The Acquisition of Math and Reading Skills in Developing Countries: What Explains Differences between *Kinh* and Ethnic Minority Students in Vietnam?

Paul Glewwe
Qihui Chen
Bhagyashree Katare

Department of Applied Economics
University of Minnesota

Incomplete first draft
September, 2009

I. Overview of Education in Developing Countries

The main stylized facts are:

- Enrollment has increased, but some regions (South Asia and Sub-Saharan Africa) still lag behind.

- Gross enrollment rates can exaggerate how many children complete a given level (primary or secondary) of schooling, since they do not account for repetition. They also overlook delayed enrollment which, with repetition, leads to overage enrollment.

- Students in developing countries often perform very poorly on standardized tests, despite the fact that they may be a “select” group.

For a nice overview, see Hanushek & Woessman, “The Role of Cognitive Skills in Economic Development.” *Journal. of Econ. Lit.* Sept., 2008.

Table 1. Primary School Gross Enrollment Rates

Area	***1960***	***1970***	***1980***	***1990***	***2000***
World	80	87	97	102	104
Country group					
Low-income	65	77	94	102	102
Middle-income	83	103	101	103	110
High-income	109	100	101	102	102
Region					
Sub-Saharan Africa	40	51	80	74	77
Middle East/North Africa	59	79	89	96	97
Latin America	91	107	105	106	127
South Asia	41	71	77	90	98
East Asia	87	90	111	120	111
East Europe/FSU	103	104	100	98	100
OECD	109	100	102	103	102

Table 2. Primary School Enrollment, Repetition, and Grade 4 Survival Rates (percents), in 2000

Area	*Gross enrollment*	*Net enrollment*	*Repetition*	*On-time enrollment*	*Grade 4 survival*
Country group					
Low-income	102	85	4	55	80
Middle-income	110	88	10	61	88
High-income	102	95	2[a]	73[b]	98[b]
Region					
Sub-Saharan Africa	77	56	13	30	76
Mid. East/N. Africa	97	84	8	64	96
Latin America	127	97	12	74	86
South Asia	98	83	5	-	55
East Asia	111	93	2	56	97
East Europe/FSU	100	88	1	67[a]	97[b]
OECD	102	97	2[a]	91[a]	99[b]

Table 3. Secondary School Gross Enrollment Rates

Area	*1960*	*1970*	*1980*	*1990*	*2000*
World	29	36	49	55	67
Country group					
Low-income	14	21	34	41	54
Middle-income	21	33	51	59	77
High-income	63	74	87	92	101
Region					
Sub-Saharan Africa	5	6	15	23	27
Middle East/North Africa	13	25	42	56	66
Latin America	14	28	42	49	86
South Asia	10	23	27	39	47
East Asia	20	24	44	48	67
East Europe/FSU	55	64	93	90	88
OECD	65	77	87	95	107

Table 4. Mean Mathematics and Reading Achievement, TIMSS and PIRLS Studies

	1999 Mathematics (TIMSS)		*2001 Reading (PIRLS)*
Country	*Grade 7*	*Grade 8*	*Grade 4*
U.S.	-	502	542
Argentina	-	-	420
Belize	-	-	327
Chile	-	392	-
Colombia	-	-	422
Indonesia	-	403	-
Iran	-	422	414
Jordan	-	428	-
Korea (South)	-	587	-
Kuwait	-	-	396
Malaysia	-	519	-
Morocco	337	-	350
Philippines	345	-	-
South Africa	-	275	-
Thailand	-	467	-
Turkey	-	429	449

Table 5. Math and Reading Achievement of 15 Year Olds, PISA Study

	Mathematics	*Reading*	
Country	Mean score	Mean score	Percent with very low skills
France	517	505	4.2
Japan	557	522	2.7
United Kingdom	529	523	3.6
United States	493	504	6.4
Argentina[a]	388	418	22.6
Brazil	334	396	23.3
Chile[a]	384	410	19.9
Indonesia[a]	367	371	31.1
Mexico	387	422	16.1
Peru[a]	292	327	54.1
South Korea	547	525	0.9
Thailand[a]	432	431	10.4

II. Analysis of Determinants of Learning of *Kinh* and Ethnic Minority Students in Vietnam

A. Data

From the “Young Lives” Panel Survey conducted in Vietnam (add website address here!)

