vector to be non-discriminatory, these sub-components measure the part of the mean difference in outcomes that originates from discrimination in favor of Group A and the part that comes from discrimination against Group B, respectively.

Again, a detailed, variable-by-variable decomposition can also be estimated:

$$\underbrace{(\bar{\boldsymbol{X}}_A - \bar{\boldsymbol{X}}_B)'\hat{\boldsymbol{\beta}}_R}_{\text{explained}} = \underbrace{(\bar{X}_{1A} - \bar{X}_{1B})\hat{\beta}_{1R}}_{\text{variable 1}} + \underbrace{(\bar{X}_{2A} - \bar{X}_{2B})\hat{\beta}_{2R}}_{\text{variable 2}} + \dots \tag{6}$$

$$\underbrace{\bar{\boldsymbol{X}}'_A(\hat{\boldsymbol{\beta}}_A - \hat{\boldsymbol{\beta}}_R)}_{\text{unexplained A}} = \underbrace{\bar{X}_{1A}(\hat{\beta}_{1A} - \hat{\beta}_{1R})}_{\text{variable 1}} + \underbrace{\bar{X}_{2A}(\hat{\beta}_{2A} - \hat{\beta}_{2R})}_{\text{variable 2}} + \dots \tag{7}$$

$$\underbrace{\bar{\boldsymbol{X}}'_B(\hat{\boldsymbol{\beta}}_R - \hat{\boldsymbol{\beta}}_B)}_{\text{unexplained B}} = \underbrace{\bar{X}_{2B}(\hat{\beta}_{2R} - \hat{\beta}_{2B})}_{\text{variable 1}} + \underbrace{\bar{X}_{2B}(\hat{\beta}_{2R} - \hat{\beta}_{2B})}_{\text{variable 2}} + \dots \tag{8}$$

The choice of the reference coefficients is generally up to the researcher. In the literature on labor market discrimination, it is often assumed that only one of the two groups faces discrimination – for instance, that only women or members of ethnic minorities are discriminated against. In such cases, the reference coefficients will simply be the coefficients from a regression on observations from one of the groups: either $\hat{\boldsymbol{\beta}}_R = \hat{\boldsymbol{\beta}}_A$ or $\hat{\boldsymbol{\beta}}_R = \hat{\boldsymbol{\beta}}_B$.

Some researchers have instead used a weighted average of $\hat{\boldsymbol{\beta}}_A$ and $\hat{\boldsymbol{\beta}}_B$ as the set of reference coefficients. Reimers1983, for example, proposes giving equal weight to coefficients from regressions on Group A and Group B observations:

$$\hat{\boldsymbol{\beta}}_R = 0.5\hat{\boldsymbol{\beta}}_A + 0.5\hat{\boldsymbol{\beta}}_B \tag{9}$$

[**?**] suggests weighting the coefficients by the proportion of observations in the corresponding group:

$$\hat{\boldsymbol{\beta}}_R = \frac{n_A}{n_A + n_B}\hat{\boldsymbol{\beta}}_A + \frac{n_B}{n_A + n_B}\hat{\boldsymbol{\beta}}_B \tag{10}$$

Others still have advocated the use of coefficient estimates from a regression that pools observations from both Groups A and B, and includes Jann2008 or does not include Neumark1988 the group indicator variable as an additional regressor. The oaxaca package estimates results for all of the aforementioned choices of $\hat{\boldsymbol{\beta}}_R$, and also enables users to specify their own custom weights for $\hat{\boldsymbol{\beta}}_A$ and $\hat{\boldsymbol{\beta}}_B$ to construct a weighted average-based set of reference coefficients.

