Asia Pacific Educ. Rev. (2011) 12:463–474
DOI 10.1007/s12564-011-9151-3

Decomposing achievement gaps among OECD countries

Liang Zhang · Kristen A. Lee

Received: 4 May 2010 / Revised: 8 January 2011 / Accepted: 4 February 2011 / Published online: 23 February 2011

Abstract In this study, we use decomposition methods on PISA 2006 data to compare student academic performance across OECD countries. We first establish an empirical model to explain the variation in academic performance across individuals, and then use the Oaxaca-Blinder decomposition method to decompose the achievement gap between each of the OECD countries and the OECD average. Results indicate that the explained portion of the achievement gap varies across countries. In some countries, our empirical models are able to account for almost all the achievement gap, while unexplained country-specific effects still dominate in other countries. Finally, we use two Asian countries, Japan and Korea, to demonstrate how to identify major factors that have contributed to the observed achievement gap across countries.

Keywords Academic performance · International comparison · Decomposition

Introduction

The study on student achievement gaps across countries has recently been fueled by the availability of international student-level datasets such as the Trends in International Mathematics and Science Study (TIMSS), OECD's Program for International Student Assessment (PISA), and Progress in International Reading Literacy Study (PIRLS). Much of the empirical work in this area has focused on explaining the variation in student achievement (e.g., Ammermüller 2007; Baker et al. 2002; Kotte et al. 2005; Fertig and Schmidt 2002). In a nutshell, student academic performance is significantly related to their individual, family, teacher, school, and country-level variables; however, even after controlling for these variables, a large proportion of individual-level variance remains unexplained. The typical R-squared for an individual-level model is about 30–40% (e.g., Fuchs and Wößmann 2007). These models can be further aggregated to explain the variation in average student academic performance across countries. For example, Fuchs and Wößmann (2007) found that the aggregated student, family, teacher, and school characteristics explain a vast majority (over 85%) of country-level variation in student performance.

Focusing on variance reduction helps build a theoretical model that maximizes the power of explaining the variation across students and countries; however, such analyses cannot quantify the extent to which empirical models explain the gaps among countries. In other words, variance reduction does not provide useful information for policymakers of a particular country who wish to better understand how the achievement gap between one country and others can (and cannot) be explained by empirical models. This question requires a different approach. In this study, we use decomposition methods to examine how observed gaps across countries in student achievement can be explained by differences in observed characteristics of their students, families, and schools. Specifically, we will use the student test scores in mathematics, science, and reading in PISA 2006 and model each of them as a function of a host of explanatory variables. We then decompose the achievement gap between each of the 30 OECD countries and the OECD average into a proportion explained by our empirical models and a country-specific proportion.

L. Zhang (✉) · K. A. Lee
Department of Education Policy Studies, The Pennsylvania State University, 310E Rackley Building, University Park, PA 16802, USA
e-mail: lxz19@psu.edu

Finally, using Korea and Japan as examples, we examine the major factors that contribute to the observed student achievement gaps across countries.

Literature review

For decades, there has been wide interest from educators, researchers, and policymakers in the comparison of educational achievement within and between countries. In recent years, the availability of international assessment studies has further promoted research in this area. Most studies use data from international tests to substantiate which predictors affect student achievement within and across countries. These variables include student characteristics (Williams et al. 2005; Shin et al. 2009), intelligence (Lynn and Mikk 2007, 2009; Rinderman 2007), access to modern information technology (Bielefeldt 2005), curriculum and instruction (Schuab and Baker 1991; Westbury 1992; MacNab 2000; Baker et al. 2003; Tatsuoka et al. 2004; Wang and Lin 2005), and school characteristics (Fuchs and Wößmann 2007).

Although there has been no agreed-upon set of variables that exhibit the greatest effect on student achievement, certain factors appear to contribute, at different strengths, to student achievement. Individual characteristics such as gender, age, and parents' origin and education seem to matter. Using data from international assessments, researchers have attempted to explicate differences in achievement between subgroups by gender (Hanna 1989; Casey et al. 2001) and immigration status (Marks 2005). While individual characteristics are powerful predictors for individual-level differences, they predict differences across countries poorly. School resources and institutional characteristics are better predictors in this regard. For example, Anderson et al. (2007) use all countries that participated in PISA 2000 and 2003 to examine both country-level and school-level variations of student performance. Their results suggest that school-level variables explain a large portion of the variation in the differences of student performance for all countries. On average, 34% of the variance in mathematical performance is due to school characteristics, while 17% of the variation can be explained by the expenditure per student. Similarly, Wößmann (2000) argues that the effects on German students' performances on the PISA in 2000 are largely attributed to the country's educational institutions. Including institutional level factors such as the autonomy of a school and school funding sources, Fuchs and Wößmann (2007) attribute 25% of the variation in student achievement cross-country to school-level characteristics in their full model.

Models based on variance reduction, however, cannot quantify the degree to which the gaps among countries can be explained. Therefore, in this analysis, we use the decomposition method in economics to examine the extent to which the observed gaps across OECD countries in student achievement can be explained by differences in observed characteristics of their students, families, and schools. We are not the first to use the decomposition method to study the achievement gaps across countries. Previous studies utilize this method to analyze and compare among student groups. Ammermüller (2007) uses student achievement data in Germany and Finland from PISA 2003 to explain the gap between these countries. McEwan and Marshall (2004) use the decomposition method to explain the achievement gap between Cuba and Mexico. Finally, Hernandez-Zavala et al. (2006) use data from Guatemala, Peru, and Mexico to explain the achievement gap between indigenous and non-indigenous students in Latin America. Unlike previous studies that are limited by the number of countries (or groups), in this analysis, we compare each of the OECD countries with the OECD average, thus providing a more systemic and thorough analysis of the achievement gaps across countries.