- 2000 children age 1 in 2002 (Round 1) and age 5 in 2006 (Round 2)
- 1000 children age 8 in 2002 (Round 1) and age 12 in 2006 (Round 2)
- Not a random sample of the Vietnamese population, but roughly representative of the country as a whole
- Extremely detailed health and education data, including test scores

B. Methodology (Oaxaca-Blinder Decomposition)

The objective is to estimate a "learning production function", which can be depicted as:

$$A = a(S, \mathbf{Q}, \mathbf{C}, \mathbf{H}, \mathbf{I}) \qquad (1)$$

A is skills learned ("achievement")

S is years of schooling

Q is all school and teacher characteristics ("quality") that affect learning

C is all child characteristics (including "ability") that affect learning

H is all household characteristics that affect learning

I is educational "inputs" from households (children's daily attendance, textbooks and other school supplies, etc.)

A simple linear specification of (1) is:

$$A = \beta_0 + \beta_1 S + \beta_{Q1} Q_1 + \beta_{Q2} Q_2 + \ldots + \beta_{C1} C_1 + \beta_{C2} C_2 + \ldots \quad (1')$$

$$+ \beta_{H1} H_1 + \beta_{H2} H_2 + \ldots + \beta_{I1} I_1 + \beta_{I2} I_2 + \ldots + u_A$$

Assuming linearity is not restrictive if one adds squared and interaction terms to the variables in (1).

For the Blinder-Oaxaca decomposition, consider estimates of equation (1′) separately for the *Kinh* and ethnic minority populations:

$$A_k = \beta_{0k} + \boldsymbol{\beta}_k' \mathbf{x}_k + u_{Ak} \qquad (5) \quad (\textit{Kinh})$$

$$A_m = \beta_{0m} + \boldsymbol{\beta}_m' \mathbf{x}_m + u_{Am} \qquad (6) \quad (\text{ethnic minority})$$

Averaging these 2 relationships for their respective populations gives:

$$\overline{A}_k = \beta_{0k} + \boldsymbol{\beta}_k'\overline{\mathbf{x}}_k \qquad (5')$$

$$\overline{A}_m = \beta_{0m} + \boldsymbol{\beta}_m'\overline{\mathbf{x}}_m \qquad (6')$$

The difference in the mean test scores between *Kinh* children and ethnic minority children can be expressed as:

$$\overline{A}_k - \overline{A}_m = (\beta_{0k} - \beta_{0m}) + (\boldsymbol{\beta}_k'\overline{\mathbf{x}}_k - \boldsymbol{\beta}_m'\overline{\mathbf{x}}_m) \qquad (7)$$

Blinder and Oaxaca both showed how the difference in the terms in the second set of parentheses can be decomposed into two parts:

$$\overline{A}_k - \overline{A}_m = (\beta_{0k} - \beta_{0m}) + (\boldsymbol{\beta}_k'\overline{\mathbf{x}}_k - \boldsymbol{\beta}_m'\overline{\mathbf{x}}_m) + \boldsymbol{\beta}_k'\overline{\mathbf{x}}_m - \boldsymbol{\beta}_k'\overline{\mathbf{x}}_m \qquad (8)$$

$$= (\beta_{0k} - \beta_{0m}) + \boldsymbol{\beta}_k'(\overline{\mathbf{x}}_k - \overline{\mathbf{x}}_m) + (\boldsymbol{\beta}_k - \boldsymbol{\beta}_m)'\overline{\mathbf{x}}_m$$

The first part, $\boldsymbol{\beta}_k'(\overline{\mathbf{x}}_k - \overline{\mathbf{x}}_m)$, reflects the difference in the mean values of the $\mathbf{x}$ variables across the two ethnic groups (which is multiplied by $\boldsymbol{\beta}_k$).

The second part, $(\boldsymbol{\beta}_k - \boldsymbol{\beta}_m)'\overline{\mathbf{x}}_m$, reflects the difference in the coefficients across the two ethnic groups (which is multiplied by $\overline{\mathbf{x}}_m$).

There is also the “unexplained” component, $(\beta_{0k} - \beta_{0m})$, which is a “fixed” disadvantage (or perhaps advantage) for ethnic minority groups.

In fact, this decomposition can be done in another, analogous, way, which multiplies the difference in the means across the two groups by $\boldsymbol{\beta}_m$ and multiplies the differences in the $\boldsymbol{\beta}$'s of the two groups by $\bar{\mathbf{x}}_k$:

$$\bar{A}_k - \bar{A}_m = (\beta_{0k} - \beta_{0m}) + (\boldsymbol{\beta}_k'\bar{\mathbf{x}}_k - \boldsymbol{\beta}_m'\bar{\mathbf{x}}_m) + \boldsymbol{\beta}_m'\bar{\mathbf{x}}_k - \boldsymbol{\beta}_m'\bar{\mathbf{x}}_k \qquad (9)$$

$$= (\beta_{0k} - \beta_{0m}) + \boldsymbol{\beta}_m'(\bar{\mathbf{x}}_k - \bar{\mathbf{x}}_m) + (\boldsymbol{\beta}_k - \boldsymbol{\beta}_m)'\bar{\mathbf{x}}_k$$

Ideally, these two different ways to decompose the difference in mean test scores of *Kinh* and ethnic minority students in Vietnam will give similar results, but this is not guaranteed.

