Education Economics, Vol. 12, No. 3, December 2004

Why Does Academic Achievement Vary Across Countries? Evidence from Cuba and Mexico

PATRICK J. MCEWAN & JEFFERY H. MARSHALL

ABSTRACT *International assessments of academic achievement are common. They are usually accompanied by attempts to infer the determinants of cross-country achievement gaps, but these inferences have little empirical foundation. This paper applies the Blinder–Oaxaca decomposition to the problem of explaining why primary students in Cuban schools score than Mexican students, on average, 1.3 standard deviations higher. The results suggest that no more than 30% of the difference can be explained by differing endowments of family, peer, and school variables. Of these, peer-group variables and, to a lesser extent, family variables explain the largest portion of the gap.*

Introduction

International assessments of academic achievement are used for many purposes; not least is the comparison of achievement across countries. With large inter-country differences in average test scores—and results that are widely disseminated in the press—debates about why some countries 'lag behind' others are inevitable. This has been particularly evident in the United States, which has consistently scored lower than Japan in mathematics. The possible causes of this test score gap have stimulated considerable debate.[1]

But what is missing from almost all discussions of inter-country differences in academic achievement is a rigorous framework that allows for multiple explanations. The preferred approach has focused on a single variable, such as the length of the school year, teacher education and training levels, or curriculum implementation. However, these 'explanations' ignore the multivariate process that determines outcomes such as test scores, and assume that the samples of students and schools in the countries in question are identical in all regards except one. Accordingly, this paper details a method for evaluating international achievement differentials that explicitly recognizes the existence of multiple sources of variation, and applies it to recently collected assessment data from Latin America.

In 1997, the Santiago office of UNESCO and 13 Ministries of Education applied language and mathematics tests to representative samples of third-grade and fourth-

P. J. McEwan, Department of Economics, 106 Central Street, Wellesley, MA 02481, USA. E-mail: pmcewan@stanfordalumni.org.
J. H. Marshall, School of Education, Stanford, CA, USA.

ISSN 0964-5292 Print; 1469-5782 Online/04/020205-13
DOI: 10.1080/0964529042000258572

grade students.[2] They further applied background surveys to students, parents, teachers, and principals. This paper uses data from two countries: Cuba and Mexico. Primary students in Cuba have notably higher academic achievement, on average, than students in Mexico (in fact, Cuba scores higher than every country in the sample) (Willms & Somers, 2001). This paper inquires as to the source of this advantage.

Can it be traced to differences in family characteristics between countries, to differences in peer groups, to differences in schools, or to unobserved factors? To answer this question, the paper applies the Blinder–Oaxaca decomposition (Blinder, 1973; Oaxaca, 1973). The method is typically used to assess the determinants of mean wage differences between two populations of workers (e.g. males and females). In this context, the technique is used to decompose the mean achievement gap between countries into the components just described.

The paper proceeds as follows. The next section reviews the literature on cross-national comparisons of academic achievement, and justifies the need for a more comprehensive empirical framework to explain cross-national differences. The third section describes the decomposition method that is employed, while the subsequent section, provides background on the data. The fifth section estimates the determinants of Spanish and mathematics achievement, followed by a decomposition exercise that accounts for the sources of variation in academic achievement. Finally, the results are summarized and conclusions presented.

Comparing achievement across countries

In recent years, the International Association for the Evaluation of Educational Achievement (IEA) has undertaken three international assessments of mathematics and science instruction.[3] Other large attempts at data collection, such as UNESCO's recent survey, are directly inspired by these efforts. Much of the attention given to the IEA results– especially in the popular press, but also among researchers– -has focused on comparisons between countries.

Most attempts to explain these cross-country differences have focused on a single explanation or a broad category of explanations. These include the length of the school year and hours of instruction (Finn, 1991; National Education Commission on Time and Learning, 1994), cultural and institutional factors (Hess & Azuma, 1991; Stigler & Stevenson, 1991), and curriculum (McKnight *et al.*, 1987; Westbury, 1992, 1993).[4] The issue of curriculum has received considerable attention in recent years, in part because one can compellingly argue that curriculum affects outcomes on international testing, but also because this 'solution' to low achievement is appealing to education scholars and policymakers (Stedman, 1994).

