Sociology of Education
2017, Vol. 90(1) 64–88

DOI: 10.1177/0038040716683779
journals.sagepub.com/home/soe

Families, Schools, and Student Achievement Inequality: A Multilevel MIMIC Model Approach

Shu-Ling Tsai[1], Michael L. Smith[2], and Robert M. Hauser[3]

Abstract

This article examines inequality in different dimensions of student academic achievement (math, science, and reading) by family background and school context in three East Asian (Taiwan, Japan, and South Korea) and three Western (United States, Germany, and the Czech Republic) nations. Building on Hauser (2009), we develop a novel multiple-indicator multiple-cause (MIMIC) model with a two-level hierarchical linear modeling specification, which allows us to explicitly test whether the several academic achievement constructs respond similarly to variation in family background and variation among schools and countries. The two-level MIMIC model is specified in detail and applied to 2012 Programme for International Student Assessment data. The analysis reveals new empirical insights, such as substantive differences within countries in performance inequality by subject, particularly among East Asian countries. While the data do not support the view of a "virtuous" relationship between excellence and equity in education, nor do they lend strong support to a "vicious" relationship either.

Keywords

MIMIC model, educational inequality, academic performance, PISA, educational stratification

Sociologists often assume that the impact of family background and schools on student performance is generally the same for different dimensions of performance, such as math, science, and reading (e.g., Organization for Economic Cooperation and Development [OECD] 2007). Due to this assumption, as well as convenience and custom, it has become commonplace in sociological practice for researchers to select only one dimension of performance as the dependent variable and then draw conclusions about educational inequality beyond the domain of that one subject. There are two aspects to this assumption: the effect of family background and school factors is assumed to be the same across different subjects assessed in standardized tests, and those academic achievement constructs are assumed to respond similarly to family background and school factors across different countries.

If these assumptions are unfounded, the implications are profound: it means we have let subject-specific inequalities bias our understanding of

[1]Academia Sinica, Taipei, Taiwan
[2]CERGE-EI, a joint workplace of Charles University and the Economics Institute of the Czech Academy of Sciences, Prague, Czech Republic
[3]National Academies of Sciences, Engineering, and Medicine, Washington, DC, USA

Corresponding Author:
Michael L. Smith, CERGE-EI, Politickych veznu 7, 111 21 Prague, Czech Republic
E-mail: michael.smith@cerge-ei.cz

inequality in academic achievement in general; it means we have ignored cross-national variation in the size of family and school effects on different academic subjects assessed simultaneously; and it means we need to find a way to rectify our biases through better model specification. In this article, we address this status quo by testing for heterogeneity in the effect of families and schools on different dimensions of academic achievement across countries using the 2012 Programme for International Student Assessment (PISA) data of the OECD. We show that the above assumptions are unfounded, and we specify a multilevel multiple-indicator multiple-cause model—hereafter, the multilevel MIMIC model—that addresses these issues and has wide application in sociological research.

We proceed by first highlighting key trends and problems in the literature on the cross-national variation in family background and school effects on student achievement. We take a close look at whether sociologists have neglected possible variation in the impact of family background and school factors on different components of academic performance and, second, whether possible variation in the effects of components of family background on student achievement across countries has also been neglected. We develop hypotheses that revisit the relationship between excellence and equity in student performance. We specify the multilevel MIMIC model and explain its advantages over standard practice and report results of the application of that model to 2012 PISA data. Although we examine only six economically developed countries that vary greatly by their educational systems (the United States, Japan, Germany, Korea, Taiwan, and the Czech Republic), this is sufficient to demonstrate cross-national variation in family background and school factors across subjects. The inclusion of more countries would undoubtedly yield even greater cross-national variation across subjects, attesting to the relevance of the multilevel MIMIC model for the study of educational inequality.

PREVIOUS RESEARCH

Families and schools are two of the most important determinants of student educational outcomes. The family-effect literature has long acknowledged a positive association between parental socioeconomic status (SES) and students' academic performance (e.g., Hauser, Sewell, and Alwin 1976). Beyond parental occupation and education, recent research also estimates strong effects for books in the home (Evans et al. 2010), parenting practices (Roksa and Potter 2011), and parental investments in shadow education (Buchmann, Condron, and Roscigno 2010; Kuan 2011; S. Lee and Shouse 2011; Liu 2012). Building on DiMaggio (1982), who theorized that such mechanisms express status group membership, Lareau (2002) argues that high-SES parents communicate their status through a "concerted cultivation" style of parenting, whereas other social classes use other parenting styles. Through concerted cultivation, high-SES families project their higher educational and occupational expectations (Carolan and Wasserman 2015) and create stronger parent–teacher relationships (Lareau 1987), which cumulatively translate into higher academic performance for their children (Roksa and Potter 2011). Contrary to such theories of status maintenance, other sociologists contend that books in the home, educational resources, and other dimensions of "scholarly culture" imbue children with information, vocabulary, and skills that directly contribute to their academic performance, independent of social class (Evans, Kelley, and Sikora 2014).

In terms of school effects, recent studies point to a number of factors that account for variation in academic performance. First, differences in disciplinary environments between schools strongly affect academic performance in combined math and science test scores in nine countries participating in the Trends in International Mathematics and Science Study (TIMSS) 2003 survey (Arum and Velez 2012). Second, in terms of the role of formal tracking, Hanushek and Woessmann (2008) found that early tracking increased performance dispersion from fourth to eighth grades without raising mean performance, whereas some school systems without early tracking reduced performance dispersion during that time. Schools can also differ in their degree of homogenous or heterogeneous ability grouping at the classroom level. Huang (2009) measured classroom homogeneity in terms of the degree of total variance in academic performance accounted for by the between-classroom component of that variance, finding that homogeneous grouping, relative to heterogeneous grouping, increased performance inequality by benefiting the high achievers at the expense of the low achievers, thus having a neutral impact on mean performance at the school level.

Bringing together the analysis of family and school effects, cross-national comparative studies of performance inequality have closely examined the role of school systems on student performance, particularly in terms of the standardization of educational provisions and the stratification of educational opportunities (e.g., Allmendinger 1989; Buchmann and Park 2009; Grodsky, Warren, and Felts 2008; Jackson, Jonsson, and Rudolphi 2012; Kerckhoff 1995, 2001; Montt 2011; Shavit, Arum, and Gamoran 2007). For example, Brunello and Checchi (2007) find that the effect of family background on different indicators of educational attainment increases with the length of tracking across countries. Van de Werfhorst and Mijs's (2010) review of the literature also concludes that school-type differentiation magnifies performance inequality by family background, which was similarly confirmed in Le Donné's (2014) comparative analysis of European PISA results from 2000 to 2009. School types even affect how students assess their own performance, at least for mathematics (Mijs 2016). By contrast, the effect of family background on student performance is weaker in countries with standardized, central exit exams administrated by an agency external to schools, compared to countries without these exams (Bishop 1998; Park and Kyei 2011; Van de Werfhorst and Mijs 2010). Those findings are echoed by Montt (2011), who found that decreasing the variability in opportunities to learn (equivalently, more standardization) and more intense schooling within a school system may reduce achievement inequality in students' math test scores.