The results presented today do not use the school quality data (this will be done soon!). To avoid omitted variable bias community fixed effects are used to control for differences in school quality. (The evidence suggests that kids in the same commune usually attend the same school.)

III. Results for Younger Cohort (5 years old when tested)

Some notes on the younger cohort:

- Very few have started school, though many have been to preschool

- Math test: CDA test of basic quantitative skills (designed by the International Evaluation Association). There are 15 questions, but one question was dropped because it was not correlated with the average of the other questions.

- Reading test: Peabody Picture Vocabulary Test (PPVT).

- Ethnic minority children had the option of taking the tests in Vietnamese or in their native language.

Table 1: Mean Test Scores for Ethnic Majority & Ethnic Minority Children

(Younger Cohort, 5 years old)

Student Type	Variable	Mean	Standard Dev.	Observations
All Communes:				
Full Sample	CDA-Q score	9.79	2.51	1906
	PPVT score	36.97	18.18	1747
Kinh	CDA-Q score	10.20	2.29	1631
	PPVT score	39.40	18.03	1480
Ethnic Minority	CDA-Q score	7.36	2.34	275
	PPVT score	23.52	12.15	267
Mixed Communes:				
Full Sample	CDA-Q score	8.99	2.40	445
	PPVT score	32.12	14.64	428
Kinh	CDA-Q score	10.03	2.05	230
	PPVT score	38.03	14.28	221
Ethnic Minority	CDA-Q score	7.88	2.26	215
	PPVT score	25.81	12.20	207

Table 2: Regression Estimates for CDA-Q Test, Younger Cohort

Variables	$\boldsymbol{\beta}_k$	$\boldsymbol{\beta}_m$	$\boldsymbol{\beta}_k$-$\boldsymbol{\beta}_m$	$\bar{\mathbf{x}}_k$	$\bar{\mathbf{x}}_m$	$(\boldsymbol{\beta}_k-\boldsymbol{\beta}_m)'\bar{\mathbf{x}}_k$	$\boldsymbol{\beta}_m'(\bar{\mathbf{x}}_k-\bar{\mathbf{x}}_m)$	$(\boldsymbol{\beta}_k-\boldsymbol{\beta}_m)'\bar{\mathbf{x}}_m$	$\boldsymbol{\beta}_k'(\bar{\mathbf{x}}_k-\bar{\mathbf{x}}_m)$	$\boldsymbol{\beta}_k$ (=$\boldsymbol{\beta}_m$)	$\boldsymbol{\beta}_k'(\bar{\mathbf{x}}_k-\bar{\mathbf{x}}_m$
Lpcexp	0.054	1.375***	-1.321***	1.943	1.135	-2.567	1.111	-1.499	0.044	--	--
Daded	0.021**	same	0.0	8.37	3.24	0.0	0.108	0.0	0.108	0.027***	0.138
Mumed	0.030***	same	0.0	7.72	2.11	0.0	0.168	0.0	0.168	0.029***	0.163
Girl	0.012	same	0.0	0.49	0.463	0.0	0.000	0.0	0.000	0.014	0.000
Agechild	0.035***	same	0.0	15.28	13.71	0.0	0.055	0.0	0.055	0.034***	0.054
Zhaz	0.002	same	0.0	3.977	2.863	0.0	0.003	0.0	0.003	0.017	0.019
Lnedxkid	0.009	same	0.0	5.541	2.501	0.0	0.027	0.0	0.027	0.061	0.185
Crechtim	0.000	same	0.0	6.739	0.555	0.0	0.001	0.0	0.001	0.000	0.001
Presctim	0.004	same	0.0	17.62	11.48	0.0	0.023	0.0	0.023	0.004	0.021
Avg. cons. (segreg.)	-1.057	-3.214									
Avg. cons. (mixed)	-0.943	-3.228									