Even when researchers use approaches that allow for multiple explanations, their models do not always take full advantage of the data. Schaub and Baker (1991) combine data from the United States and Japan from the SIMS data collection to explore the effects of teaching methodology on mathematics achievement. They find that indicators of classroom management and instructional methods are significant predictors of average achievement in the two-country sample of 430 middle-school classrooms, and that these measures perform better than simple measures of school and teacher resources. However, by restricting the analysis to classroom averages and a very limited set of covariates, it is difficult to fully assess the extent to which the Japanese advantage in mathematics is attributable to these qualitative classroom practices. No effort is made to identify intervening variables that are related either to the school populations or the schools (or teachers) themselves.

Using the TIMMS data, Schmidt *et al.* (2001) explore the determinants of cross-national variation in achievement. As a dependent variable, they use national average gain scores.[5] They regress these gain scores on national averages for teacher responses about curriculum coverage and additional curriculum measures taken from textbooks. They find that their curriculum measures are generally significant predictors of academic achievement in a sample of 31 countries. However, the use of national averages and the almost total exclusion of control variables limit the utility of the findings, since we know nothing about how international variation in parental education, class size, and availability of learning materials (among many other factors) explain these differences.[6]

Empirical framework

Stedman (1994) and Jaeger (1992) have argued for a "richer formulation" (Stedman, 1994, p. 30) of cross-national analyses of academic achievement differentials that more accurately captures the numerous factors that explain these observed differences. This section describes an empirical framework for assessing the determinants of the achievement gap between two countries. It allows one to determine what portion of the gap—if any—is attributable to different endowments of family, peer, and school variables in each country.

In order to decompose the achievement gap into these components, one can employ a common technique from labor economics. The Blinder–Oaxaca decomposition was originally developed to study male–female and black–white wage discrimination in the United States, and it has been extended to many countries.[7] The technique can be usefully extended to the present context of achievement comparisons.[8] We begin with a basic education production function:

$$A_{ij} = \mathbf{X}_{ij}\beta_j + \varepsilon_{ij} \tag{1}$$

where the achievement (A_{ij}) of student i in country j is determined by a vector of family, peer, and school variables ($\boldsymbol{X}_{ij}$). An error term (ε_{ij}) captures unmeasured variables.

In the following discussion, we focus on a high-achieving country (indicated by the subscript H) and a low-achieving country (indicated by the subscript L). Equation (1) is separately estimated in each country by ordinary least squares, yielding a set of estimated regression coefficients. Since regression lines pass through the means of variables, it is true that:[9]

$$\bar{A}_H = \bar{X}_H\hat{\beta}_H \tag{2}$$

$$\bar{A}_L = \bar{X}_L\hat{\beta}_L \tag{3}$$

where $\bar{A}$ and $\bar{X}$ are means for high-achieving and low-achieving countries, and $\hat{\beta}$ values are coefficient estimates from ordinary least squares regressions. To explore the achievement gap between high-achieving and low-achieving countries, one can subtract equation (3) from equation (2):

$$\bar{A}_H - \bar{A}_L = \bar{X}_H\hat{\beta}_H - \bar{X}_L\hat{\beta}_L \tag{4}$$

Equation (4) can be re-written in two ways:

$$\bar{A}_H - \bar{A}_L = (\bar{X}_H - \bar{X}_L)\hat{\beta}_H + \bar{X}_L(\hat{\beta}_H - \hat{\beta}_L) \quad (5)$$

or

$$\bar{A}_H - \bar{A}_L = (\bar{X}_H - \bar{X}_L)\hat{\beta}_L + \bar{X}_H(\hat{\beta}_H - \hat{\beta}_L) \quad (6)$$

In each equation, the first term represents the achievement difference due to varying endowments of the independent variables in each country. The empirical analysis will further divide this component into the portions attributable to differing endowments of each independent variable, such as parent education and teacher training.

The second term is the achievement difference that is attributable to differences in the marginal products of the independent variables across countries. It is also a measure of our ignorance, in that it subsumes the influences of unmeasured variables (Filer, 1983).[10] The empirical analysis does not present detailed decomposition results for this second term, instead treating it as a single residual. This is for a specific empirical reason, which is frequently ignored in the empirical literature on the Blinder–Oaxaca decomposition. When decomposing this term, the portions attributable to individual variables and the constant are sensitive to choices about excluded categories of dummy variables (Jones, 1983; Oaxaca & Ransom, 1999).