ASSUMPTION OF HOMOGENOUS EFFECTS

Despite the progress made in recent years in understanding student achievement inequality, two methodological problems persist in the literature. First, international comparative studies of student academic achievement typically examine the determinants of student performance in only one subject area at a time, an approach conventionally used by PISA analysts themselves. When examining "quality and equality in the performance of students and schools," the authors of the PISA 2006 analysis report claim that "the overall impact of home background on student performance tends to be similar for science, mathematics, and reading in PISA 2006," and thus they restrict their analysis to science only (OECD 2007:170). Three years later, the same methodological approach was used to examine the effect of school characteristics on only reading performance in the 2009 data (OECD 2010), and then, three years later, on only math performance in the 2012 data (OECD 2013). However, the authors never establish, within a single model, that performance in one academic area can be taken as an indicator of academic achievement as a whole.

Other sociologists also commit to the assumption of homogenous effects across subjects. Le Donné's (2014:339) results on the role of educational systems on PISA student performance in Europe is based largely on reading performance, with the assumption that "the social gradient of cognitive ability in reading should be highly comparable with the social gradient of cognitive ability in mathematics or in sciences." Although she tests for robustness by switching out reading for mathematics or science performance as the dependent variable in her hierarchical linear modeling (HLM) models, those results are not reported, and the approach cannot establish strict comparability in family and school effects within and across countries on the same student sample. In a similar vein, a number of cross-national studies of performance inequality examine each academic subject separately (Bol et al. 2014; Hanushek and Woessmann 2008; Marks 2006; Roksa and Potter 2011), even though these authors are clearly interested in the latent construct of academic achievement.

Of equal concern, sociologists using PISA data seem to select one assessed subject as a matter of convenience, often without justification, and do not report tests of whether their results would hold if other subjects were analyzed instead. For example, Montt's (2011) important analysis of teacher quality homogeneity and the absence of tracking on performance inequality examined only performance in mathematics, even though science and reading scores were available in the same data set. Would the inferred cross-national patterns be the same or different had all three subjects been analyzed? Scholars have legitimate reasons to focus on specific academic subjects, but the literature discussed here is concerned with educational inequality more generally, not with specific subjects. Thus, an overall composite measure of academic achievement based on PISA scores in all three subjects (math, science, and reading) would be more reliable than an analysis of only one measure of achievement and would also provide a more accurate view of the relative importance of between- and within-school variances across the three subjects (Hauser 2009).

In addition to the assumption of homogeneous effects across subjects, some cross-national studies of family and school effects (e.g., Bol et al. 2014; Woessmann 2007) aggregate family background factors into a single variable, which in turn masks how the effects of components of that variable vary across countries and subjects. For example, the widely used PISA "index of economic, social, and cultural status" (ESCS) does not include information on how its components weigh on the index differently in different countries and in relation to academic achievement, something of utmost policy relevance. The index also masks the specific role of each component variable in contributing to academic achievement within and across countries. Models that measure the contribution of these components simultaneously could address the problem, but they are not utilized.

For these reasons, we depart from the conventional HLM approach used in much of the literature and instead follow Hauser's (2009) recommendation to include all three standardized PISA test scores (math, science, and reading) as outcome variables within a multilevel MIMIC model. This approach enables us to address three unresolved questions. First, is student academic achievement one-dimensional in relation to families and schools? Second, do academic achievement constructs respond similarly to family background characteristics and school factors in each country? And third, how do these relationships vary across countries?

EQUITY AND EXCELLENCE

If current approaches to analyzing cross-national differences in family and school effects on academic performance are as problematic as we suggest, it means we must revisit the central policy question of the literature: the relationship between equity and excellence in student performance across educational systems. The gold standard of educational policy with respect to academic performance is to achieve two ideals at once: a high average level of academic performance ("excellence") as well as a relatively high degree of "equity" in educational outcomes. Recent analytic reports from OECD PISA suggest there is a "virtuous" excellence–equity relationship (OECD 2013), although such reports are based on a problematic measurement of family background (ESCS), ignore institutional differentiation, and consider only one dimension of performance. For example, the PISA 2009 analysis report (OECD 2010), which focused on reading, found a lower degree of performance inequality by socioeconomic background among high-performing countries, such as Finland, Japan, and Korea. But if we examine the role of family and school effects on academic performance across countries and all three subjects at the same time, does this virtuous efficiency–equality relationship still hold?

We hypothesize that after taking into account both family- and school-level effects, the relationship between excellence and equity in educational outcomes may not be as virtuous as the OECD suggests. The vision of complementarity between excellence and equity is a *morally* desirable policy goal, but it has been questioned on *empirical* grounds. Woessmann (2004, 2007) found that the association between different family background variables and mean performance is often negative in Europe and across German states. In addition, a number of scholars have noted in both TIMSS (Huang 2009) and PISA (Montt 2011; OECD 2013) data that Taiwanese students' levels of math achievement are near the top of the world ranking, yet Taiwan's levels of dispersion in math test scores are among the highest in the world. One reason for the lack of a virtuous excellence–equity relationship is the mediating role of school-level effects, such as tracking mechanisms, which *reinforce* performance inequality by sorting students with high-SES backgrounds to the best-performing tracks (Gamoran and Mare 1989).

By utilizing the MIMIC model, we can advance this debate by comparing family and school effects on different subjects within the same country. In the pursuit of achieving high levels of academic performance, families employ a variety of mechanisms that may lead to increased inequality: parents invest in educational resources at home or in shadow education, and they deploy strategies to place their children in schools or tracks believed to give them a comparative advantage. However, there is no reason to presume that family effects are uniform across subjects: depending on the cultural importance of given subjects or the perceived relevance of a subject—for example, the heightened importance of mathematics performance in university applications in some countries—families may choose to invest their resources in one subject more than another. Because such choices can be culturally or nationally idiosyncratic, there is likely cross-national variance in the order with which family and school effects influence performance in different subjects. By extension, the degree of complementarity or trade-off between equality and efficiency can vary within countries in terms of math, science, and reading performance.

Institutional differences in school context contribute to differentials in academic performance. A vocational school that dedicates less time to academic subjects, compared to a more academic high school, will likely have lower average performance in standardized achievement tests than the latter. The sociological literature suggests several stylized facts about these kinds of links: (1) countries where students are sorted into different types of school at an early age tend to have larger disparities in student achievement than do countries without tracking or with tracking only at later ages, (2) the effect of family background on student achievement is stronger in countries with early selection than in countries without early selection, and (3) the effect of family background on student achievement is weaker in countries with standardized, central exit exams administrated by an agency external to schools than in countries without these exams (Bishop 1998; Park and Kyei 2011; Van de Werfhorst and Mijs 2010). Because of the importance of school differentiation, early selection, and standardization on academic performance, we selected countries with substantial variation in these factors and incorporated school-type effects into the MIMIC model.

Our main hypothesis is that the size of family and school effects on performance in a given subject (indicated by its factor loading on latent academic achievement) will be larger in subjects where average academic performance is higher—indicated by the subject-specific intercept and "adjusted" by the parameters of the model, not the raw average country scores often reported in the media. Although unobservable, this relationship is likely due to the role of culturally specific priorities and investments in education that drive both family choices and mean performance. Because we examine only a small set of countries, we are cautious not to make any definitive conclusions about the relationship between equity and excellence in academic performance, although we do hope the approach can be replicated later when using a larger set of countries.

> **Hypothesis 1:** For most countries, the size of family and school effects on student achievement will be largest in the subject where mean student achievement is the highest (*excellence with inequality*).