Data and methods

The data used in this analysis comes from PISA 2006. Since 2000, the OECD has administered PISA every 3 years. This data collection effort focuses on 15-year-olds' capacities in mathematics, science, and reading literacy near the end of mandatory education to assess how well these young adults are prepared to meet the challenges of the knowledge-based economy (OECD 2004). The 2006 wave of PISA data includes all 30 OECD member countries and 27 non-OECD countries and regions. PISA provides very rich information on students, their parents and families, and schools. In this analysis, we only included the 30 OECD countries. As members of OECD, these countries share similar social and political characteristics, making it meaningful to estimate a common model for these countries while simultaneously being sufficiently heterogeneous in terms of their educational and economic development, making it possible to compare across these countries.

The dependent variables used in the analysis are student test scores in mathematics, science, and reading. They are measured by multiple test items, which test students' basic knowledge and their ability to use their knowledge and skills in solving real-life problems. The relevance of these test scores has been confirmed by studies that have followed up students in the first PISA 2000 study over years. For example, Knighton and Bussiere (2006) found that in Canada there is a strong positive relationship between reading proficiency achieved at age 15 and the probability of attending college at age 19. Student test scores in PISA are computed by item response models. On each subject, five plausible values are reported that represent the range of test scores that a student might reasonably have. While this presents methodological advantages over the traditional point estimates, it also poses challenges to researchers when using these plausible values. For a detailed discussion on plausible values, see Wu and Adams (2002) and OECD (2009). In this analysis, we adopt the technical procedures recommended by OECD (2009). To estimate sample statistics, each plausible value was used separately and then averaged. Table 1 reports the average test scores on each of the three subjects across the 30 OECD countries.

One technical issue that has important implications in this analysis is the weight in PISA data. Weight is necessary due to various reasons. Countries differ in the size of their 15-year-olds' population, and the number of students in the PISA data is not proportional to the population in their respective countries. Consequently, one might expect that when all students in OECD countries are viewed as a whole, the weight of a particular student should be proportional to the student population in his/her country. For example, the average weight of students from Luxembourg is 1.036 while that of the United States is 637.683. In addition, the weight within a particular country varies because of unequal probability of selection for students and schools within the same country, adjustments for non-responses, and intentional oversampling of certain strata for national reporting purposes (OECD 2009). In short, weighting makes the PISA sample representative of the student population from all OECD countries.

There is, however, a slight complication when we compare across countries. For example, Table 1 shows that the average math score in Luxembourg is 490.00 while that in the United States is 474.35. Because the average student weight in the United States is much larger than that in Luxembourg, the average test score of these two countries when weighted by student weights in PISA is 474.38, which is very close to the U. S. average. This is against our intuitive interpretation of averages among countries, which usually assumes equal contribution of each country, regardless of the size of population. To compute the average among countries, one can simply average the country-averages across countries. Alternatively, one can use the student weights in PISA and scale them to allow equal total weights in each country. This rescaled student weight can then be used to compute averages across countries. In Table 1, the OECD average weighted by the original PISA student weights is 483.67, while the average weighted by the rescaled student weights is 497.68. The latter is used to compute the achievement gap between each OECD country and the overall OECD averages.

A variety of independent variables are used in our empirical model. The first two columns of Table 2 provide the descriptive statistics of these variables classified into four conceptual groups—individual characteristics, study time and activities, family backgrounds, and school characteristics. Individual characteristics consist of basic demographic information such as the grade that a student is in, age in months, and gender. These demographic

Table 1 Mean test scores across OECD countries, PISA 2006

Country	Math		Science		Read	
	Mean[a]	SE[b]	Mean[a]	SE[b]	Mean[a]	SE[b]
Finland	548.36	2.30	563.32	2.02	546.87	2.15
Korea	547.46	3.76	522.15	3.36	556.02	3.81
Netherland	530.65	2.59	524.86	2.74	506.75	2.92
Switzerland	529.66	3.15	511.52	3.16	499.28	3.06
Canada	527.01	1.97	534.47	2.03	527.01	2.44
Japan	523.10	3.34	531.39	3.37	497.96	3.65
New Zealand	521.99	2.39	530.38	2.69	521.03	2.99
Belgium	520.35	2.95	510.36	2.48	500.90	3.04
Australia	519.91	2.24	526.88	2.26	512.89	2.06
Denmark	513.03	2.62	495.89	3.11	494.48	3.18
Czech Republic	509.86	3.55	512.86	3.48	482.72	4.18
Iceland	505.54	1.81	490.79	1.64	484.45	1.95
Austria	505.48	3.74	510.84	3.92	490.19	4.08
Germany	503.79	3.87	515.65	3.80	494.94	4.41
Sweden	502.36	2.41	503.33	2.37	507.31	3.44
Ireland	501.47	2.79	508.33	3.19	517.31	3.54
France	495.54	3.17	495.22	3.36	487.71	4.06
United Kingdom	495.44	2.14	514.77	2.29	495.08	2.26
Poland	495.43	2.44	497.81	2.34	507.64	2.79
Slovak Republic	492.11	2.82	488.43	2.59	466.35	3.06
Hungary	490.94	2.89	503.93	2.68	482.37	3.28
Luxembourg	490.00	1.07	486.32	1.05	479.37	1.28
Norway	489.85	2.64	486.53	3.11	484.29	3.18
Spain	479.96	2.33	488.42	2.57	460.83	2.23
United States	474.35	4.02	488.91	4.22	[e]	
Portugal	466.16	3.07	474.31	3.02	472.30	3.56
Italy	461.69	2.28	475.40	2.02	468.52	2.43
Greece	459.20	2.97	473.38	3.23	459.71	4.04
Turkey	423.94	4.90	423.83	3.84	447.14	4.21
Mexico	405.65	2.93	409.65	2.71	410.50	3.06
Weighted average[c]	483.67	1.15	490.84	1.17	483.82	1.03
Simple average[d]	497.68	1.13	500.00	1.21	491.79	1.11