As equations (5) and (6) show—and as the literature has frequently emphasized (Neumark, 1988)—two alternative decompositions can be derived. In equation (5), the mean differences in the independent variables are multiplied by the coefficients from the high-achieving country. In equation (6), they are multiplied by coefficients from the low-achieving country. There is no unambiguous method for choosing between the two decompositions.[11] Hence, the empirical analysis will present results from both, in order to assess whether they yield a similar pattern of results.

Data

The UNESCO Survey

In 1997, the Santiago office of UNESCO implemented an assessment of student achievement in Latin America, working in collaboration with the Ministries of Education of 13 countries. Using a common sampling methodology and set of survey instruments, researchers in each country collected representative samples of data on third-grade and fourth-grade achievement in language and mathematics, as well as data from background surveys of students, parents, teachers, and principals.

The sample was stratified and multi-stage. In each country, five strata were established: (1) public schools in cities with more than one million inhabitants; (2) private schools in cities with more than one million inhabitants; (3) public schools in cities with between 2500 and one million inhabitants; (4) private schools in cities with between 2500 and one million inhabitants; and (5) schools in cities with less than 2500 inhabitants. In Cuba, of course, private schools were not sampled because they do not exist.

Within each stratum, the sampling was conducted in two stages. Schools were defined as the primary sampling unit and, in the first stage, schools were sampled with a probability proportional to their enrollments. In the second stage, a fixed number of students within each school were sampled.[12]

This paper analyzes data from two countries in the UNESCO survey: Cuba and Mexico. They were chosen for several reasons.[13] First, there are large differences in mean Spanish and mathematics achievement across the countries. On average, Cuban students score 100 points higher in mathematics (or, when divided by the pooled standard deviation, 1.3 standard deviations). In language, the Cuban advantage is approximately 89 points—also 1.3 standard deviations. As noted previously, Cuban students score higher, on average, than every country in the UNESCO data (Willms & Somers, 2001).[14] Second, there are suggestive cross-country differences between Cuba and Mexico in the mean values of many family and school variables (see later). This study's goal is to assess whether differing endowments of these variables are, in part, responsible for achievement gaps. Third, the samples of both countries are relatively high quality, especially when compared with other countries in the UNESCO data. For example, both the Honduran and Venezuelan samples contained a large proportion of missing data for key independent variables such as parent education—upwards of 50% of total observations.

Independent Variables

Table 1 presents the definitions of the independent variables used in the analysis. They are divided into three categories: student and family variables, peer variables, and school and teacher variables.

Family influences include parental schooling and the number of books in the household, both proxies for the educational environment in the home and family income. Peer variables include the average socioeconomic status of a student's classmates, and the the average frequency of disciplinary problems in the classroom. Both are constructed from individual level variables.[15] The school variables describe students' prior educational experiences, the availability of instructional materials, the quality of school infrastructure, teachers' qualifications, and principals' autonomy in school management. They also include dummy variables that measure whether schools are urban or rural.

The means and standard deviations of these variables are presented in Table 2. They reveal considerable differences across countries in all three categories of variables. For example, the parents of Cuban students have markedly higher levels of education than their Mexican counterparts—roughly 12 years in Cuba compared with eight years in Mexico. Cuban parents also have a greater number of books in the household, on average. Not surprisingly, the average of peer variables is also quite different. Cuban classrooms have higher levels of socioeconomic status, on average, and fewer classroom disturbances.

Among school and teacher variables, several patterns can be noted. Students in Cuba are somewhat more likely to have attended preschool, and their teachers are much more likely to hold a post-secondary degree. However, Cuban and Mexican students attend classrooms with fairly similar levels of instructional materials and in similar physical conditions. Notwithstanding Cuba's nominally centralized system and Mexico's decentralization in 1993, principals in each country report having comparable levels of autonomy in managing personnel, budgets, and other school affairs; Cuba is even a bit higher.[16]