Given the evidence in the literature reviewed earlier on the role of tracking mechanisms on educational inequality, we also hypothesize that student academic achievement should be influenced by structural forces to a larger extent in more diversified and stratified school systems, because in such systems lower-SES students are often channeled into vocational tracks or tracks limiting university entry. Moreover, vocational students tend to have lower scores in achievement tests than do academic students. Accordingly, we expect variance in academic achievement could be explained by *family background* characteristics to a larger extent in more diversified and stratified school systems; that is, the explained variance (percent) should be larger in more diversified and stratified school systems, like the Czech Republic and Germany, which we included in this study precisely to test this kind of hypothesis. Similarly, we expect to find that the variance in academic achievement can be explained by *school factors* to a larger extent in more diversified and stratified school systems. For example, the explained variance for Taiwan should be larger than that for South Korea, and the explained variance in Germany and the Czech Republic should be larger than that in the United States.

> **Hypothesis 2:** (a) Variance in student achievement explained by between-school factors should be larger than that explained by within-school factors in more diversified and stratified school systems; in addition, (b) the combined share of within- and between-school variance in latent student achievement explained by the model should be larger in such school systems (*stratification with inequality*).

COUNTRY SELECTION

To test our two hypotheses, we assess the extent to which family background characteristics and type of school attended explain the variation in academic achievement among students in six highly industrialized nations representing different types of educational systems: Japan, Korea, Taiwan, the Czech Republic, Germany, and the United States. On the one hand, we selected East Asian countries because their secondary education is highly standardized (Japan and Korea, in particular) and *performance based*; that is, selection into higher levels of education is based primarily on performance on high-stakes tests and, by extension, the role of family background on performance on those tests (Jackson 2013). The Japanese, Korean, and Taiwanese systems are embedded in a highly competitive educational setting in which cultural expectations for student achievement are high, and hence parents are willing to invest in shadow education to help their children do better on exams, especially in math (Huang 2013; Kuan 2011; S. Lee and Shouse 2011; Park, Byun, and Kim 2011). East Asian countries also have high levels of student achievement and low levels of dispersion compared to other regions of the world (Park 2010, 2013; Park and Sandefur 2006). These similarities may be rooted in those countries' common Confucian legacy, which values learning, hard work, and disciplined student behavior (Ishida and Miwa 2012; Park 2012).

East Asian educational systems, however, differ substantially from each other. In addition to academic and vocational high schools, Taiwan also offers comprehensive high schools, which provide both kinds of tracks. Korea and Japan have similar types and distributions of high schools, but Korea uses a unique equalization policy through which students are randomly assigned to local schools, whereas parents and children in Japan play a much larger role in school selection (Park 2013). Although shadow education is prevalent in all three countries, these investments are associated with high SES in Japan (Stevenson and Baker 1992), whereas in Taiwan, students' participation in shadow education does not reflect specific patterns of social stratification (Liu 2012).

Turning to Western countries, we selected Germany and the Czech Republic because their school systems are more diversified and vocationally specific: they sort students into different types of schools at an early age (10 or 11), have dead-end tracks in secondary education, and maintain strong traditions of promoting vocational training in addition to academic tracks. We also include the United States, which has a distinctive system of comprehensive secondary schools characterized by relatively little institutional differentiation but substantial within-school tracking (Gamoran 1987; Lucas 2001). All three Western countries have *choice-based* secondary educational systems: selection into higher levels of education is based primarily on the choice of school or track attended and, by extension, the role of family background on school or track choice. Performance and choice effects are present to some degree in every country and at each educational transition (Jackson et al. 2012), but analytically distinguishing school systems by the purportedly dominant mechanism of inequality transfer can be useful for interpreting results of analyses of the role of family background, schools, and institutional context on academic performance.

RESEARCH DESIGN

Research involving school effects typically involves HLM, in which students are nested in schools nested in countries (Lee and Bryk 1989; Raudenbush and Bryk 1986, 1992). Because the conventional HLM approach uses only one dimension of academic performance at a time, we develop a MIMIC model with two-level HLM specifications, which allows us to explicitly test whether the several academic achievement constructs, estimated simultaneously, respond similarly to variation in family background and variation among schools for each of the countries examined. To simplify the analysis, we do not consider a three-level model, as that would involve standardized indicators of school types across countries, which poses a problem if some countries have little or no variation in school types. However, by imposing equality constraints on parameters of interest (an approach feasible with a small number of countries, such as those examined here), we can test for whether between-country differences in coefficients are statistically significant and thus can compare across countries in ways similar to a three-level HLM.

The MIMIC model was first introduced into sociology by Hauser to study unobservable variables in path analysis as both causes and effects of observables (Hauser 1972; Hauser and Goldberger 1971); it was later elaborated within the LISREL framework for the study of social stratification

(Hauser and Wong 1989; Warren and Hauser 1997). An important feature of the MIMIC model is the specification of an unobservable variable (η), which is determined by multiple exogenous variables (X) and determines multiple outcome variables (Y_c). In general, a MIMIC model is estimated on one-level (individual) data and can be expressed as

$$\eta = \Sigma\beta_k X_k + \xi,$$
$$\text{and } Y_c = \lambda_c \eta + \varepsilon_c,$$

where Y_c is a vector of observable indicators of the unobservable variable η; X_k is a vector of observable determinants of that η; λ_c represents factor loadings; and ξ and ε_c are mutually uncorrelated error terms, which means COV (ζX) = 0, COV ($\eta\varepsilon_c$) = 0 for all c, and COV ($\zeta\varepsilon_c$) = 0 for all c. We also have two possible error variance-covariance specifications: COV ($\varepsilon_c\varepsilon_{c'}$) = 0 for $c \neq c'$, and VAR (ζ) = 0.

Compared to a conventional structural equation model, the MIMIC model presents a more parsimonious specification, in which proportionality constraints are placed on the effects of X variables on outcome variables (Y). The MIMIC model is overidentified and incorporates two kinds of restrictions: (1) the regression coefficient matrix has rank one, restricting the reduced-form coefficients relating Y_c to X_k—namely, b_{kc}—such that they are combinations of structural coefficients; and (2) the residual variance-covariance matrix satisfies a confirmatory factor analysis model with one common factor, restricting the covariance matrixes such that they are built up from factor loadings, factor variance, and unique variances (Hauser and Goldberger 1971). Here, the first econometric restriction matters more than the second factor-analytic restriction, because the force of the MIMIC model is to place constraints on the entire matrix of coefficients estimated in a single population. That is, if the model is true, then in the population

$$b_{kc} = \beta_{kc}\lambda_c,$$

there are proportionality constraints on the reduced form coefficients:

$$b_{kc}/b_{k'c} = \beta_{kc}/\beta_{k'c} (\text{for } k \neq k'),$$
$$\textit{and } b_{kc}/b_{kc'} = \lambda_c/\lambda_{c'} (\text{for } c \neq c').$$

Specification of Two-level MIMIC Model

In algebraic form, the two-level MIMIC model of student academic achievement consists of two levels and two equations in each level, which can be expressed as follows:

Level 1: Within-school Model

$$\eta_{ij} = \Sigma\beta_{kj} X_{kij} + \zeta_{ij}, \quad (1)$$

$$Y_{cij} = \lambda_{cj}\eta_{ij} + \varepsilon_{cij}, \quad (2)$$

where η_{ij} is a latent variable representing general academic achievement for student i in school j, which is influenced by the student's family background characteristics (X_{kij}) and indicated by the student's test scores in math, science, and reading (Y_{cij}, c = 1, 2, 3); β_{kj} are the coefficients of background variables among students in school j; λ_{cj} are the student-level factor loadings for school j; and ε_{cij} and ξ_{ij} are error terms for student i in school j in the two within-school equations.