[a] All means are computed using 5 plausible values of scores for each student, weighted by student final weight (w_fstuwt)

[b] Standard errors are computed with 80 replicates

[c] Weighted OECD average is computed using the student final weight (w_fstuwt)

[d] Simple OECD average normalizes the total weight from each member country to a constant, that is, equivalent to the arithmetic average of country means reported in this table

[e] Reading scores for the United States are not reported in PISA 2006

Table 2 Descriptive statistics of main variables and determinants of students' math test scores

Variable	Descriptive statistics		Regression analysis		
	Mean	SD	Regression coeff.[a]	Standardized beta	t-test[b]
Constant			−23.509		−0.97
Individual characteristics					
Grade	9.79	0.73	12.576	0.124	15.29
Age	15.78	0.3	6.503	0.021	4.62
Boy	0.5	0.5	23.066	0.122	30.80
First generation immigrant	0.04	0.19	−4.106	−0.008	−1.70
Second generation immigrant	0.05	0.22	−2.376	−0.005	−1.22
Other national language	0.02	0.13	−2.400	−0.003	−0.82
Other foreign languages	0.05	0.23	−6.146	−0.015	−3.05
Occupational aspiration	59.93	17.72	0.672	0.208	24.95
Study time and activities					
School time in science	3.21	2.1	8.128	0.185	9.99
School time in science^2			−0.662	−0.105	−5.50
Out-of-school time in science	0.7	1.24	−1.482	−0.019	−1.78
Out-of-school time in science^2			−0.049	−0.003	−0.27
Homework in science	1.56	1.56	1.039	0.017	1.02
Homework in science^2			−0.286	−0.027	−1.90
School time in math	3.89	1.89	8.092	0.170	7.72
School time in math^2			−0.770	−0.119	−5.66
Out-of-school time in math	1.07	1.54	−2.659	−0.043	−3.16
Out-of-school time in math^2			0.498	0.043	3.40
Homework in math	1.97	1.73	7.404	0.136	8.28
Homework in math^2			−0.707	−0.080	−4.38
School time in language	3.84	1.88	12.234	0.256	12.22
School time in language^2			−1.753	−0.269	−13.94
Out-of-school time in language	0.92	1.44	−8.971	−0.135	−7.95
Out-of-school time in language^2			0.788	0.060	3.89
Homework in language	1.78	1.63	−3.531	−0.061	−3.74
Homework in language^2			−0.007	−0.001	−0.05
School time in other subjects	3.98	2.26	2.989	0.074	3.36
School time in other subjects^2			0.516	0.100	4.87
Out-of-school time in other subjects	1.16	1.64	−4.703	−0.081	−7.19
Out-of-school time in other sub.^{2}			0.250	0.024	2.24
Homework in other subjects	2.07	1.85	5.008	0.098	6.87
Homework in other subjects^2			−0.577	−0.072	−5.29
Extra class time with instructor	0.29	0.45	−15.204	−0.073	−16.20
Extra class time no with instructor	0.27	0.44	−5.250	−0.025	−5.76
Required science courses	0.9	0.31	23.944	0.078	18.56
Optional science courses	0.37	0.48	−11.189	−0.057	−10.07
Family background					
Mother's occupational status	44.88	16.19	0.179	0.045	5.52
Father's occupational status	43.99	17	0.333	0.078	10.69
Parent's highest education	12.87	3.28	2.386	0.096	3.41
Parent's highest education^2			−0.013	−0.011	−0.41
Wealth	−0.19	1.03	0.506	0.006	0.84
Cultural possessions	−0.09	0.98	1.135	0.012	2.77

Table 2 continued

Variable	Descriptive statistics		Regression analysis		
	Mean	SD	Regression coeff.[a]	Standardized beta	t-test[b]
Educational resources	−0.18	1.03	6.668	0.072	14.47
Number of books (in 10)	13.39	15.05	2.652	0.421	23.30
Number of books (in 10)^2			−0.034	−0.260	−14.75
School characteristics					
Percentage of girls	0.49	0.17	4.650	0.011	0.71
Public school	0.78	0.42	−11.629	−0.051	−6.05
Percentage of students repeating a grade	3	5.32	−1.065	−0.060	−5.56
Class size	28.47	8.77	3.827	0.473	7.50
Class size^2			−0.054	−0.347	−7.25
Student–teacher ratio	15.53	6.85	−0.436	−0.039	−3.06
Percentage of teachers with certificate	0.92	0.22	11.721	0.056	3.17
Percentage of teachers with BA	0.83	0.32	7.403	0.036	3.02
Percentage of computers with internet	0.86	0.25	21.582	0.082	6.51
Short of science teachers	0.73	0.45	2.855	0.013	1.00
Short of math teachers	0.74	0.44	4.768	0.022	1.44
Short of language teachers	0.8	0.4	8.593	0.036	3.14
Number of observations			251,278		
Average R-squared			0.4792		

Regression model includes missing value indicators

[a] Regression model is estimated using 5 plausible values of scores for each student, weighted by student final weight (w_fstuwt)

[b] *t*-statistics (i.e., standard errors) are computed with 80 replicates

variables have been shown to have a significant relationship with test scores (Fuchs and Wößmann 2007; Marks 2005; Fertig 2003; Schütz et al. 2007). In addition, we use information on the birth places of students and their parents to construct three immigration categories, namely, first generation immigrants, second generation immigrants, and native students. Dummy variables are used to indicate first and second generation immigrants. Previous research has shown that the language(s) spoken at home also affect student academic performance. We include dummy variables to indicate that a student speaks a language other than the test language, and we further differentiate whether the student speaks another national language or a foreign language. Finally, we include student occupational aspiration measured by the ISEI scale (Ganzeboom et al. 1992).