Table 1. Variable definitions

Variable	Definition
Family and student variables	
FEMALE	1 = female, 0 = male
PARENTED	Average years of schooling for both parents
BOOKS	Ordinal measure of the number of books in household: 1 = "no books"; 2 = "less than 10 books"; 3 = "10–50 books"; 4 = "more than 50 books"
Peer variables	
CLASSES	Classroom average for principal component factor analysis using PARENTED and BOOKS
DISTURB	Classroom averages for student responses on two questions: "in our class some students disturb the class very much" (1 = yes; 0 = no); and "we always get into fights" (1 = yes; 0 = no)
School and teacher variables	
GRADE	1 = student in fourth grade; 0 = student in third grade
PRESCHOOL	1 = Student attended preschool; 0 = not
TEXTBOOK	1 = student indicated in interview he/she has a Spanish or mathematics textbook; 0 = no
TRAIN	Number of training courses taken by teacher in last three years
DEGREE	1 = teacher has post-secondary degree, 0 = teacher has secondary degree
CONDITION	Average condition of classroom (according to teacher) for lighting, temperature, hygiene, safety and acoustics: 1 = "adequate"; 0 = "inadequate"
MATERIALS	Sum of materials available in classroom (according to teacher): blackboard, classroom library, calculators, games, maps/globes, overhead projector, slide projector, geometry materials, textbooks, computer, television, and videotape record
AUTONOMY	Average degree of autonomy (according to principal) for hiring/firing teachers, budget allocation, textbook/material selection, student admissions/suspensions, student promotion, rules, pedagogical prioritizing, planning extra-curricular activities: 1 = "no autonomy"; 2 = "partial autonomy"; 3 = "full autonomy"
LARGECITY	1 = school located in city with population over one million, 0 = not
URBAN	1 = school located in city with population between 2500 and one million, 0 = not

Results

Achievement Regressions

Table 3 presents estimates of country-specific regressions, the results of which are generally consistent with production function analyses in other countries of Latin America (Vélez *et al.*, 1996). Parental education, not surprisingly, is positively related to achievement in both countries, as is the number of books in the household. With a few exceptions, these coefficients are statistically significant at 5%.

The results also suggest that individual achievement is affected by characteristics of peer groups. The socioeconomic status of classrooms is positively related to achievement, although coefficients are only statistically significant in Mexico. As the frequency of classroom disturbances increases, achievement declines substantially. All of these coefficients are statistically significant at 5%.

In comparison with the student and family variables, the school and teacher variables generally exhibit less consistent links to academic achievement. Three results should be emphasized. First, the number of teacher training courses is positively related to achievement, but only in Cuba. Second, the presence of textbooks

Table 2. Means and standard deviations, by country and subject

	Mathematics		Spanish	
	Cuba	Mexico	Cuba	Mexico
MATH/	356.70	256.94	341.51	253.78
SPANISH	(68.29)	(43.92)	(51.20)	(53.87)
Student and family variables				
FEMALE	0.52	0.49	0.52	0.49
PARENTED	12.08	8.17	11.98	8.19
	(3.25)	(3.80)	(3.27)	(3.78)
BOOKS	2.67	2.30	2.65	2.30
	(0.90)	(0.90)	(0.90)	(0.90)
Peer variables				
SCHOOLSES	0.39	-0.14	0.37	-0.14
	(0.33)	(0.48)	(0.32)	(0.48)
DISTURB	0.27	0.57	0.27	0.57
	(0.21)	(0.16)	(0.21)	(0.16)
School and teacher variables				
GRADE	0.49	0.48	0.50	0.47
PRESCHOOL	0.94	0.85	0.94	0.85
TEXTBOOK	0.95	0.92	0.97	0.96
TRAINING	4.05	4.55	4.08	4.62
	(6.68)	(3.36)	(6.94)	(3.41)
DEGREE	0.92	0.40	0.92	0.39
CONDITION	0.78	0.76	0.78	0.77
	(0.28)	(0.26)	(0.28)	(0.26)
MATERIALS	5.89	6.39	5.84	6.41
	(1.26)	(2.22)	(1.18)	(2.22)
AUTONOMY	2.43	2.21	2.42	2.20
	(0.32)	(0.43)	(0.32)	(0.42)
LARGECITY	0.33	0.24	0.31	0.24
URBAN	0.31	0.44	0.31	0.44
n	3353	1649	2977	1547

Note: Standard deviations are not reported for dummy variables.

is positively related to achievement in both countries, but only for Spanish. Third, the presence of other instructional materials in the classroom is always positively associated with achievement, although two coefficients are not statistically significant.