Table 13: Add caption

		Mean value		Albania		Colombia	
Ed wealth		*Dev 7*	*Vietnam*	*Endowments*	*Coefficients*	*Endowments*	*Coefficients*
	INTERCEPT				100.7672		222.6017
Students	PRESCHOOL	0.7888	0.9120	3.2606	14.9314	0.1767	21.2995
	LATESCHOOL	1.5131	1.1872	2.5351	-16.1377	2.7914	-7.0084
	NOREPEAT	0.8085	0.9321	-1.1335	81.2394	9.752	5.6154
	SHRS	3.7566	3.9597	8.103	18.4864	2.1314	10.2102
Parents	OUTMATH	1.8280	3.1305	8.4386	5.9606	6.3176	13.8406
	PARPRESSURE	0.2665	0.3837	3.0612	6.3037	5.2187	5.0041
	TIGERMOM	52.4472	62.4183	0.0071	-3.1614	0.1958	-14.0056
	TEACHMOM	12.1764	38.2821	2.4287	4.431	2.0458	2.2844
Teachers	PROPCERT	0.6757	0.7961	-0.8449	6.3679	-	-
	MATHPROFDEV	40.5068	49.0086	0.0438	-0.353	-0.2447	-1.0775
	TCH_INCENTV	-0.0317	0.2687	3.4264	-2.4877	1.9205	-0.4223
	TCM_INSPE	0.5882	0.8664	-5.3929	-17.0471	-13.3928	-0.5966
	TCM_OBSER	0.8015	0.9785	-	-	-9.9929	-5.2289
	COMP_USE	0.4345	0.6447	-0.4169	-0.3193	-0.6	3.1921
	STU_FEEDB	0.7105	0.8419	-0.5089	7.7288	-0.3485	-10.1593
Schools	EXC6_MATHCOMP	0.6268	0.8032	0.3129	-9.663	-0.6487	-8.5158
	SCMATBUI	-0.6322	-0.3988	0.388	-4.8621	-0.0363	-0.4844
	SCL_EXTR_CL	0.6538	0.9584	-3.5828	-2.9217	-8.4506	-9.4397
	SCORE_PUBLIC	0.3450	0.7567	5.3549	3.4256	2.6639	4.0431
				25.4804	91.9218	-0.5007	8.5509
Wealth				18%	64%	0%	8%
Students	EXAPPLM	0.1111	-0.2418	1.9364	-0.5576	0.2362	1.6552
	EXPUREM	-0.1384	0.1587	-0.2203	1.3971	2.9206	0.9846
	LHRS	3.5990	3.2207	-3.359	-34.8084	9.528	-44.012
	MHRS	3.8960	3.7878	4.424	14.1238	-3.6621	14.9222
Parents	HISEI	40.4196	26.6023	-	-	-6.3504	7.3598
	MISCED	3.1193	2.1744	-5.6001	18.8463	-2.143	-1.2957
	WEALTH	-1.4606	-2.1343	0.4117	-1.4204	2.4134	17.4408
	CULTPOS	-0.1424	-0.2361	0.3533	0.5941	0.3183	0.327
	HEDRES	-0.7427	-1.0743	-3.3855	-5.5085	-4.8774	-7.6374
	BOOK_N	53.6393	50.7860	0.0053	-0.4736	-0.0504	-5.2598
Teachers	TXT_BOOK	0.7905	0.7855	-	-	-5.6426	-5.2097
	CLSIZE	35.0130	42.5043	-8.6779	-23.4445	-0.5804	-46.7664
	TCFOCST	0.4975	0.1402	2.9242	0.3422	2.7312	0.8924
	TCMORALE	0.0376	-0.2941	-6.2933	-0.0748	-7.7804	1.8946
	TCHPARTI	-0.2169	-1.6445	11.3136	4.9119	6.7157	1.8929
Schools	TOWN	0.4508	0.3101	5.0475	-6.7474	-3.0845	-2.9438
	VILLAGE	0.1403	0.4584	-7.49	-0.6774	-11.2676	-1.7407
	PRIVATESCL	0.1714	0.0832	-0.0055	-0.0083	4.4454	-2.5786
	STU_FEES	25.7233	16.6104	0.671	-10.8914	4.5328	-23.7404
	RATCMP15	0.3909	0.2216	2.9895	-4.1902	7.6894	-12.933
	SCHAUTON	-0.2542	-1.0419	-4.4263	-8.8247	-10.9579	-0.3955
	EXC1_BAND	0.4710	0.1678	0.5318	-8.6246	0.29099	-2.4563
				-8.8496	-66.0364	-14.57471	-109.5998
				-6%	-46%	-14%	-103%
				16.6308	25.8854	-15.07541	-101.0489
	Total				**143.2834**		**106.4774**