Table 3: Regression Estimates for PPVT Test, Younger Cohort

Variables	$\boldsymbol{\beta}_k$	$\boldsymbol{\beta}_m$	$\boldsymbol{\beta}_k$-$\boldsymbol{\beta}_m$	$\bar{\mathbf{x}}_k$	$\bar{\mathbf{x}}_m$	$(\boldsymbol{\beta}_k-\boldsymbol{\beta}_m)'\bar{\mathbf{x}}_k$	$\boldsymbol{\beta}_m'(\bar{\mathbf{x}}_k-\bar{\mathbf{x}}_m)$	$(\boldsymbol{\beta}_k-\boldsymbol{\beta}_m)'\bar{\mathbf{x}}_m$	$\boldsymbol{\beta}_k'(\bar{\mathbf{x}}_k-\bar{\mathbf{x}}_m)$	$\boldsymbol{\beta}_k$ [$\boldsymbol{\beta}_m$]	$(\boldsymbol{\beta}_k-\boldsymbol{\beta}_m)'\bar{\mathbf{x}}_k$ [$(\boldsymbol{\beta}_k-\boldsymbol{\beta}_m)'\bar{\mathbf{x}}_m$]	$\boldsymbol{\beta}_m'(\bar{\mathbf{x}}_k-\bar{\mathbf{x}}_m)$ [$\boldsymbol{\beta}_k'(\bar{\mathbf{x}}_k-\bar{\mathbf{x}}_m)$]
Lpcexp	0.338***	0.921***	-0.583**	1.943	1.135	-1.133	0.774	-0.662	0.273	--	--	--
Daded	0.019**	Same	0.0	8.37	3.24	0.0	0.097	0.0	0.097	0.029***	0.0	0.149
Mumed	0.033***	Same	0.0	7.72	2.11	0.0	0.185	0.0	0.185	0.035***	0.0	0.196
Girl	-0.041	Same	0.0	0.49	0.463	0.0	-0.001	0.0	-0.001	-0.027	0.0	-0.001
Agechild	0.054***	0.019	0.035***	15.28	13.71	0.535	0.030	0.480	0.085	0.052*** [0.015]	0.565 [0.507]	0.030 [0.082]
Zhaz	0.050	Same	0.0	3.977	2.863	0.0	0.056	0.0	0.056	0.017	0.0	0.019
Lnedxkid	0.008	Same	0.0	5.541	2.501	0.0	0.024	0.0	0.024	0.061	0.0	0.185
Crechtim	-0.002	Same	0.0	6.739	0.555	0.0	0.011	0.0	0.011	0.000	0.0	-0.001
Presctim	0.004	Same	0.0	17.62	11.48	0.0	0.023	0.0	0.023	0.004	0.0	-0.004
Av. cons. (segreg.)	-2.084	-2.451										
Av. cons. (mixed)	-1.902	-2.347										

Table 4: Mean Test Scores for Ethnic Majority and Ethnic Minority Children
(Older Cohort, 12 years old)

Student Type	Variable	Mean	Standard Dev.	Observations
All Communes:				
Full Sample	Math (IEA) score	7.44	1.92	981
	PPVT score	137.6	26.1	945
Kinh	Math (IEA) score	7.75	1.51	855
	PPVT score	142.3	18.8	827
Ethnic Minority	Math (IEA) score	5.28	2.78	126
	PPVT score	104.3	41.5	118
Mixed Communes:				
Full Sample	Math (IEA) score	6.62	2.32	217
	PPVT score	130.4	29.1	206
Kinh	Math (IEA) score	7.44	1.58	118
	PPVT score	141.8	18.6	113
Ethnic Minority	Math (IEA) score	5.64	2.66	99
	PPVT score	116.6	33.3	93