Level 2: Between-school Model

$$\eta_j = \Sigma\gamma_p Z_{pj} + \zeta_j, \quad (3)$$

$$Y_{cj} = \lambda_{c0j} + \lambda_c\eta_j + \upsilon_{cj}, \quad (4)$$

where the four dependent variables (η_j and Y_{cj}, c = 1, 2, 3) are all unobservable variables; η_j represents general academic achievement for school j, which is influenced by school factors (Z_{pj}) and indicated by the average test scores in math, science, and reading for school j (Y_{cj}), which are actually the latent intercepts estimated in Equation 2; γ_p are the coefficients of p school factors; λ_c are the school-level factor loadings; and ξ_j and υ_{cj} are error terms for school j in the two between-school equations.

Three notes are important here. First, similar to the HLM specifications reported in Arum and Velez (2012:12), at the between-school level, the coefficients of the student background variables are assumed not to vary across schools:

$$\beta_{kj} = \beta_k. \quad (5)$$

Second, the regression of the within-school factors on X variables is assumed to have a random intercept varying across schools. Accordingly, the factor loadings are constrained to be equal across the within and the between levels:

$$\lambda_{cj} = \lambda_c. \tag{6}$$

And third, within the MIMIC model, one outcome variable is set as a reference index (i.e., $\lambda_c = 1$ if $c = 1$) to identify the relative factor loadings of other outcome variables (i.e., $\lambda_c \neq 1$ if $c \neq 1$). In this analysis, math test scores are set as the reference index at both student and school levels. We also test for the possibility of homogeneity in factor loadings across subjects, that is, $\lambda_{cj} = \lambda_c = 1$ for all c ($c = 1, 2, 3$), namely, homogeneous family effects across subjects within countries.

Variations of the Two-level MIMIC Model

To deal with error variances and covariances, we consider and estimate (via MPlus 7) two versions of the two-level MIMIC Model: one with covariances among outcome variables but no disturbances in latent factors (η) and another, vice versa, with disturbances in the latent factors (η) but without covariances among outcome variables. The two models—Model A and Model B—are depicted in Figures 1 and 2, respectively. In the two figures, the rectangles represent observed variables, including background variables (X) and outcome variables (Y) at the student level and school factors (Z) at the school level. The circles represent latent variables. The arrows represent regression relationships between variables, and the curved, double-headed arrows represent covariances between variables. In the within-school model, the filled small circles at the end of the arrows from the within-school factor η_{ij} to Y_{1ij}, Y_{2ij}, and Y_{3ij} represent random intercepts, which are referred to as Y_{1j}, Y_{2j}, and Y_{3j} in the between-school model. These intercepts, varying across schools, are indicators of the between-school factor η_j.

Model A in Figure 1 allows indicator disturbances (i.e., error terms in Y) to be freely correlated at both student and school levels, because one might expect positive correlations between student test scores in the three subjects. In accordance with this expectation, the unobservable variable (η) is specified as an exact function of its causes at each level, absorbing the disturbance ξ into the ε's or υ's, and thus $\xi = 0$ at both levels. In contrast, Model B in Figure 2 allows the indicator disturbances to be mutually independent and thus specifies $\xi \neq 0$ at both levels. The basic idea of Model B is that at each level, η is a single latent variable that accounts for the covariance of the Y variables. Once the effects of the X and Z variables via η are removed, there should no longer remain any correlation among the Y variables. It is important to note that the reduced-form disturbances under Model B are not independent of each other, since they all have the disturbance ξ in common, whereas the correlations of the reduced-form disturbances under Model A are not patterned this way, since $\xi = 0$. In a nutshell, Model B is stronger and more restrictive, using four fewer parameters than Model A. Which MIMIC model fits better in which country is an empirical question.

DATA, VARIABLES, AND DESCRIPTIVE STATISTICS

Data for this study are derived from the 2012 PISA survey, which is an internationally standardized assessment that was jointly developed by some 60 participating countries and administered to students age 15 to 16 years at the time of the assessment. The PISA survey also contains a student background questionnaire and a school questionnaire completed by principals, with information on school context. In this study, we use only the family and school variables in the 2012 PISA data commonly available for the six countries examined. The sample is double weighted, by student and school sampling probabilities. Because almost all Korean and Japanese students (and most U.S. students) participating in the 2012 PISA were in grade 10, we restrict the analysis sample to 10th graders and include only students with complete information on family background characteristics and school variables of concern. That is, we exclude students with missing data on the variables of concern from the analysis in a listwise fashion. The share of the sample omitted due to missing data is modest for most countries (10.1 percent in Korea, 14.3 percent in Taiwan, 17.5 percent in the United States, 18.8 percent in Japan, and 21.1 percent in the Czech Republic) but reaches 39.8 percent in Germany, which is mainly due to students' nonreporting of parental

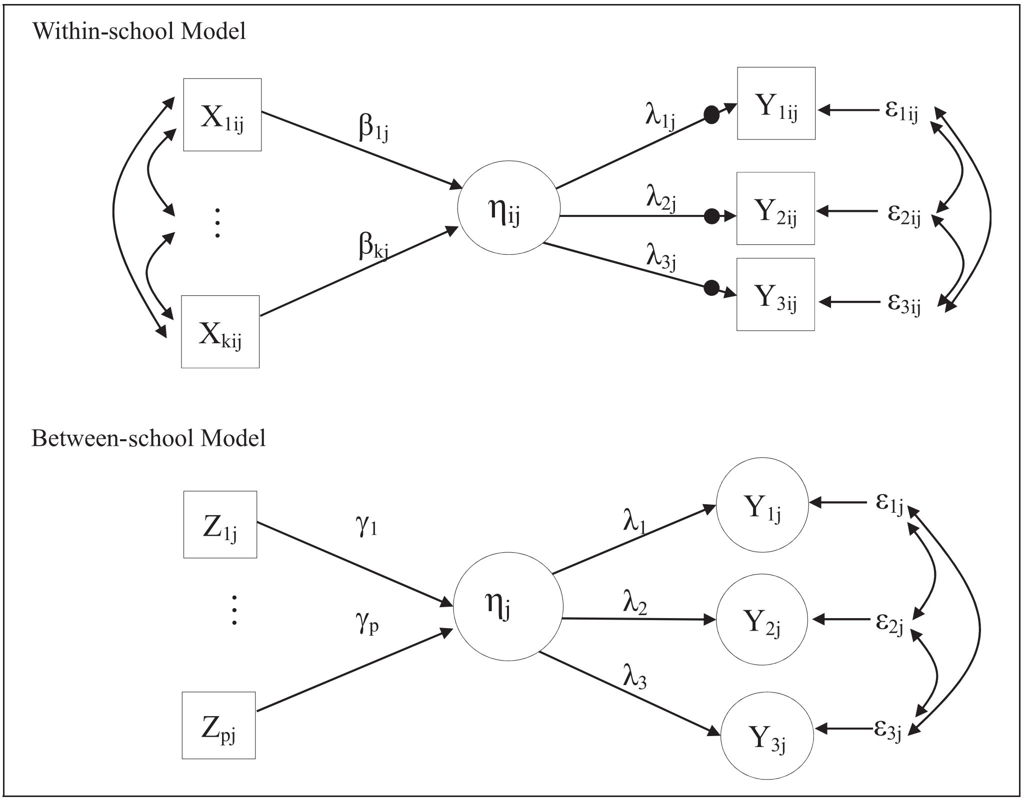

Figure 1. Diagrams for Model A: two-level multiple-indicator multiple-cause model with covariances among outcome variables (Y), no disturbances in latent variables (η).
Note: The curved, double-headed arrows represent covariances between variables. In the within-school model, the filled small circles at the end of the arrows from the within-school factor η_{ij} to Y_{1ij}, Y_{2ij}, and Y_{3ij} represent random intercepts, which are referred to as Y_{1j}, Y_{2j}, and Y_{3j} in the between-school model, and are indicators of the between-school factor η_j.