Table 14: Add caption

		Indonesia		Jordan		Peru	
Ed wealth		*Endowments*	*Coefficients*	*Endowments*	*Coefficients*	*Endowments*	*Coefficients*
	INTERCEPT		181.3475		232.8262		247.8394
Students	PRESCHOOL	6.2309	6.0373	2.9304	8.8395	0.8096	6.8201
	LATESCHOOL	1.3545	0.1656	4.0339	-17.9123	5.3749	-6.8398
	NOREPEAT	3.0327	15.4797	-0.5863	-0.5305	5.1146	8.7662
	SHRS	3.6461	3.2498	-3.4578	-1.8296	2.56	19.0496
Parents	OUTMATH	9.9458	3.9547	8.4886	12.3205	3.6561	14.5438
	PARPRESSURE	0.2358	7.19	2.9216	8.3922	1.9524	4.8938
	TIGERMOM	-1.4622	-5.0425	-1.456	0.7928	-0.6088	3.4834
	TEACHMOM	2.1566	1.7254	2.0207	0.5374	1.4841	3.9414
Teachers	PROPCERT	1.0728	7.8387	-	-	-	-
	MATHPROFDEV	-0.0778	-4.2427	0.1404	0.0539	0.0423	-3.9528
	TCH_INCENTV	0.2442	0.6929	-0.3897	1.002	2.6139	-1.3882
	TCM_INSPE	-1.7713	-7.5635	2.0905	4.5083	-6.3926	-11.6551
	TCM_OBSER	-0.0742	11.6516	0.0738	-11.522	-1.9658	-37.8481
	COMP_USE	-0.6291	1.714	0.1317	-6.769	-0.3406	1.8251
	STU_FEEDB	0.0796	7.9599	-0.4144	-4.888	-0.6405	5.056
Schools	EXC6_MATHCOMP	-0.3133	1.5705	-0.8381	-2.5399	0.0019	-10.8913
	SCMATBUI	0.0048	-1.1337	0.2473	-3.7528	-0.0379	-0.7149
	SCL_EXTR_CL	-2.9043	-20.2148	-4.3968	-9.0748	-8.1854	-11.5939
	SCORE_PUBLIC	5.7895	1.2721	5.9804	4.7486	7.0514	2.4778
		26.5611	32.305	17.5202	-17.6237	12.4896	-14.0269
Wealth		19%	23%	15%	-15%	10%	-11%
Students	EXAPPLM	1.2618	-0.0988	2.5477	-1.9561	1.7418	-0.0448
	EXPUREM	2.1991	0.1956	2.1748	2.0725	-0.0688	-1.3801
	LHRS	-2.0941	-29.3342	15.7577	-58.088	14.0996	-56.6772
	MHRS	1.0459	14.6798	-0.1466	43.9865	-4.8692	13.7176
Parents	HISEI	-1.7446	-0.6415	-	-	-3.7414	3.6262
	MISCED	-0.4967	6.245	-5.2108	-3.2033	-1.2325	-10.9724
	WEALTH	-0.5319	-1.6968	1.5112	0.8037	1.4103	17.2629
	CULTPOS	-0.3684	0.7266	-0.1767	0.9315	0.4633	1.8325
	HEDRES	2.5025	-8.659	-5.509	-3.607	-6.0766	-3.1882
	BOOK_N	-0.062	1.7571	-0.2911	-0.1599	0.0089	-6.5238
Teachers	TXT_BOOK	1.2766	-17.5263	3.5046	-16.7213	0.629	-14.3821
	CLSIZE	-4.7921	-13.5878	-5.4853	-59.9741	-9.4773	-36.5348
	TCFOCST	4.5553	-4.0316	1.1106	-0.8656	1.4043	-0.7333
	TCMORALE	-9.9416	1.028	-0.4259	-1.4027	-1.9897	-0.2643
	TCHPARTI	18.5498	-10.6116	0.5336	8.9057	14.9263	-0.3443
Schools	TOWN	5.1411	-1.5651	2.5742	-2.1225	1.4527	-6.9018
	VILLAGE	-7.3432	-1.496	-10.417	-1.5033	-8.1092	-1.6368
	PRIVATESCL	2.0065	0.0116	6.1701	-12.8976	10.9349	-11.4833
	STU_FEES	11.4805	-29.6824	-	-	-	-
	RATCMP15	-1.56	-11.0681	3.8324	-25.5703	3.6968	-11.777
	SCHAUTON	-17.2045	8.6425	1.6804	5.5624	-10.3639	0.5217
	EXC1_BAND	0.4597	-7.9357	0.0348	-2.8344	-0.118	0.1559
		4.3397	-104.6487	13.7697	-128.6438	4.7213	-125.7274
		3%	-75%	12%	-109%	4%	-100%
		30.9008	-72.3437	31.2899	-146.2675	17.2109	-139.7543
	Total		**139.9046**		**117.8486**		**125.2960**