Table 5: Regression Estimates for Mathematics (IEA) Test, Older Cohort										
Variables	$\boldsymbol{\beta}_k$	$\boldsymbol{\beta}_m$	$\boldsymbol{\beta}_k$-$\boldsymbol{\beta}_m$	$\bar{\mathbf{x}}_k$	$\bar{\mathbf{x}}_m$	$(\bar{\mathbf{x}}_k$-$\bar{\mathbf{x}}_m)$	$(\boldsymbol{\beta}_k$-$\boldsymbol{\beta}_m)'\bar{\mathbf{x}}_k$	$\boldsymbol{\beta}_m'(\bar{\mathbf{x}}_k$-$\bar{\mathbf{x}}_m)$	$(\boldsymbol{\beta}_k$-$\boldsymbol{\beta}_m)'\bar{\mathbf{x}}_m$	$\boldsymbol{\beta}_k'(\bar{\mathbf{x}}_k$-$\bar{\mathbf{x}}_m)$
Lpcexp	0.264**	same	0.0	2.085	1.384	0.701	0.0	0.185	0.0	0.185
Daded	0.025**	same	0.0	8.515	2.902	5.613	0.0	0.140	0.0	0.140
Mumed	0.024***	same	0.0	7.651	1.619	6.032	0.0	0.145	0.0	0.145
lnedxkid	0.016	same	0.0	6.027	2.905	3.122	0.0	0.050	0.0	0.050
Girl	-0.011	0.287*	-0.298*	0.502	0.503	-0.001	-0.150	-0.001	-0.150	0.000
agechild	0.010	same	0.0	15.163	13.669	1.494	0.0	0.015	0.0	0.015
yrs_sch	0.234***	0.368**	-0.134**	5.954	5.133	0.821	-0.799	0.302	-0.688	0.192
hrs_sch	0.140***	same	0.0	4.504	4.000	0.504	0.0	0.071	0.0	0.071
hrs_stud	0.010	same	0.0	2.901	1.579	1.322	0.0	0.013	0.0	0.013
hrs_work	-0.049*	same	0.0	1.826	3.495	-1.669	0.0	0.082	0.0	0.082
exclsmth	0.004	same	0.0	1.913	0.291	1.622	0.0	0.006	0.0	0.006
Haz	0.065**	same	0.0	3.728	2.721	1.007	0.0	0.065	0.0	0.065
hearprob	-0.023	same	0.0	0.208	0.007	0.201	0.0	-0.005	0.0	-0.005
undrstpr	-0.661***	same	0.0	0.015	0.031	-0.016	0.0	0.011	0.0	0.011
lnghlth8	-0.057	same	0.0	0.063	0.086	-0.023	0.0	0.001	0.0	0.001
mightdie12	-0.258**	same	0.0	0.056	0.055	0.001	0.0	-0.000	0.0	-0.000
Avg. const. (segregated)	-3.205	-4.632								
Avg. const.	-3.339	-3.499								

(mixed)

Table 6: Regression Estimates for PPVT Test, Older Cohort										
Variables	$\boldsymbol{\beta}_k$	$\boldsymbol{\beta}_m$	$\boldsymbol{\beta}_k$-$\boldsymbol{\beta}_m$	$\bar{\mathbf{x}}_k$	$\bar{\mathbf{x}}_m$	$(\bar{\mathbf{x}}_k-\bar{\mathbf{x}}_m)$	$(\boldsymbol{\beta}_k-\boldsymbol{\beta}_m)'\bar{\mathbf{x}}_k$	$\boldsymbol{\beta}_m'(\bar{\mathbf{x}}_k-\bar{\mathbf{x}}_m)$	$(\boldsymbol{\beta}_k-\boldsymbol{\beta}_m)'\bar{\mathbf{x}}_m$	$\boldsymbol{\beta}_k'(\bar{\mathbf{x}}_k-\bar{\mathbf{x}}_m)$
Lpcexp	0.396***	same	0.0	2.085	1.384	0.701	0.0	0.278	0.0	0.278
Daded	0.024***	same	0.0	8.515	2.902	5.613	0.0	0.135	0.0	0.135
Mumed	0.007	0.102**	-0.095**	7.651	1.619	6.032	-0.727	0.615	-0.154	0.042
lnedxkid	-0.029	same	0.0	6.027	2.905	3.122	0.0	-0.091	0.0	-0.091
Girl	-0.081	same	0.0	0.502	0.503	-0.001	0.0	0.000	0.0	0.000
agechild	0.025***	same	0.0	15.163	13.669	1.494	0.0	0.037	0.0	0.037
yrs_sch	0.319***	same	0.0	5.954	5.133	0.821	0.0	0.262	0.0	0.262
hrs_sch	0.036	same	0.0	4.504	4.000	0.504	0.0	0.018	0.0	0.018
hrs_stud	0.001	same	0.0	2.901	1.579	1.322	0.0	0.001	0.0	0.001
hrs_work	-0.012	same	0.0	1.826	3.495	-1.669	0.0	0.020	0.0	0.020
exclsmth	0.022*	same	0.0	1.913	0.291	1.622	0.0	0.036	0.0	0.036
Haz	0.037	same	0.0	3.728	2.721	1.007	0.0	0.037	0.0	0.037
hearprob	-0.614***	same	0.0	0.208	0.007	0.201	0.0	-0.123	0.0	-0.123
undrstpr	-0.188	same	0.0	0.015	0.031	-0.016	0.0	0.003	0.0	0.003
lnghlth8	-0.153*	same	0.0	0.063	0.086	-0.023	0.0	0.004	0.0	0.004
mightdie12	-0.110	same	0.0	0.056	0.055	0.001	0.0	-0.000	0.0	-0.000
Avg. const. (segregated)	-3.263	-5.399								
Avg. const.	-3.345	-2.981								