Decomposition of Achievement Gaps

Table 4 presents the results of four achievement decompositions. The first two are based on equation (5), utilizing the Cuban coefficients to weight mean differences in the independent variables. The final two are based on equation (6) and the Mexican coefficients.

There is a mean difference of 100 points in mathematics achievement, favoring Cuba. The decompositions suggest that between 7.4 and 10.7 points of this difference is explained by differences in the endowments of student and

Table 3. Determinants of mathematics and Spanish achievement, by country

	Mathematics		Spanish	
	Cuba	Mexico	Cuba	Mexico
Student and family variables				
FEMALE	3.41	0.26	7.17	7.48
	(1.31)	(−0.11)	(3.83)	(3.19)
PARENTED	1.90	1.55	0.28	2.41
	(3.79)	(4.18)	(0.73)	(5.46)
BOOKS	8.66	3.74	2.76	2.31
	(5.09)	(2.56)	(2.31)	(1.40)
Peer variables				
SCHOOLSES	12.47	16.54	8.47	17.25
	(1.17)	(3.11)	(0.96)	(3.13)
DISTURB	53.49	26.04	28.21	23.38
	(−3.14)	(−2.71)	(−2.51)	(−2.45)
School and teacher variables				
GRADE	1.23	19.21	5.63	21.20
	(0.90)	(5.91)	(1.75)	(7.13)
PRESCHOOL	10.6	8.36	2.37	5.52
	(−1.08)	(2.54)	(0.39)	(1.57)
TEXTBOOK	−5.73	5.14	14.64	17.11
	(−0.91)	(1.12)	(2.23)	(3.35)
TRAIN	1.6	0.52	0.08	−0.28
	(3.91)	(−1.20)	(2.32)	(−0.57)
DEGREE	5.02	5.42	5.70	5.38
	(0.50)	(1.79)	(0.83)	(1.44)
CONDITION	−5.83	5.02	17.93	−1.54
	(−0.44)	(0.67)	(2.00)	(−0.24)
MATERIALS	6.59	1.02	2.53	3.63
	(2.16)	(1.26)	(1.27)	(3.76)
AUTONOMY	0.88	−4.03	1.94	0.69
	(0.08)	(−1.01)	(−0.21)	(0.14)
LARGECITY	−5.95	5.03	11.37	6.73
	(−0.62)	(1.10)	(1.39)	(1.25)
URBAN	−4.88	6.3	7.88	9.54
	(−0.46)	(1.43)	(1.06)	(2.03)
Constant	287.58	226.73	289.21	178.16
	(8.66)	(16.85)	(11.00)	(11.05)
n	3353	1649	2977	1547
R^2	0.13	0.233	0.083	0.271

Note: *t*-statistics are in parentheses. Standard errors are adjusted for stratified and multi-stage sampling (see text for details).

family variables –notably PARENTED and, to a lesser extent, BOOKS. Overall, this represents between 8% and 11% of the total achievement gap. An even larger portion, between 16.6 and 22.7 points, is explained by differences in the endowments of peer variables. This represents 17–23% of the total achievement gap. The logic of this finding is straightforward. On average, Cuban students are exposed to higher levels of desirable attributes (such as parental education) and