occupation or education. We consider the potential issue of selectivity bias due to missing data a serious limitation of the current study. Whether and how missing data on parental background in PISA data bias estimates of educational inequality is an important topic for future research, which we cannot adequately address here.

This analysis uses three kinds of observed variables. First, our outcome variables (Y) are the means of five plausible values in student test scores in math, science, and reading in PISA 2012. These plausible values were scaled to have an average score of 500 and a standard deviation of 100 across all students of OECD countries participating in PISA. Ideally, all five plausible values should be used simultaneously to obtain the estimates of population parameters, but we were not able to program this into our model specification. By taking their mean, a practice common in the literature (e.g., Buchmann and Park 2009), we lose information on the dispersion across plausible values and thus run the risk of understating standard errors in our results. We do not think the regression estimates or factor loadings are affected by this choice.

Second, based on the literature reviewed earlier, variables related to family background (X) include the following: (1) father's years of schooling; (2) mother's years of schooling; (3) the highest International Socio-economic Index of Occupational Status (ISEI) of either parent; (4) number of books at home, with response categories 1 = 0 to 10 books,

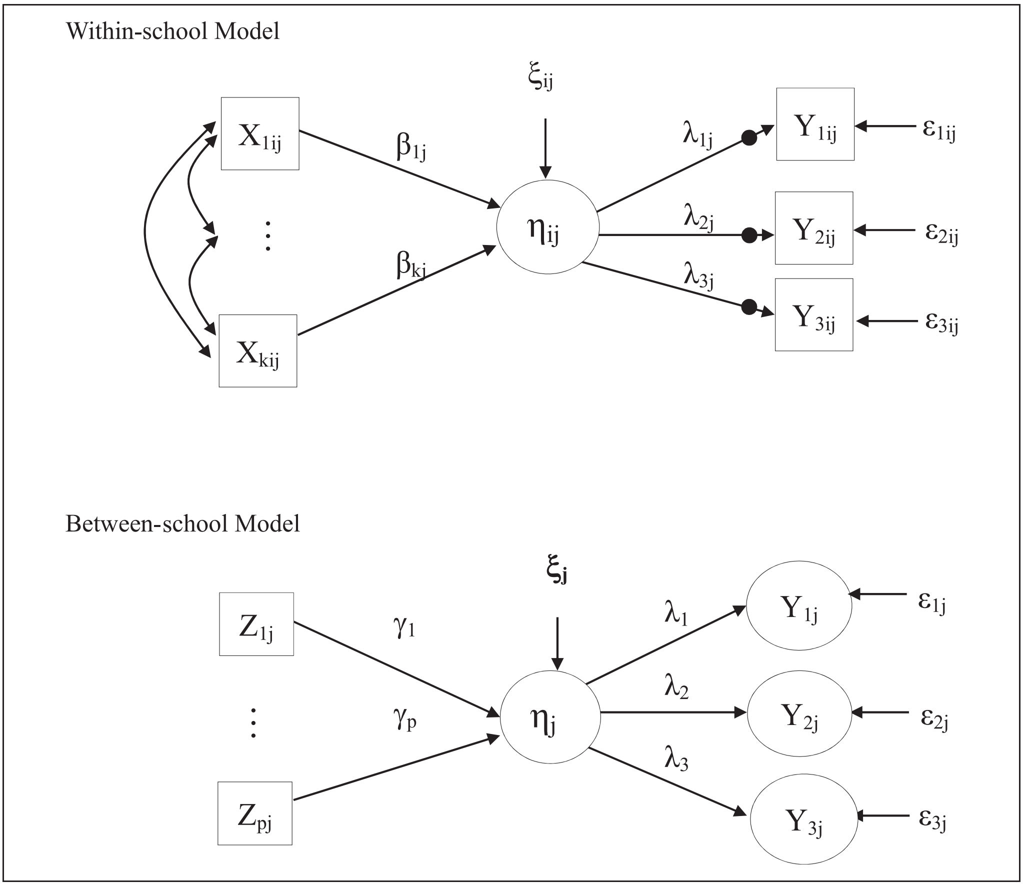

Figure 2. Diagrams for Model B: two-level multiple-indicator multiple-cause model with disturbances in latent variables (η), no covariances among outcome variables (*Y*).

Note: In the within-school model, the filled small circles at the end of the arrows from the within-school factor η_{ij} to Y_{1ij}, Y_{2ij}, and Y_{3ij} represent random intercepts, which are referred to as Y_{1j}, Y_{2j}, and Y_{3j} in the between-school model, and are indicators of the between-school factor η_j.

2 = 11 to 25 books, 3 = 26 to 100 books, 4 = 101 to 200 books, 5 = 201 to 500 books, and 6 = more than 500 books; and (5) home educational resources, with items indicating the total sum of whether a student has a desk to study at, a quiet place to study, a computer to use for schoolwork, educational software, textbooks to help with schoolwork, technical reference books, and a dictionary.

Third, school factors (*Z*) include two dummy variables indicating whether the school attended is private or public and whether the school is located in a rural (village, small town, or town) or urban (city or large city) area; we include other dummy variables indicating country-specific distinctive types of schools in upper-secondary education. We consider the following school types: one type for the United States (comprehensive schools), two types for Japan and Korea (academic/general and vocational high schools), and three types for Taiwan (senior high schools, senior vocational schools, and comprehensive high schools, in which both academic and vocational tracks are available), the Czech Republic (*Gymnasium*, technical schools with the school-leaving exam, and vocational schools without the school-leaving exam, including special schools), and Germany (*Gymnasium*, comprehensive schools [*Realschule*], and vocational schools [*Hauptschule*], including basic schools). Although the terminology of “comprehensive school” is officially used in the Taiwanese, German, and U.S. systems, its meaning and content differs across these countries. Use of country-specific dummy variables departs from strict comparability across countries (i.e., from a three-level MIMIC model analysis), but they are necessary to take into account important school factors on performance in different subjects within countries.

Table 1 reports descriptive statistics of variables used in this analysis, weighted by the student and school sampling probabilities provided in the PISA data, along with the sample size of students and of schools. As the table shows, Taiwan is a high achiever in math but has an extremely high dispersion in math achievement. The Czech Republic and Germany are well known for having highly differentiated upper-secondary educational tracks, but their dispersion in academic performance in the three subjects does not appear to be materially different than in countries with more standardization. Germany is also the highest performer in science and reading.