Study time and activities variables include information on student time allocation and types of out-of-school activities. PISA asks students the amount of time they typically spend per week studying each of the four subject areas (e.g., science, mathematics, language and literature, and other subjects) in the following three activities: regular classes, out-of-school classes, and homework. Student time allocation is measured by students' report on their time use in each of the 12 categories: no time, less than 2 h a week, 2–4 h a week, 4–6 h a week, and more than 6 h a week. Treating these data as categorical information will result in a large number of dummy variables. Consequently, we converted the information from the categories above into a continuous variable using the mid-point of each category. The squared term of time spent in each area is also used to capture the possibility of changing marginal effects of study time on academic performance. Furthermore, two dummy variables (with no out-of-school classes as the reference category) were created to distinguish between different types of out-of-school classes since students may have their out-of-school classes with their school teachers, or with someone who is not a teacher in their school.

Family background information consists of parental education, occupational status, and family wealth especially in those aspects relevant to education. The occupational status of each parent is measured according to the ISEI scale. We also use the educational level of the better-educated parent because our model does not detect statistically different effects between a mother's and father's education on student academic performance. Parental education is measured by the number of schooling years recoded from the International Standard Classification of Education. One important factor that has been consistently shown to affect student academic performance is family wealth. PISA collects data on household assets because they capture family wealth better than family income. In PISA 2006, students report the availability and quantities of about two dozen household items such as computers, works of arts, dishwashers, and guest rooms. These data were then used to derive three separate indices: family wealth possessions, cultural possessions, and home educational resources. OECD (2009) provides detailed information on how these indices are derived from the original item responses. Finally, the number of books at home is recoded from the 6 categories reported in PISA into a continuous variable, with its squared term to capture the changing marginal effect of the number of books on academic performance.

The final block of variables is school characteristics. We include basic information of schools such as whether it is a public school or not, the percentage of girls, the percentage of students who repeated a grade within the past 2 years prior to testing, class size, and student–teacher ratio. In addition, to capture the technology aspect of the learning environment, we include the proportion of computers with internet connection available in the school. Finally, teacher quality and quantity are measured by the proportion of teachers with certificates or BA degrees and whether a school is in short of science, mathematics, and language teachers.

Our analysis consists of two distinct steps. In the first step, we establish a baseline model of student-level academic performance. OLS regressions are used to estimate the effects of individual characteristics, study time and activities, family background, and school characteristics on student test scores. Because the multistage sampling design adopted by PISA violates independent observations assumption in OLS, the estimated standard errors (and hence statistical tests based on these standard errors) may be incorrect. Recent research in this area has adopted variations of multilevel modeling (e.g., HLM, cross-classified models, etc.) to address problems associated with data collected through complex sample designs and to bring empirical models into closer congruence with the multilevel theoretical models. Because the focus of this analysis is to explain the achievement gap by using a decomposition method that has been developed using the OLS framework, we choose to use the more conventional OLS estimates. To address the issue of incorrect standard errors using OLS regression, we use the Balanced Repeated Replication method recommended by OECD (2009). Results using HLM models are very similar to the OLS models reported in this paper; they are available upon request.

In the second step, we examine the extent to which cross-country gaps in each of the three tests can be explained by those variables in the empirical model. One could, for example, first estimate an individual-level model with only country-level dummies to obtain cross-country gaps with reference to a particular country, and then re-estimate the equation by adding all independent variables to obtain the cross-country gaps after controlling for these variables. Based on the country-specific effects estimated by these two models, one could have a rough idea of how much of the cross-country gaps are explained by those variables. This approach, however, cannot identify factors that contribute most to the observed gaps in test scores.

To that end, we bring together the empirical model for test score determination with well-established decomposition methods in econometrics. Since Oaxaca's (1973) seminal work on decomposing male and female wage differentials, there have been literally hundreds of studies that use the decomposition method to better understand gaps (usually wage gaps) among groups. To illustrate how the decomposition method is used in this analysis, we will first describe the original Oaxaca (1973) method and its extensions as in Neumark (1988). Then we will modify the method to fit our empirical context of decomposing cross-country gaps in test scores.

Unlike methods that use a single equation, decomposition methods are usually based on a system of multiple equations (most often two or three). Oaxaca (1973) and Blinder (1973) used two-equation methods that estimated separate earnings equations for male and female groups. To illustrate, we will use two countries—country A and country B—as our groups of interest. In the first step, multiple regression analysis is used to estimate individual test scores in these two countries separately. For simplicity's sake, we represent all independent variables by a single matrix X.

$$T_{Ai} = X'_{Ai}\beta_A + \mu_{Ai} \tag{1}$$

$$T_{Bi} = X'_{Bi}\beta_B + \mu_{Bi} \tag{2}$$

The average achievement gap can then be expressed as:

$$\bar{T}_A - \bar{T}_B = \bar{X}'_A\hat{\beta}_A - \bar{X}'_B\hat{\beta}_B \tag{3}$$

where $\bar{T}_A$ and $\bar{T}_B$ are the average test scores in country A and B. $\hat{\beta}_A$ and $\hat{\beta}_B$ are estimated coefficients from Eqs. (1) and (2). Following Oaxaca (1973), we decompose the achievement gap into two parts:

$$\bar{T}_A - \bar{T}_B = (\bar{X}_A - \bar{X}_B)'\hat{\beta}_B + \bar{X}'_A(\hat{\beta}_A - \hat{\beta}_B) \tag{4}$$

This decomposition uses the country B as the reference. That is, the differences in independent variables between country A and B are weighed by the coefficient estimated from country B. Similarly, the coefficients from country A can also be used as the reference. Later, researchers (Cotton 1988; Neumark 1988; Reimers 1983) have proposed different weighted averages of the coefficients from different groups as the benchmark. For example, in studying male and female wage differentials, Neumark (1988) uses the male and female pooled estimates as the benchmark for decomposition based on the premise that the pooled model gives the prevailing market price.