Table 4. Decomposition of achievement differential between Cuba and Mexico

	Using Cuban coefficients		Using Mexican coefficients	
	Mathematics	Spanish	Mathematics	Spanish
Explained by endowments of student and family variables				
FEMALE	0.10	0.22	-0.01	0.22
PARENTED	7.43	1.06	6.06	9.13
BOOKS	3.20	0.97	1.38	0.81
Subtotal (a)	10.74	2.24	7.44	10.17
Explained by endowments of peer variables				
SCHOOLSES	6.61	4.32	8.77	8.80
DISTURB	16.05	8.46	7.81	7.01
Subtotal (b)	22.66	12.78	16.58	15.81
Explained by endowments of school and teacher variables				
GRADE	0.04	0.17	0.19	0.73
PRESCHOOL	-0.95	−0.21	0.75	0.50
TEXTBOOK	−0.17	0.15	0.15	0.17
TRAIN	−0.80	0.37	0.26	0.15
DEGREE	2.61	3.02	2.82	2.85
CONDITION	0.12	0.18	0.10	0.02
MATERIALS	-3.30	−1.44	0.51	2.07
AUTONOMY	0.19	0.43	−0.89	0.15
LARGECITY	-0.54	0.80	0.45	0.17
URBAN	0.63	−1.02	0.82	1.24
Subtotal (c)	−2.39	5.20	2.51	1.69
Total, (a)+(b)+(c) = (d)	31.00	9.82	26.53	27.67
Total Achievement Gap (e)	99.75	88.74	99.75	88.74
(a)/(e)	0.108	0.025	0.075	0.115
(b)/(e)	0.227	0.144	0.166	0.178
(c)/(e)	0.024	−0.059	0.025	0.019
(d)/(e)	0.311	0.111	0.266	0.312

lower levels of undesirable attributes (such as classroom disturbances). All of these variables have the predicted effects on achievement, and serve to widen the achievement gap.

In contrast, the difference in mathematics achievement is not explained well by the observed school and teacher variables. In fact, the decomposition with the Cuban coefficients yields a value of −2.4 points, driven by the variable MATERIALS. This stems from the fact that Cuban students have fewer classroom materials, on average. Since this variable is positively related to achievement, one would expect Mexican students to have higher achievement. The decomposition produces slightly different results with the Mexican coefficients (2.5 points). In neither case, however, is the magnitude appreciable.

With few exceptions, the results for the Spanish decompositions do not alter this general pattern of results. Overall, the results suggest that no more than 31% of the substantial differences in mean achievement between Cuba and Mexico can be explained by differing endowments of the independent variables. Of the latter, peer-group variables and, to a lesser extent, family variables explain the largest portion of the gap.

Conclusions

This paper applied the Blinder–Oaxaca decomposition to the problem of explaining the determinants of cross-country achievement differences. Using data from Cuba and Mexico, the achievement difference between each pair of countries was decomposed into four components: (1) the component attributable to differing endowments of family and student variables; (2) the component attributable to differing endowments of peer variables; (3) the component attributable to differing endowments of school and teacher variables; and (4) a residual component that cannot be explained by endowments of observed variables.

The findings suggest several conclusions. First, they confirm that socioeconomic status is a strong and consistent determinant of academic achievement in all samples. Differences in these variables account for 3–12% of the achievement gap. Second, individual achievement is strongly affected by the average characteristics of peer groups, with at least 14–23% of the achievement gap explained by differences in variable endowments.

Third, the findings suggest that observed school and teacher variables are inconsistently linked to achievement. Even when positively associated with achievement in country samples, these variables are not always distributed unevenly across countries. Thus, they can explain very little of the observed difference in achievement. It should be noted, however, that the Cuban school system still appears to be affecting the achievement gap through its long-term affects on the educational attainment of parents.

On the whole, a substantial portion of the gap—more than two-thirds—cannot be explained by differing endowments of the independent variables. This finding should be interpreted carefully.[17] It is tempting, perhaps, to infer that 'schools don't matter' (or even 'families don't matter'). Instead, the results suggest that we are doing a poor job of measuring the relevant determinants of student achievement. To provide a simple example, a recent case study of Cuban education asserts that teacher training and support networks are unique within Latin America (Gasperini, 1999). The extent and quality of these interventions are not measured in the UNESCO data, and their effectiveness cannot be empirically assessed. Is the measured advantage of Cuban students partly due to such training initiatives? Unfortunately, this paper—and others based on international surveys—are hard-pressed to answer these questions.

To construct a full explanation of achievement differences, it is clear that data need to be improved and that empirical techniques should be used that are better suited to the question at hand. Minimally, the results in this paper should make us leery of facile attempts to construct a single 'explanation' for the entirety of cross-country achievement differences.

Acknowledgements

The authors thank Luis Benveniste, Martin Carnoy, Sandra Cusato, Juan Carlos Palafox, and Marie-Andrée Somers for helpful conversations. These individuals are not responsible for any errors or interpretations in the analysis.