Table 1 also reveals major differences in explanatory variables between countries. In the sample, 51.9 percent of Korean students and 44.2 percent of Taiwanese students study at private secondary schools, which is several times more than in the Western countries. Similarly, the great majority of students in the East Asian countries study in urban settings, whereas the trend is the opposite in Western countries. Besides the United States, Korea and Japan have the greatest share of secondary-school pupils in academic or general high schools; only 30.2 percent of Czech pupils and 28.9 percent of German pupils study in *Gymnasia*. In terms of student-level variables, Germany has particularly high averages in all indicators of family background, although we do not see any substantial differences in these distributions between countries.

RESULTS

Analysis of Model Fit

Whether the MIMIC model offers a superior approach to analyzing student achievement inequality compared to the subject-by-subject approach in OECD PISA analyses depends on the validity of the proportionality constraints in the MIMIC model. To test for this, we estimate a variety of models, including Models A and B depicted in Figures 1 and 2. To evaluate model fit, we use the sample size–adjusted Bayesian information criterion (BIC) statistic; the model with a lower BIC provides a better fit to the data. To test for the statistical significance of the model contrast—in case of a small BIC contrast—we also carried out a chi-square test for nested models with the MLR estimators, using the TRd test statistic (University of California–Los Angeles 2013).

Table 2 describes the fit of six models for the six countries under study. Model 1 is a conventional null baseline in which no parameters are fitted except the means and variances for the three subjects at two levels, yielding nine free parameters. Model 2 continues to allow for intercorrelations between background variables, using another 20 free parameters. Next, we estimate two different MIMIC model specifications in a parallel fashion, one with covariances among outcome variables (Panel II) and the other with free variances in latent factors (Panel III). That is, Model 3 adds to Model 2 the intercorrelations between student test scores in the three subjects at two levels, yielding six free parameters in total, whereas Model 5 adds to Model 2 the variance for the latent factor (ξ) at two levels and thus fits two more free parameters than Model 2. Model 3 is used as the preferred baseline of Model A (i.e., Model 4 in the table), because it is based on the specification of Model A, but it fixes all factor loadings (λ) at one and all structural coefficients—that is, β in Equation 1 and γ in Equation 3 —at zero, just like the two null models in Panel I. By the same logic, Model 5 serves as the preferred baseline of Model B (i.e., Model 6 in the table), because Model B allows parameters of factor loadings that are fixed at one in Model 5—except math (the reference index)—and parameters of structural coefficients that are fixed at zero in Model 5 to be free. The difference between Model 3 and Model 4 (= Model A), or between Model 5 and Model 6 (= Model B), is that family and school effects are allowed to be free in Models A and B.

We compare the fit of Model A and Model 3, finding a large decrease in BIC for the six countries examined in Table 2 (see Panel IV). Similarly, the contrasts in BIC of Model B and Model 5 are all negative. The significance test on TRd also indicates a substantial improvement in fit for every country when moving from Model 3 to Model A or moving from Model 5 to Model B.

Table 1. Double Weighted Descriptive Statistics, by Country.

	East Asian countries			Western countries		
Variable	Japan	Korea	Taiwan	Czech Republic	Germany	United States
PISA 2012 test scores						
Math	522.670	557.632	580.447	522.623	561.841	494.302
	(90.322)	(98.446)	(114.046)	(87.154)	(80.886)	(77.412)
Science	537.386	540.428	535.021	526.750	567.663	513.281
	(90.489)	(79.100)	(82.150)	(79.726)	(77.239)	(80.423)
Reading	528.651	538.570	537.536	515.483	549.576	511.896
	(93.280)	(81.988)	(87.019)	(80.691)	(76.241)	(78.788)
Student-level variables						
Father's schooling	13.420	13.522	12.469	12.937	13.813	13.038
	(2.194)	(2.638)	(2.783)	(1.974)	(3.572)	(2.835)
Mother's schooling	13.221	13.210	12.388	13.009	13.276	13.465
	(1.800)	(2.416)	(2.717)	(1.939)	(3.242)	(2.662)
Highest ISEI of either parent	48.598	52.876	48.100	48.822	53.458	55.670
	(20.076)	(18.940)	(19.930)	(18.935)	(20.566)	(20.235)
Number of books in home	3.378	3.921	3.299	3.423	3.809	3.085
	(1.371)	(1.332)	(1.483)	(1.341)	(1.371)	(1.428)
Educational resources in home	5.022	5.627	5.426	6.330	6.150	5.563
	(1.234)	(1.285)	(1.418)	(0.869)	(0.946)	(1.553)
School-level variables						
School sector (%)						
Public	75.2	48.1	55.8	77.7	94.1	80.9
Private	24.8	51.9	44.2	22.3	5.9	19.1
School location (%)						
Urban	63.6	69.9	59.7	22.3	22.7	17.7
Rural	36.4	30.1	40.3	77.7	77.3	82.3
School type (%)						
Academic schools	64.6	76.7	44.5	—	—	—
Vocational schools	35.4	23.3	27.4	—	63.0	—
Comprehensive	—	—	28.1	—	8.1	—
Gymnasium	—	—	—	30.2	28.9	—
Technical schools with the school-leaving exam	—	—	—	46.2	—	—
Vocational schools without the school-leaving exam	—	—	—	23.6	—	—
Sample size						
Students	5,155	4,250	3,473	1,883	1,101	2,998
Schools	190	140	113	132	169	152

Note: Numbers in parentheses are standard deviations. Data are weighted by both student and school sampling probabilities. PISA = Programme for International Student Assessment; ISEI = International Socio-economic Index of Occupational Status.

These results indicate that the proportionality constraints in the MIMIC model are valid. However, this raises an important question: which type of MIMIC model fits better in which country? Inspection of the last three rows in Table 2 indicates that in terms of BIC, Model A fits better than Model B for Japan, Korea, Czech Republic, and Germany. Conversely, Model B fits better than Model A for Taiwan and the United States, yet the BIC contrasts are very small (5.936 for the former; 6.037 for the latter). The chi-square significance test reveals that the corresponding TRd statistic for the United States, with four degrees of freedom, is not statistically significant, in which case Model B is preferred over Model A, consistent with the BIC result. However, the TRd

Table 2. Model Fit Statistics, by Country.