To apply the decomposition method in the analysis of cross-country gaps in test scores among OECD countries, one has to first decide benchmark estimates. Cross-country gaps can be calculated in different ways. One may construct gaps between any two countries or between any one country and a carefully selected benchmark country. Because in most cases, policymakers are interested in knowing how their country fairs relative to other countries, it is probably appropriate to construct the achievement gap between each of the OECD country and the OECD average. One might wonder whether to include the country of interest in the calculation of the OECD average. In other words, is it better to compare country "A" with the rest of OECD countries with the exclusion of country "A" or to compare it with all OECD countries including country "A"? This is usually not an issue in the decomposition literature because in most cases there are only two groups (e.g., men vs. women) involved. We prefer including all OECD countries in the calculation in this study because it provides a common benchmark for each member country in OECD. Conceptually, one might view OECD as one group instead of a collection of 30 countries. For similar reasons, we also chose to use pooled estimates based on all OECD countries as the benchmark estimates. To illustrate, the gap in test scores between country "A" and the OECD average can be decomposed as follows:

$$\bar{T}_A - \bar{T}_{\mathrm{OECD}} = (\bar{X}_A - \bar{X}_{\mathrm{OECD}})'\hat{\beta}_{\mathrm{OECD}} + \bar{X}_A'(\hat{\beta}_A - \hat{\beta}_{\mathrm{OECD}}) \tag{5}$$

The left-hand side of the equation represents the observed average gap in student academic performance between country "A" and the OECD as a whole. The first term on the right-hand side is usually called "explained portion" in the literature because it is explained by the differences in the averages of the independent variables (e.g., individual characteristics, study time and activities, family background, and school characteristics) between country "A" and OECD. The second term is usually called the "unexplained portion" or "discrimination" in studies on wage differentials by gender; in this study, we call it "country-specific portion" because it is due to differences in model intercepts and regression coefficients.

Results

A student-level model of academic achievement

The last three columns of Table 2 report results from our fully specified models of students' math test scores. Regression models for science and reading test scores show very similar patterns. Results for these two test scores are not reported here due to space limitations but are available upon request. Since all students in OECD countries are viewed as one group in this analysis, each model is estimated using 5 plausible values of corresponding test scores and then weighted by the student final weight provided in PISA. We also run a stepwise regression where each of the four blocks of variables enters into the model separately. This exercise suggests that each block of variables contributes significantly in explaining individual differences in math test scores.[1] Taken as a whole, our regression model performs quite well, explaining slightly less than 50% of the total variance in student-level test scores. Adding in country fixed effects would increase the coefficient of determination to slightly over 50%; however, adding these fixed effects would defeat the purpose of our analysis because we are interested in using individual differences to explain cross-country gaps. Finally, we also report standardized beta coefficients for the effect of our independent variables in math test scores. These standardized coefficients may be more useful than the regression coefficients in comparing the effects across independent variables measured in different units (Goldberger and Duncan 1973).

Results in Table 2 are generally consistent with earlier research on the determinants of student academic performance. For example, both grade and age are positively related to math test scores. Being one grade level higher (e.g., 10th grade when compared with 9th grade) is associated with about a 13-point increase in math scores. Age is also positively related to test scores, which is an interesting finding given that all students in PISA are 15-year-olds. It suggests that students who start their schooling at an older age due to the schooling cycle have certain advantages. Boys perform significantly better in mathematics. Additional analyses also reveal that boys perform better in science, but not as well in reading. Immigrant students appear to have lower test scores. Even after controlling for immigration status, students who do not speak the test language perform significantly worse, especially for those who speak foreign languages. Finally, students with higher occupational aspirations perform significantly better. Although the estimated coefficients are relatively small, these actually represent large impact considering that the scale of the ISEI is from 16 to 90. The standardized beta coefficient suggests that one standard deviation increase (which is about 18 points in ISEI) in occupational aspiration is associated with 0.208 standard deviation increase in math test scores, or approximately 20 points given the standardized deviation of math test scores in our final sample is about 95 points.

[1] Results from these stepwise regressions are available upon request.

Our regression models also yield interesting results regarding student time use. Students who spend more time in school have higher test scores. Although the effect is much smaller than time spent directly on these three subjects, school time spent on other subjects also has a positive impact on test scores of other subject areas.[2] The negative and significant coefficients of the square terms suggest diminishing marginal return of study time. The positive relationship between time spent in school and academic performance, however, may simply reflect student self-selection within and between schools. Within a school, high-performing students are more likely to take optional or extra classes. Students and their family also self-select schools based on their academic performance. It could be the case that high-performing students are more likely to attend schools that require more study time in school.

Students who spend more time in out-of-school classes, especially on language, appear to have lower test scores. This negative relationship, however, may also be a result of sorting. For example, low-performing students might need tutors to help them with class materials, leading to a negative relationship between the time spent in out-of-school classes and test scores across individuals. Unfortunately, given our cross-section data, we are unable to separate the within- and between-individual relationship. The relationship between time spent on homework and academic performance suffers from the same problem. One might expect that more time spent on homework can improve academic performance; however, low-performing students probably need more time to finish the same amount of homework than others, leading to a negative relationship between time-use and learning outcomes across individuals.

The relationship between family characteristics and student academic performance is consistent with findings in the literature. Both parental occupational status and education have positive relationships with student academic performance. The economic condition also matters significantly. Cultural possessions, educational resources, and especially the number of books a family has, all have positive impacts on student academic performance. Family wealth, however, does not seem to have a positive relationship with student academic performance. This might be a result of cross-country variations.