Notes

1. For example, see Baker (1993) and Westbury (1992, 1993).
2. The project is referred to as the Latin American Laboratory for the Evaluation of School Quality. See Willms and Somers (2001) for a descriptive analysis of the results, as well as within-country statistical analyses of the correlates of achievement.
3. These include the First International Mathematics Study (FIMS), the Second International Mathematics Study (SIMS), and the Third International Mathematics and Science Study (TIMMS). For an overview of IEA studies and other international assessments, see Mullis (1997), Garden (1987), Husen (1974), Postlethwaite (1987), and Theisen *et al.* (1983).
4. For a critique of Westbury (1992, 1993), see Baker (1993).
5. As the authors note, this is not a gain score in the true sense, since the data are not longitudinal. However, since the TIMMS study includes two adjacent grades in each country it was possible to subtract the average score of the lower grade from the higher grade to arrive at a rough indicator of average learning for the year.
6. They do include the Gross National Product (GNP) as a control in some of these regressions, but it is doubtful that the total size of the national economy is an adequate control for differences in parental education and many other determinants of academic achievement. For example, Spain has a larger economy than Switzerland, and South Africa's GNP is almost as large as Norway's.
7. See Blinder (1973) and Oaxaca (1973) For an application to black white earnings differentials in the United State see Ashraf (1994); on indigenous and non-indigenous earnings differentials in Latin America, see Psacharopoulos and Patrinos (1994).
8. The theoretical foundations of the present application are not as rich as those of wage discrimination. Nonetheless, the Blinder–Oaxaca technique can be easily applied to familiar– and theoretically defensible specifications of education production functions. If nothing else, it is intuitively appealing and it maps well onto a empirical question that has not been well addressed in the empirical literature: what are the determinants of cross country achievement variation?
9. For a similar discussion, see Neumark (1988).
10. In wage decompositions, this term is interpreted as an upper-bound estimate of discrimination. It is inclusive of different returns to worker characteristics, as well as unmeasured variables that determine wages.
11. In wage comparisons, the choice depends on whether the high-wage structure or low-wage structure is thought to prevail in the absence of discrimination. Neumark (1988) derives an alternative estimator, in which differences in variable endowments are weighted by coefficients from a pooled regression of the two groups. However, this is based on an explicit theory of wage determination and discrimination, and has no obvious relation to the present empirical context.
12. If one applies ordinary least squares to data with a stratified or multi-stage sample design, it is well known that the standard errors of regression coefficients will be incorrect. This paper's regression analyses use the modified formulas for standard errors described in Cochran (1977) and Deaton (1997), and implemented in the statistical package Stata (StataCorp, 2001).
13. To some extent, of course, the choice of Cuba and Mexico is arbitrary. The same decomposition techniques could be applied to data from any pair of countries.
14. The large Cuban advantage has been the source of some debate. For example, it spurred a World Bank working paper that sought to explain the gap as a function of innovative schooling policies in Cuba, as compared with other Latin American countries (Gasperini, 1999). These findings are largely anecdotal, however.
15. CLASSES was constructed with principal components analysis using PARENTED and BOOKS using a two-step process. In the first, principal component analysis was used for individual data to create a family SES factor using these two variables. The resulting factor explains approximately 75% of the variance, and the two loadings for PARENTED and BOOKS are each 0.857. This variable was then averaged at the classroom level in order to create a classroom-specific SES factor. All of the results in the text still hold if the variables are not combined in an index, however.
16. Anecdotal evidence suggests that some functions in Cuba like teacher evaluation are conducted mainly at the local level (Gasperini, 1999). Moreover, the Mexican decentralization transferred decision-making authority from the national to state governments: it is perhaps not surprising, then, that schools still have minimal authority.

17. One problem that we have not emphasized is the assumption that all independent variables are exogenous (and its corollary, that variable coefficients are unbiased estimates). It is increasingly recognized that sorting by families and resource allocation decisions by schools and governments may create correlations between independent variables and the error term, biasing coefficients in unknown directions. This paper has not addressed this issue.

References

Ashraf, J. (1994) Recent trends in white-black earnings differentials, *Economics Letters*, 45, pp. 487–493.

Baker, D. (1993) Compared to Japan, the U.S. is a low achiever ... really, *Educational Researcher*, 22(3), pp. 18–20.

Blinder, A. S. (1973) Wage discrimination: reduced form and structural estimates, *Journal of Human Resources*, 8(4), pp. 436–455.