	East Asian countries			Western countries		
Model and contrast	Japan	Korea	Taiwan	Czech Republic	Germany	United States
I. Null baseline						
Model 1: Fit the means and the variances at two levels						
Sample size–adjusted BIC	174478.599	143549.806	115251.920	62012.653	36891.852	102818.167
Loglikelihood H0 value	−87215.134	−71751.606	−57603.571	−30986.690	−18428.701	−51387.356
H0 scaling correction factor for MLR	2.7745	2.4657	2.0596	1.1420	1.8234	2.9974
(Number of free parameters) [*df*]	(9) [30]	(9) [30]	(9) [33]	(9) [33]	(9) [33]	(9) [27]
Model 2 = (Model 1 + intercorrelations between background variables at first level						
Sample size–adjusted BIC	295073.354	244401.568	200570.197	104125.936	64099.371	177451.156
Loglikelihood H0 value	−147458.811	−122125.716	−100212.957	−51999.695	−31994.183	−88655.568
H0 scaling correction factor for MLR	2.8298	4.5704	2.8046	2.1619	1.7868	3.6301
(Number of free parameters) [*df*]	(29) [30]	(29) [30]	(29) [33]	(29) [33]	(29) [33]	(29) [27]
II. Models with covariances between outcome variables						
Model 3 = Model 2 + covariances between outcome variables at two levels						
Sample size–adjusted BIC	278572.866	230499.250	188668.147	99089.030	61038.684	164965.279
Loglikelihood H0 value	−139192.457	−115159.026	−94247.006	−49468.151	−30452.356	−82398.144
H0 scaling correction factor for MLR	2.5526	4.1586	2.4434	2.0405	1.8411	3.4546
(Number of free parameters) [*df*]	(35) [24]	(35) [24]	(35) [27]	(35) [27]	(35) [27]	(35) [21]
Model 4 = Model A = Model 3 + family and school effects						
Sample size–adjusted BIC	278407.271	230246.746	188308.469	98767.185	60804.623	164599.223
Loglikelihood H0 value	−139082.809	−115006.889	−94039.803	−49283.229	−30314.274	−82193.389
H0 scaling correction factor for MLR	2.4456	3.7523	2.2036	1.9029	1.7623	3.3020
(Number of free parameters) [*df*]	(45) [14]	(45) [14]	(46) [16]	(46) [16]	(46) [16]	(44) [12]
III. Models with disturbances in latent variables (η)						
(Model 5) = (Model 2) + free variance for latent factor at two levels						
III. Models with disturbances in latent variables (η)						
Model 5 = Model 2 + free variance for latent factor at two levels						
Sample size–adjusted BIC	278899.933	230872.905	189685.287	99247.707	61171.187	165187.431
Loglikelihood H0 value	−139366.731	−115356.208	−94765.527	−49556.217	−30526.264	−82518.877
H0 scaling correction factor for MLR	2.6278	4.4045	2.5860	2.1211	1.6978	3.5545
(Number of free parameters) [*df*]	(31) [28]	(31) [28]	(31) [31]	(31) [31]	(31) [31]	(31) [25]

(continued)

Table 2.
(continued)

Model and contrast	East Asian countries			Western countries		
	Japan	Korea	Taiwan	Czech Republic	Germany	United States
Model 6 = Model B = Model 5 + family and school effects						
Sample size–adjusted BIC	278504.410	230282.380	188302.533	98794.288	60840.186	164593.186
Loglikelihood H0 value	−139142.119	−115035.059	−94046.785	−49305.508	−30339.711	−82200.027
H0 scaling correction factor for MLR	2.5036	3.9142	2.2955	1.9437	1.7115	3.3723
(Number of free parameters) [*df*]	(41) [18]	(41) [18]	(42) [20]	(42) [20]	(42) [20]	(40) [16]
IV. Contrasts						
Model A vs. Model 3: BIC	−165.595	−252.504	−359.678	−321.845	−234.061	−366.056
TRd	105.884***	130.576***	287.662***	252.439***	182.700***	151.191***
[*df*]	[10]	[10]	[11]	[11]	[11]	[9]
Model B vs. Model 5: BIC	−395.523	−590.525	−1382.754	−453.419	−331.001	−594.245
TRd	212.040***	268.265***	973.365***	347.301***	213.190***	232.337***
[*df*]	[10]	[10]	[11]	[11]	[11]	[9]
Model A vs. Model B: BIC	−97.139	−35.634	5.936	−27.103	−35.563	6.037
TRd	64.081***	26.920***	11.274*	30.219***	22.160***	5.108
[*df*]	[4]	[4]	[4]	[4]	[4]	[4]

Note: TRd statistic computed according to University of California–Los Angeles (2013). BIC = Bayesian information criterion.
*$p < .05$. **$p < .01$. ***$p < .001$.

Table 3. Estimates of Background Effects (β), by Country.

	East Asian countries			Western countries		
Variable	Japan	Korea	Taiwan	Czech Republic	Germany	United States
Model A						
Father's schooling	−0.047	0.894	1.352	−0.581	−0.283	0.847
	(−0.081)	(1.318)	(1.816)	(−0.715)	(−0.400)	(1.122)
Mother's schooling	0.069	0.282	0.054	−0.045	−1.534*	−0.998
	(0.077)	(0.288)	(0.069)	(−0.050)	(−2.383)	(−0.667)
Highest ISEI of either parent	−0.039	0.089	0.309***	0.122	0.200	0.704***
	(−0.669)	(0.980)	(3.837)	(1.440)	(1.893)	(5.725)
Number of books in home	7.442***	9.983***	7.085***	11.790***	12.677***	12.715***
	(8.707)	(7.368)	(6.220)	(8.681)	(6.607)	(7.379)
Educational resources in home	1.708	6.531**	5.385***	1.393	−1.855	1.568
	(1.433)	(3.120)	(4.523)	(0.812)	(−0.756)	(1.106)
Model B						
Father's schooling	−0.482	0.228	1.151	−0.926	0.770	0.817
	(−0.715)	(0.344)	(1.707)	(−1.149)	(1.180)	(1.269)
Mother's schooling	0.261	−0.256	0.074	0.003	−2.161***	0.035
	(0.391)	(−0.269)	(0.093)	(0.004)	(−3.389)	(0.037)
Highest ISEI of either parent	0.031	0.064	0.306***	0.174*	0.159	0.573***
	(0.488)	(0.756)	(3.957)	(2.039)	(1.402)	(4.921)
Number of books in home	7.610***	8.811***	6.974***	12.557***	12.990***	13.079***
	(9.079)	(6.908)	(6.542)	(10.155)	(7.650)	(9.918)
Educational resources in home	2.436*	7.318***	5.464***	0.744	−0.707	0.628
	(2.409)	(4.255)	(4.771)	(0.422)	(−0.297)	(0.433)

Note: Estimates are weighted by both student and school sampling probabilities. Numbers in parentheses are the absolute values of the ratio of the metric coefficient to its standard error. ISEI = International Socio-economic Index of Occupational Status.
*$p < .05$. **$p < .01$. ***$p < .001$.

test statistic for Taiwan is just barely statistically significant, suggesting that Model A might fit the data better than Model B, inconsistent with the BIC result.

Overall, the cases of Japan, Korea, Czech Republic, and Germany are better described by Model A (the less restrictive MIMIC model) and the U.S. case by Model B (the more restrictive, parsimonious, and stronger MIMIC model). The Taiwanese case is inconclusive, as both MIMIC models might do equally well. In what follows, we report parameter estimates yielded by Model A and Model B, in parallel, for the six countries examined.

Background Effects

Table 3 reports estimates of parameters specified in the within-school model, Equation 1, under Model A and Model B. For each of the six countries examined, the estimates differ in the magnitude between the two models, yet the pattern of the relative importance of background variables is consistent between models in most occasions.

In terms of Model B, father's and mother's education has practically no effect on student academic performance, after taking into account the dominant role of scholarly culture in the home. Parental occupation matters for performance, particularly for Taiwanese, Czech, and U.S. students. In all countries, the dominant dimension of family background is the number of books in the home, confirming the recent literature (Evans et al. 2014). In the Western countries, each shift in the response categories for books in the home increases mean academic performance in the latent variable by 10 points or more. Educational resources also strongly affect academic performance in the East Asian countries but not the Western ones. This might reflect the test-based nature of the East Asian systems: because selection into higher levels of education is based mainly on performance on high-stakes tests, parental investments in educational resources and shadow education may contribute to children's academic performance. Table 3 also shows that the substantive interpretation of the estimates for Model B hold for Model A. In both models, number of books in the home is the dominant factor driving academic performance, followed by educational resources in the East Asian countries.