Table 15: Add caption

		Thailand		Tunisia		Shanghai	
Ed wealth		*Endowments*	*Coefficients*	*Endowments*	*Coefficients*	*Endowments*	*Coefficients*
	INTERCEPT		197.3024		220.3262		88.5547
Students	PRESCHOOL	-1.3537	-17.271	6.1004	8.8937	-0.8489	-3.5329
	LATESCHOOL	2.4103	-6.5245	5.1805	-13.6441	0.3962	9.7699
	NOREPEAT	-0.8545	18.0457	11.7885	-24.2972	1.6305	-17.898
	SHRS	-4.6096	-1.3158	6.821	12.3146	-1.772	-12.5189
Parents	OUTMATH	6.7928	12.3212	1.5262	12.4973	5.2761	12.9491
	PARPRESSURE	-1.4518	4.2812	6.05	-2.0319	4.4975	1.3579
	TIGERMOM	-0.388	-2.3087	-4.5979	-4.6769	0.72802	9.9719
	TEACHMOM	2.753	0.1281	3.4577	-0.343	2.2711	-1.1408
Teachers	PROPCERT	-0.4727	31.4282	1.103	-1.8785	-0.6693	-51.1404
	MATHPROFDEV	0.2289	0.5997	-0.0296	-5.3119	0.2539	-0.8222
	TCH_INCENTV	-0.3042	2.7973	1.7686	-0.8613	-0.3001	-0.1284
	TCM_INSPE	-6.9777	-9.7581	0.5072	-16.981	1.0351	-24.5551
	TCM_OBSER	-0.0414	35.1591	-0.8747	8.2309	0.0005	65.8477
	COMP_USE	0.0932	3.2715	-1.3139	2.006	-0.6498	1.8805
	STU_FEEDB	-0.0804	-3.4671	-1.645	-0.1837	0.1813	-16.2718
Schools	EXC6_MATHCOMP	-0.6327	-15.9607	-1.3721	-0.8139	-0.456	-38.0986
	SCMATBUI	0.4918	-0.3168	1.4125	-13.0502	-0.6578	0.199
	SCL_EXTR_CL	-0.299	-31.4213	-3.8009	-17.7227	-8.1668	-15.1872
	SCORE_PUBLIC	0.1084	-1.5941	7.2212	-0.7671	7.9348	-0.2396
		-4.5873	18.0939	39.3027	-58.6209	10.68432	-79.5579
Wealth		-6%	25%	35%	-52%	-12%	92%
Students	EXAPPLM	2.6597	-0.6224	-0.0268	0.7122	1.5828	0.0217
	EXPUREM	1.1331	-0.244	3.0066	0.953	1.0186	0.3889
	LHRS	-10.6968	-28.5366	23.1941	-63.0774	1.2012	-2.4455
	MHRS	0.4215	-18.4475	-3.834	24.3011	-3.1122	-8.4981
Parents	HISEI	-3.7308	3.9065	-4.5879	-17.802	-8.0828	-1.1654
	MISCED	-1.5344	-3.2852	-0.8438	9.0345	-3.3967	3.5561
	WEALTH	1.9975	4.6253	0.8843	5.3532	2.4335	-2.6004
	CULTPOS	0.2273	-0.0479	-0.5056	0.2824	0.9242	-2.4857
	HEDRES	-6.2788	-2.0394	-2.857	-5.5976	-9.2128	-0.1094
	BOOK_N	-0.3221	-4.9515	0.1597	0.8503	-0.7705	-8.6599
Teachers	TXT_BOOK	2.749	-10.9737	3.2614	-33.4627	2.5555	-0.762
	CLSIZE	-3.667	-20.3239	-8.5485	10.7592	-2.2003	-21.9057
	TCFOCST	4.7078	-9.8512	-2.34511	1.3381	1.2575	-2.1
	TCMORALE	-4.4468	1.0381	3.9636	1.4813	-3.4884	0.0449
	TCHPARTI	27.6836	-19.9891	3.4112	7.9172	8.2979	4.5458
Schools	TOWN	3.5728	-3.3756	6.7826	6.9162	-	-
	VILLAGE	-9.3596	-2.0706	-11.82	0.9214	-	-
	PRIVATESCL	0.1877	1.3994	-	-	0.19617	-5.3329
	STU_FEES	0.2682	-8.9641	-0.355	-19.7408	0.2292	-3.3761
	RATCMP15	7.5765	-21.4644	1.8275	-7.3367	7.9599	-8.3436
	SCHAUTON	-21.8484	10.7935	-2.7901	-17.2538	-11.3664	-1.2102
	EXC1_BAND	0.6927	3.9018	0.154	-3.1367	0.7272	-32.1963
		-8.0073	-129.5225	8.13119	-96.5876	-13.24643	-92.6338
		-11%	-177%	7%	-86%	15%	107%
		-12.5946	-111.4286	47.43389	-155.2085	-2.56211	-172.1917
	Total		**73.2792**		**112.5516**		**-86.1991**