The final block of variables is school characteristics. Students in public schools on average have lower test scores. Similarly, students in schools with a high proportion of students repeating a grade show lower performances. Student performance has a concave relationship with class size. The quality of teachers also matters; the higher the proportion of teachers having certificates or BA degrees the better the student performance. However, the lack of teachers does not seem to have a detrimental effect on student learning. Finally, school resources, measured by the proportion of computers with internet connection, have a positive impact on student performance.

Decomposing country-level gaps

Table 3 reports decomposition results for math test scores.[3] Column 1, "total gap", presents the difference between the math score of each country and the OECD simple average. Again, it is worth noting that in calculating the OECD simple average, each country, is weighted equally regardless of the magnitude of their student population. The next two columns report our decomposition results based on Eq. (5). Columns 4–7 report the portion of explained gap that can be attributed to each of the four blocks of variables as described in this study, namely individual demographic, study time and activities, family background, and school characteristics.

Due to the large volume of data presented in this table, we will not go over each country and test individually; instead, we will provide three general observations. First, our empirical models explain a substantial proportion of cross-country gaps in test scores. For example, the variance of observed average math scores across OECD countries (i.e., the first column in Table 3) is 1,020.4, while the variance of unexplained achievement gap (i.e., the third column of Table 3) is 371.3, representing a 63.6% reduction in variance of the cross-country differences. Results for science and reading scores reveal a similar reduction in cross-country variances.

Second, although our empirical model explains a large proportion of cross-country difference, the reduction in the actual gap varies across countries. Results in Table 3 indicate that our model explains a large proportion of the achievement gap for some countries such as Canada, Japan, Czech Republic, Portugal, Turkey, and Mexico; while for others such as Finland, Netherland, and Italy, the explanatory power is moderate at best. Because the model estimation is based on the entire OECD sample, there is no guarantee that the country-specific portion is smaller than the observed difference. For example, Switzerland scored 32.0 points higher in math than the OECD average and after considering all factors in the model, the gap actually increases to 41.5 points.

Third, last four columns of Table 3 suggest that the observed gap across different countries could be due to very different factors. For example, while both Finland and Korea perform very well in math test, their advantage is due to quite different reasons. For Finland, it appears that

[2] Results for science and reading tests reveal similar patterns.

[3] Results for science and reading scores are not reported here and are available upon request.

Table 3 Decomposition of math test score gaps among OECD countries (Benchmark: all OECD nations)

	(1) Total gap	(2) Explained gap	(3) Unexplained gap	(4) Demographic	(5) Study time	(6) Family	(7) School
Finland	50.7	13.8	36.9	−12.2	18.3	6.5	1.2
Korea	49.8	25.2	24.6	8.8	−8.0	4.6	19.7
Netherland	33.0	6.2	26.8	−3.2	0.6	3.3	5.6
Switzerland	32.0	−9.5	41.5	−11.6	5.9	3.6	−7.3
Canada	29.3	17.1	12.2	3.9	2.2	10.1	0.9
Japan	25.4	29.8	−4.4	4.3	12.5	−2.6	15.6
New Zealand	24.3	34.3	−10.0	16.0	12.1	2.5	3.7
Belgium	22.7	−2.9	25.6	−1.1	5.0	0.8	−7.7
Australia	22.2	35.8	−13.6	6.1	6.0	6.3	17.4
Denmark	15.4	−8.6	24.0	−10.5	−4.0	2.9	2.9
Czech Republic	12.2	13.7	−1.5	−1.5	7.1	5.8	2.3
Iceland	7.9	35.2	−27.3	6.7	9.7	18.0	0.8
Austria	7.8	5.9	1.9	−4.9	8.4	4.6	−2.2
Germany	6.1	−2.9	9.0	−10.3	4.1	6.0	−2.7
Sweden	4.7	3.3	1.4	−9.7	0.2	9.0	3.8
Ireland	3.8	3.1	0.7	−1.1	−6.4	0.3	10.4
France	−2.1	0.3	−2.4	−2.7	5.7	−0.1	−2.5
United Kingdom	−2.2	20.4	−22.6	16.4	3.5	0.1	0.5
Poland	−2.2	−5.6	3.3	−6.9	0.8	−4.4	5.0
Slovak Republic	−5.6	−9.8	4.2	−0.4	−4.9	−1.3	−3.2
Hungary	−6.7	−8.7	2.0	−6.7	−12.8	2.8	8.0
Luxembourg	−7.7	−20.0	12.3	−9.0	−2.9	4.0	−12.1
Norway	−7.8	21.8	−29.6	4.4	2.0	12.5	2.9
Spain	−17.7	−12.0	−5.7	0.7	1.3	−3.4	−10.5
United States	−23.3	−3.8	−19.5	8.2	−9.8	−2.6	0.4
Portugal	−31.5	−26.9	−4.6	−2.2	−1.5	−13.2	−10.0
Italy	−36.0	−11.6	−24.4	2.8	−5.3	−1.7	−7.5
Greece	−38.5	−22.1	−16.4	6.3	−19.2	−6.9	−2.4
Turkey	−73.7	−52.5	−21.2	6.4	−11.3	−36.1	−11.6
Mexico	−92.0	−69.1	−22.9	3.0	−19.4	−31.4	−21.5

study time and activities are the main drivers, while for Korea school characteristics play an important role. In Turkey and Mexico, family characteristics including educational resources are the most important factors that explain the observed gap in student performance. Knowing which factors that contributes to the performance gap across countries has important policy implications.