Cochran, W. G. (1977) *Sampling Techniques* (New York, Wiley).

Deaton, A. (1997) *Analysis of Household Surveys: A Microeconometric Approach to Development Policy* (Washington, DC, World Bank).

Filer, R. K. (1983) Sexual differences in earnings: the role of individual personalities and tastes, *Journal of Human Resources*, 18(1), pp. 82–99.

Finn, C. (1991) *We Must Take Charge* (New York, Free Press).

Garden, R. A. (1987) The second IEA mathematics study, *Comparative Education Review*, 31(1), pp. 47–68.

Gasperini, L. (1999) *The Cuban Education System: Lessons and Dilemmas, LCSHD Paper Series No. 48* (Washington, DC, World Bank).

Hess, R. D. & Azuma, H. (1991) Cultural support for schooling: contrasts between Japan and the United States, *Educational Researcher*, 20(9), pp. 2–8.

Husen, T. (1974) Multi-national evaluation of school systems: purposes, methodology and some preliminary findings, *Scandinavian Journal of Educational Research*, 18(1), pp. 13–39.

Jaeger, R. M. (1992) Weak measurement serving presumptive policy, *Phi Delta Kappan*, 74, pp. 118–128.

Jones, F. L. (1983) On decomposing the wage gap: a critical comment on Blinder's method, *Journal of Human Resources*, 18(1), pp. 126–130.

McKnight, C. C. *et al.* (1987) *The Underachieving Curriculum: Assessing U.S. School Mathematics from an International Perspective* (Champaign, IL, Stipes).

Mullis, I. V. S. (1997) *Mathematics Achievement in the Primary School Years: IEA's Third International Mathematics and Science Study* (Chestnut Hill, MA, Boston College).

National Education Commission on Time and Learning (1994) *Prisoners of Time* (Washington, DC, US Government Printing Office).

Neumark, D. (1988) Employers' discriminatory behavior and the estimation of wage discrimination, *Journal of Human Resources*, 23(3), pp. 279–295.

Oaxaca, R. (1973) Male–female wages differentials in urban labor markets, *International Economic Review*, 14(3), pp. 693–709.

Oaxaca, R. L. & Ransom, M. R. (1999) Identification in detailed wage decompositions, *Review of Economics and Statistics*, 81(1), pp. 154–157.

Postlethwaite, T. N. (1987) Comparative education achievement research: can it be improved?, *Comparative Education Review*, 31(1), pp. 150–158.

Psacharopoulos, G. & Patrinos, H. A. (1994) *Indigenous People and Poverty in Latin America: An Empirical Analysis* (Washington, DC, World Bank).

Schaub, M. & Baker, D. P. (1991) Solving the math problem: exploring mathematics achievement in Japanese and American middle grades, *American Journal of Education*.

Schmidt, W. H. *et al.* (2001) *Why Schools Matter: Using TIMMS to Investigate Curriculum and Learning*, unpublished manuscript.

StataCorp. (2001) *Stata Statistical Software: Release 7.0* (College Station, TX, Stata Corporation).

Stedman, L. C. (1994) Incomplete explanations: the case of U.S. performance in the international assessments of education, *Educational Researcher*, 23(7), pp. 24–32.

Stigler, J. W. & Stevenson, H. W. (1991) How Asian teachers polish each lesson to perfection, *American Educator*, pp. 13–20, 43–46.

Theisen, G., Achola, P. P. W. & Musa Boakari, F. (1983) The underachievement of cross-national studies of achievement, *Comparative Education Review*, 27(1), pp. 46–68.

Vélez, E., Schiefelbein, E. & Valenzuela, J. (1996) *Factores que afectan el rendimiento académico en la educación primaria* (Mexico, Secretaría de Educación Pública).

Westbury, I. (1992) Comparing American and Japanese achievement: is the United States really a low achiever?, *Educational Researcher*, 21(5), pp. 18–24.

Westbury, I. (1993) American and Japanese achievement … again, *Educational Researcher*, 22(3), pp. 21–25.

Willms, J. D. & Somers, M.-A. (2001) Family, classroom and school effects on children's educational outcomes in Latin America, *School Effectiveness and School Improvement*, 12(4), pp. 409–445.