4.1 Sensitivity to the choice of omitted baseline category

The results of Blinder-Oaxaca decompositions have been found to be sensitive to the researcher's choice of the omitted baseline category when categorical variables are included as covariates OaxacaRansom1999. Typically, categorical explanatory variables are introduced as a set of indicator ("dummy") variables on the right hand side. To avoid perfect multicollinearity, one of the dummy variables is usually omitted, and represents the baseline category. The coefficients on the remaining dummy variables are then interpreted as deviations from this omitted baseline. A linear regression model that contains a categorical explanatory variable may thus have the following general form:

4.2 Other resources for other uses

5 Conclusion

6 Resources

6.1 Blax

• Rtf2LaTeX2e - free software to convert from rtf (MS-Word) file to LaTeX. It saves a lot of work when converting existent papers written in Word like programs, but it is not perfect (tables, graphs, equations, and formats may not convert well).

• LaTeX.org - information and free programs for lots of uses

• Ctan.org - information and free programs for lots of uses (Boston College) Economics' resources - information and links for LaTeX typesetting (includes an introduction manual)

• Sourceforge - free open source LaTeX programs for lots of uses (look for LaTex in the software search)

6.2 Links for publication of economics reseach

Search for these useful links on the web (I'll add the links to this document later).

• JEL Classification Numbers

• How to publish in Economics by Prof. Kwan Choi (Editor, Review of International

Economics)

• http://econpapers.repec.org/ or http://www.ssrn.com/ - to share working papers 1.5=one and half, 2=double, 3=triple, etc.).

6.3 New line or paragraph

To start a new line **with indent** like for a new paragraph, skip one line in your .tex file.