The case of Asian countries: Japan and Korea

In this section, we take a closer look at two Asian countries, Korea and Japan, and provide a better discussion on our decomposition results. Again, we use Eq. (5) and compute the difference in each subject (i.e., mathematics, science, and reading) between Korea/Japan and the OECD averages. The difference is then multiplied by the regression estimates from the OECD model, yielding the contribution of each independent variable to the explained gap across countries. Results for Korea and Japan are reported in Table 4. Korea scores 49.8 points higher in math than the OECD average. Our empirical model explains 25.2 points of the total gap, and the other 24.6 points remains unexplained. In other words, the difference in independent variables between Korea and the OECD averages can explain 25.2 points of the gap. The unexplained portion could be due to factors that are not controlled in our model such as cultural differences among countries or due to the fact that the same independent variables have different effects on test scores across countries. In addition, the contribution from the four categories of variables is 8.8, −8.0, 4.6, and 19.7 points for demographic characteristics, study time and activities, family background, and school characteristics, respectively. Among demographic characteristics, the average

Table 4 Decomposition of mathematics, science, and reading test scores for Korea and Japan

	Korea			Japan		
	Mathematics	Science	Reading	Mathematics	Science	Reading
Achievement gap	49.8	22.2	64.2	25.4	31.4	6.2
Decomposition of achievement gap						
Explained gap	25.2	17.9	35.3	29.8	26.0	23.2
Unexplained gap	24.6	4.2	28.9	−4.4	5.4	−17.1
Important variables						
Demographic characteristics	8.8	11.0	13.6	4.3	5.5	4.3
Study time and activities	−8.0	−13.3	−5.7	12.5	10.6	9.4
Family background	4.6	5.4	5.3	−2.6	−2.0	−5.6
School characteristics	19.7	14.9	22.1	15.6	11.8	15.2

grade of Korea students is about 0.4 years higher than the OECD average that leads to approximately 5 points in math test scores. In addition, lower proportions of immigrants and students speaking languages other than the test language also work to the advantage of Korean students. Korean students on average also have higher occupational aspiration than others. Although on average, Korean students spend more time in school, potentially leading to higher test scores, this is offset by their time spent in out-of-school class especially in language, which is unfortunately converted into lower math test scores. As previously discussed, this result might be due to student sorting. Differences in family background do not seem to contribute to the observed achievement gap, although Korean families have a larger number of books than the OECD average, leading to higher test scores. Our results suggest that the main driver of the achievement gap between Korea and the OECD average is differences in school characteristics. Further analysis identifies several factors. Schools in Korea have better qualified teachers as measured by the proportion of teachers with certificates and BA degrees and better technology resources measured by the proportion of computers with internet connection. In addition, the average class size in Korea is closer to the optimal level suggested by our model. These factors are also the main drivers for the explained gaps in science and reading scores.

Japanese students on average score 25.4 points higher on math tests than the OECD average. The differences in independent variables between Japan and other countries can explain 29.8 points of the gap. In other words, if Japanese students had similar values in individual characteristics, study time and activities, family background, and school characteristics, they would have scored 4.4 points lower than the OECD average. The contribution from the four categories of variables is 4.3, 12.5, −2.6, and 15.6 points for demographic characteristics, study time and activities, family background, and school characteristics, respectively. The most striking difference from the results for Korea is that Japanese students spend more time in school on science and less time out-of-school in language, leading to higher test math scores. Further, Japanese students do not seem to have similar educational resources as their Korean counterparts. Finally, a similar set of variables including better qualified teachers and better school resources explain the achievement gap between Japan and the OECD average.

Discussion and conclusion

In this study, we use PISA 2006 data to compare student academic performance across OECD countries. We first establish an empirical model to explain the variation in academic performance across individuals, and then use the Oaxaca-Blinder decomposition method to decompose the achievement gap between each of the OECD countries and the OECD average. Results indicate that the explained portion of the achievement gap varies across countries. In some countries, our empirical models are able to account for almost all the achievement gap, while unexplained country-specific effects still dominate in other countries. Finally, we use two Asian countries, Japan and Korea, to demonstrate how to identify major factors that have contributed to the observed achievement gap across countries.

The methods and results of this analysis is of interest to researchers who study student achievement gaps across countries and policymakers who are enthusiastic about understanding and closing the achievement gaps of their own countries with other high-achieving countries. Besides reconfirming findings from previous studies regarding the determination of student academic performance, our regression analyses have also highlighted some variables that could be potentially affected by policy interventions. For example, time spent in school has a positive and significant association with student academic performance. Student occupational aspiration stands out to have an important effect on test scores. Policy interventions specifically designed to improve these factors are likely to produce better performance. One caveat of this type of policy analysis is the difficulty in establishing causality based on cross-sectional and observational data. Recent work in this area has used various experimental and quasi-experimental methods to provide better evidence for causal relationship (Schneider et al. 2007). Obviously, more work in this area is needed to establish the causal relationship between these policy and outcome variables.

We have also used decomposition methods to partition the observed achievement gap between any particular country and others into two portions—one indicating the differences derived from observed independent variables and another part that suggests differences created from unique characteristics of the country in question. Identifying the main factors that have led to the observed achievement gap is an important step in policy making. If the observed achievement gap is mainly due to differences in the observed student, family, and school characteristics, policymakers need to focus on ways to improve these characteristics in their own countries. Our analysis on Korea and Japan further illustrates how to identify factors that contribute most to the observed achievement. On the other hand, if the observed gap is mainly due to country-specific effects, public policy needs to focus on broader and underlying economic, social, and cultural differences.

References

Ammermüller, A. (2007). PISA: What makes the difference? Explaining the gap in test scores between Finland and Germany. *Empirical Economics, 33*, 263–287.

Anderson, J., Lin, H., Treagust, D., Ross, S. P., & Yore, L. (2007). Using large-scale assessment datasets for research in science and mathematics education: Programme for International Student Assessment (PISA). *International Journal of Science and Mathematics Education, 5*, 591–614.