6.4 New line or paragraph

To start a new line **with indent** like for a new paragraph, skip one line in your .tex file.

To start a new line **without indent** add \\ at the point where you want the new line to start.

6.5 Indent

To eliminated the indent in a given paragraph (useful when preparing presentation slides), start the paragraph with \noindent

To increase the indent, add a \quad or \hspace{**X**cm}, where **X** is the number of centimeters to skip (you can use in=inch too).

6.6 Margins

To change page layout margins, alter the parameters in

\geometry{left=**1.0**in,right=**1.0**in,top=**1.0**in,bottom=**1.0**in}

Instead of inches (in), you could use centimeters (cm). You must be using the geometry package, i.e., make sure the following is in the preamble of your .tex file:

\usepackage[nohead]{geometry}

6.7 Hyphenation

To avoid excessive hyphenation (i.e., word-breaks between lines), add the following to where you want the command to start having effect (usually before the beginning of your

text):

\sloppy

This command does not completely eliminate hyphenation, but makes it very rare. LaTeX was create to generate a nice looking output, so the compiler tries the best it can to avoid hyphenation, but sometimes it would create large spaces between words, so the compiler prefers to hyphenate the last word of the line.

6.8 Justification

Justification is generally not needed for working papers, but here it is. To have text justified to the left, use \flushright at the point you want justification to start. To have text justified to the right, use \flushleft at the point you want justification to start. To have text centered, use

\begin{center}

Text that you want to be centered

\end{center}

to call the label by writing

Smith \ref{**labelforSmith**}. For references with year, like Smith (1996), using labels is not that useful because it is faster to just write the year yourself. But if you want to get the year automatically using the label, write Smith\cite{**labelforSmith**} or, if the reference is already within parentheses, write (Smith, \citeyear{**labelforSmith**}).

6.9 Figures and pictures

and then to crop it, you can try to use an eps version of the figure. I won't explain this here because I think it's too much work and confusing. Add the figure (where you want it to be) with:

\begin{figure}[**htbp**]

\caption{**Title**}

\centering \includegraphics[width=**0.75**\textwidth]{**filename.pdf**} \\

A note you want to add here (like the source of the data for a graph).

\label{**your_key**}

\end{figure}

where **htbp** is for the location on the page: here, top of the page, bottom, of floating in an exclusive page, **Title** is the title that appears at the top of the figure (automatically precedes with "Figure X:", where X is the number of the figure), **0.75**\textwidth gives the width as a proportion of the text width (you can use a measure in inches or cm instead), **filename.pdf** is the name of the file of the figure, which should be in the same folder of your .tex file, and **your_key** is the key that you can use to refer to the figure in the text (you have to write \ref{**your_key**} in order to have the reference (the number of the figure) shown in the text). Notice that you can add a note at the bottom of the figure for sources or other remarks. The example above should give something like the following figure (using the option "h", i.e., print it here).

T

7 Concluding remarks

Good luck! Yes, luck is helpful during this learning process (avoiding silly mistakes will save you a lot of time).

This is an open-source document. Feel free to write and distribute your own improved version based on this one (just don't forget to cite this document). The original .tex file of this document is available at http://faculty.gvsu.edu/ogural/

Future topics to be covered here include how to use Bibtex ...

Table 16: Summary statistics - Additional Variables used for regressions

		Dev7 countries		Vietnam	
Variable	Description	MS	Valid N	MS	Valid N
TCSHORT	Shortage of teaching staff	0.4742 (1.2601)	43144	0.418 (1.1628)	4959
TCFOCST	Teacher focus	0.4932 (1.0049)	43422	0.1321 (0.8347)	4959
Quality assurance of mathematics teachers through ...					
TCM_STUASS	test or assessment of student achievement	0.8734 (0.3325)	43048	0.9821 (0.1328)	4959
Assessment used to					
ASS_PROG	inform parents about childs progress	0.9669 (0.179)	42703	0.9929 (0.0837)	4959
ASS_PROM	decide on students? retention or promotion	0.8998 (0.3002)	42478	0.9516 (0.2146)	4959
ASS_NAT	compare school to national performance	0.6951 (0.4604)	42450	0.8804 (0.3245)	4959
ASS_CUR	identify improvements in the curriculum	0.8978 (0.3029)	42475	0.9141 (0.2803)	4959
School policy related factors					
EXC11_UNICORN	School offers 'country specific item'	0.7108 (0.4534)	41907	0.9635 (0.1875)	4959
LEADINST	Promotion of instructional leadership	0.0732 (1.0797)	43253	-0.0465 (0.9424)	4959
QUAL_RECORD	Systematic recording of data for quality assurance	0.8824 (0.3221)	42939	0.9821 (0.1328)	4959
SCHSEL	School selectivity/ student admission policies	2.3036 (0.7997)	43296	2.8411 (0.4074)	4959

References

Leslie Lamport, *LaTeX: a document preparation system* Addison Wesley, Massachusetts, 2nd edition, 1994.

OECD *PISA 2012 Results: Excellence Through Equity: Giving Every Student the Chance to Succeed (Volume II)*, 2013a, PISA, OECD Publishing. OECD PISA 2012 Results Volume II

Roland G. Fryer Jr. and Steven D.Levitt, Understanding the Black-White Test Score Gap in the First Two Years of School *The Review of Economics and Statistics* May 2004, 86 (2):447-464

Andreas Ammermueller PISA: What makes the difference? Explaining the gap in test scores between Finland and Germany *Empirical Economics* (2007) 33:263287

Alan S. Blinder Wage Discrimination: Reduced Form and Structural Estimates *Journal of Human Resources*, 1973, 8(4):436-455.

Ronald L. Oaxaca Male-Female Wage Differentials in Urban Labor Markets *International Economic Review*, 1973, 14(3):693-709.

W.Steven Barnett Long-Term Outcomes of Early Childhood Programs *The Future of Children*, 1995, 5(3):25-50

Lawrence J. Schweinhart, Jeanne Montie, Zongping Xiang, W. Steven Barnett, Clive R. Belfield, and Milagros Nores Lifetime Effects: The High/Scope Perry Preschool Study Through Age 40, *Monographs of the High/Scope Educational Research Foundation, 14*, 2005, Ypsilanti, MI: High/Scope Press

Russell J.Dalton and Nhu-Ngoc T.Ong Authority Orientations and Democratic Attitudes: A Test of the Asian Values Hypothesis *Japanese Journal of Political Science* 2005, 6(2):1-21

OECD *PISA 2012 Technical Report*, 2014, PISA, OECD Publishing OECD PISA 2012 Technical Report

Tran Thu Ha and Trudy Harpham Primary education in Vietnam: Extra classes and outcomes *International Education Journal*, 2005, 6(5), 626-634.

Hai-Anh Dang The determinants and impact of private tutoring classes in Vietnam *Economics of Education Review* 2007 26:684699

Tuan Anh Le *Applying realistic mathematics education in Vietnam : teaching middle school geometry*, Doctoral Dissertation, 2007, Universität Potsdam Doctoral dissertation Tuan Le 2007

Danh Nam Nguyen and Trung Tran Recommendations for Mathematics Curriculum Development in Vietnam, *Proceedings of the 6th International Conference on Educational Reform (ICER 2013): ASEAN Education in the 21st Century* Nguyen and Tran 2013 Conference Paper

Amy Chua *Battle Hymn of the Tiger Mother* 2011 Penguin Press, New York

Amy Hsin and Yu Xie Explaining Asian Americans academic advantage over whites *Proceedings of the National Academy of Sciences*, May 2014 111 (23) 8416-8421; Hsin and Xie, PNAS 2014

Vu Dinh Phuong *Using Video Study to Investigate Eighth-grade Mathematics Classrooms in Vietnam*, Doctoral Dissertation, 2014, Universität Potsdam Doctoral Dissertation Vu Dinh Phuong 2014