Baker, D. P., Goesling, B., & Letendre, G. K. (2002). Socioeconomic status, school quality, and national economic development: A cross-national analysis of the "Heyneman-Loxley effect" on mathematics and science achievement. *Comparative Education Review, 46*(3), 291–312.

Baker, D. P., Lee, J., & Heyneman, S. P. (2003). Should America be more like them? Cross-national achievement and U.S. policy. *Brookings Papers on Education Policy, 6*, 309–338.

Bielefeldt, T. (2005). Computers and student learning: interpreting the multivariate analysis of PISA 2000. *Journal of Research on Technology in Education, 37*(4), 339–347.

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

Casey, M. B., Nuttall, R., & Pezaris, E. (2001). Spatial-mechanical reasoning skills versus mathematical self-confidence as mediators of gender differences on mathematics subtests using cross-national gender-based items. *Journal for Research in Mathematics Education, 32*(1), 28–57.

Cotton, J. (1988). On the decomposition of wage differentials. *The Review of Economics and Statistics, 70*(2), 236–243.

Fertig, M. (2003). *Who's to blame? The determinants of German students' achievement in the PISA 2000 study*. IZA discussion paper no. 739, IZA-Bonn.

Fertig, M., & Schmidt, C. (2002). *The role of background factors for reading literacy: Straight national scores in the PISA 2000 study*. IZA Discussion Paper No. 545, IZA-Bonn.

Fuchs, T., & Wößmann, L. (2007). What accounts for international differences in student performance? A re-examination using PISA data. *Empirical Economics, 32*, 433–464.

Ganzeboom, H., De Graaf, P., & Treiman, D. (1992). A standard international socio-economic index of occupational status. *Social Science Research, 21*, 1–56.

Goldberger, A., & Duncan, O. D. (1973). *Structural equation models in the social sciences*. New York: Seminar Press.

Hanna, G. (1989). Achievement of girls and boys in grade eight: results from twenty countries. *Educational Studies in Mathematics, 20*(2), 225–232.

Hernandez-Zavala, M., Patrinos, H, Sakellariou, C., & Shapiro, J. (2006). *Quality of schooling and quality of schools for indigenous students in Guatemala, Mexico and Peru*. Policy research working paper no. 3982. Washington, DC: The World Bank.

Knighton, T., & Bussiere, P. (2006). *Educational outcomes at age 19 associated with reading ability at age 15. Culture, Tourism and the Centre for Education Statistics Research Paper (Statistics Canada Catalogue no. 81–595-MIE2006043)*. Ottawa, ON: Statistics Canada.

Kotte, D., Lietz, P., & Lopez, M. (2005). Factors influencing reading achievement in Germany and Spain: Evidence from PISA 2000. *International Education Journal, 6*(1), 113–124.

Lynn, R., & Mikk, J. (2007). National differences in intelligence and educational attainment. *Intelligence, 35*, 115–121.

Lynn, R., & Mikk, J. (2009). National IQs predict educational attainment in math, reading, and science across 56 nations. *Intelligence, 37*, 305–310.

MacNab, D. (2000). Raising standards in mathematics education: Values, vision, and TIMSS. *Educational Studies in Mathematics, 42*(1), 61–80.

Marks, G. N. (2005). Accounting for immigrant non-immigrant differences in reading and mathematics in twenty countries. *Ethic and Racial Studies, 28*(5), 925–946.

McEwan, P., & Marshall, J. (2004). Why does academic achievement vary across countries? Evidence from Cuba and Mexico. *Education Economics, 12*(3), 205–217.

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

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

OECD. (2004). *Learning for tomorrow's world: First results from PISA 2003*. Paris: OECD.

OECD. (2009). *PISA data analysis manual: SPSS* (2nd ed.). Paris: OECD.

Reimers, C. (1983). Labor market discrimination against Hispanic and black men. *The Review of Economics and Statistics, 65*(4), 570–579.

Rinderman, H. (2007). The g-factor of international cognitive ability comparisons: the homogeneity of results in PISA, TIMSS, PIRLS, and IQ-tests across nations. *Europena Journal of Personality, 21*, 667–706.

Schneider, B., Carnoy, M., Kilpatrick, J., Schmidt, W., & Shavelson, R. (2007). *Estimating causal effects using experimental and observational designs*. Washington, DC: American Educational Research Association.

Schuab, M., & Baker, D. P. (1991). Solving the math problem: exploring mathematics achievement in Japanese and American middle grades. *American Journal of Education, 99*(4), 623–642.

Schütz, G., West, M. R., & Wößmann, L. (2007). *School accountability, autonomy, choice and the equity of student achievement: International evidence from PISA 2003*. OECD education working paper no. 12, EDU/WKP(2007)7. Paris: OECD.

Shin, J., Lee, H., & Kim, Y. (2009). Student and school factors affecting mathematics achievement: International comparisons between Korea, Japan and the USA. *School Psychology International, 30*, 520–537.

Tatsuoka, K., Corter, J., & Tatsuoka, C. (2004). Patterns of diagnosed mathematical content and process skills in TIMSS-R across a sample of 20 countries. *American Educational Research Journal, 41*(4), 901–926.

Wang, J., & Lin, E. (2005). Comparative studies on U.S. and Chinese mathematics learning and the implications for standards-based mathematics teaching reform. *Educational Researcher, 34*(5), 3–13.

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

Williams, T., Williams, K., Kastberg, D., & Jocelyn, L. (2005). Achievement and affect in OECD nations. *Oxford Review of Education, 31*(4), 517–545.

Wu, M., & Adams, R. (Eds.). (2002). *PISA 2000 technical report*. Paris: OECD.

Wößmann, L. (2000). *Schooling resources, educational institutions, and student performance: The international evidence*. Kiel working paper no. 983. Kiel: Kiel Institute of World Economics.