School Effects

Table 4 reports school-level effects. In the more restrictive Model B, we see large cross-national differences in the key parameters that affect performance, reflecting the institutional particularities of each educational system. Urban–rural differences particularly matter in Taiwan and to a lesser degree in Korea, perhaps due to the presence of geographically remote schools that may have difficulty attracting the best teachers or keeping the most ambitious students. This finding is consistent with Tsai, Gates, and Chiu (1994), who found strong effects of the degree of local industrialization on the educational achievement of Taiwanese students across educational transitions. Although 51.9 percent of Korean 10th graders and 24.8 percent of Japanese 10th graders attend private schools, performance by students in those schools is no different from that by students in public schools. At least the effect is not negative, as is the case in Taiwan, where 44.2 percent of 10th graders in the sample attend private schools, yet they scored 68 points less, on average, than public school students on the latent variable of academic performance. Taiwanese high schools are clearly stratified by prestige ranking and school quality, and students are pushed hard by their parents and teachers to compete against each other over limited spots in good public schools located in major cities. Students whose grades are insufficient for acceptance into a preferred public high school often attend public vocational schools or lower-quality private high schools.

Czech students in private schools (22.3 percent of the weighted sample) also score 20 points lower than students in public schools. Czech private schools may underperform because these students have initially weaker academic potential, and their parents sort them into private schools to help mitigate that weakness; alternatively, Czech private schools may use alternative pedagogies that do not "teach to the test" as many public schools do.

Nonetheless, the most critical school-level factor for student performance in all countries besides the United States is the type of school attended. Performance inequality is particularly stark in the Czech Republic, where under Model B, vocational school students score about 145 points lower in mean performance compared to *Gymnasium* students. Performance gaps are large in other countries as well. For example, even though students at Taiwanese senior vocational schools receive credentials enabling them to apply to a university, their ability to do so might be constrained by their significantly lower performance (by 114 points under Model B) compared to students at Taiwanese senior high schools. These results are robust under the different conditions of Models A and B.

Factor Loadings

Table 5 shows country differences in the magnitude of factor loadings on math, science, and reading. Larger coefficients imply more performance inequality—that is, family- and school-level variables affect that component of academic performance to a larger degree. We posed the question of whether the impact of family background and school attended on student performance is generally the same for different dimensions of performance, such as math, science, and reading—or in other words, whether academic performance is one-dimensional with respect to those background conditions. We can confirm that this is not the case. In Taiwan and Korea, for example, our explanatory variables affect math performance to a greater degree than reading, and they affect reading performance more than science. In Germany and the Czech Republic, the coefficients for math and science are statistically similar in Model B yet larger than the loading for reading performance. In Japan and the United States, the coefficients for science performance are generally the strongest.

To confirm that the coefficients for math, science, and reading performance are statistically different from each other, Appendix Table A provides the results of tests for equality constraints imposed on the factor loadings across subjects for the six countries examined, using BIC as a model selection criterion. The results are already incorporated into the symbols provided in Table 5 ($>$ indicates statistically significant differences; $\approx$ does not). Therefore, we can confirm that family

Table 4. Estimates of School Effects (γ) and Between-level Intercepts, by Country.

	East Asian countries			Western countries		
Variable	Japan	Korea	Taiwan	Czech Republic	Germany	United States
Model A						
School sector (public)						
Private	−18.326	8.538	−71.710***	−20.748*	9.588	4.648
	(−1.929)	(0.709)	(−8.194)	(−2.134)	(1.312)	(0.387)
School location (urban)						
Rural	−12.355	−30.113*	−35.808***	−8.162	8.675	6.319
	(−1.462)	(−2.012)	(−3.669)	(−1.415)	(0.888)	(0.769)
School type (academic)						
Vocational schools	−5.794	−58.205***	−116.960***	—	−87.103***	—
	(−0.752)	(−4.514)	(−9.688)		(−14.881)	
Comprehensive	—	—	-67.112***	—	−106.369***	—
			(-7.179)		(−6.024)	
Technical schools with the school-leaving exam	—	—	—	−79.618***	—	—
				(−10.856)		
Vocational schools without the school-leaving exam	—	—	—	−158.113***	—	—
				(−18.537)		
Between-level intercepts						
Math	491.236***	472.894***	591.132***	554.310***	578.895***	399.762***
	(33.946)	(18.863)	(44.336)	(32.589)	(30.243)	(22.205)
Science	501.566***	478.596***	541.884***	557.918***	583.570***	415.067***
	(29.345)	(28.126)	(57.016)	(37.513)***	(31.556)***	(21.650)
Reading	497.166***	471.763***	545.553***	546.404	564.502	420.821***
	(31.816)	(23.242)	(53.414)	(35.798)	(28.799)	(26.286)

(continued)

Table 4.
(continued)

Variable	East Asian countries			Western countries		
	Japan	Korea	Taiwan	Czech Republic	Germany	United States
Model B						
School sector (public)						
Private	-1.856	0.985	-68.763^{***}	-20.394^{*}	7.534	3.338
	(-0.134)	(0.071)	(-7.052)	(-2.223)	(1.079)	(0.320)
School location (urban)						
Rural	-19.381	-35.673^{*}	-35.451^{***}	-3.608	8.516	3.964
	(-1.531)	(-2.047)	(-3.853)	(-0.602)	(1.018)	(0.546)
School type (academic)						
Vocational schools	-24.372^{*}	-59.966^{***}	-114.591^{***}	—	-83.388^{***}	—
	(-2.272)	(-5.548)	(-8.759)		(-14.641)	
Comprehensive	—	—	-65.723^{***}	—	-103.852^{***}	—
			(-7.124)		(-6.343)	
Technical schools with the school-leaving exam	—	—	—	-74.431^{***}	—	—
				(-10.507)		
Vocational schools without school-leaving exam	—	—	—	-145.770^{***}	—	—
				(-18.298)		
Between-level intercepts						
Math	492.417^{***}	496.426^{***}	590.973^{***}	548.429^{***}	565.439^{***}	399.379^{***}
	(25.295)	(23.231)	(46.224)	(30.989)	(28.334)	(28.921)
Science	505.214^{***}	488.697^{***}	542.016^{***}	555.972^{***}	571.941^{***}	408.867^{***}
	(23.876)	(28.645)	(58.443)	(31.623)	$(27.711)^{***}$	(27.835)
Reading	495.960^{***}	485.966^{***}	545.938^{***}	539.018^{***}	549.086	418.294^{***}
	(23.654)	(26.632)	(53.952)	(36.501)	(30.659)	(29.553)

Note: Estimates are weighted by both student and school sampling probabilities. Numbers in parentheses are the absolute values of the ratio of the metric coefficient to its standard error.
$^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

Table 5. The Magnitude of Factor Loadings (λ) on Three Subjects, by Country.

Country	Model A	Model B
East Asian countries		
Japan	Science > Math ≈ Reading	Science > Reading > Math
	(1.260) (1.000) (0.974)	(1.128) (1.085) (1.000)
Korea	Math > Reading > Science	Math > Reading > Science
	(1.000) (0.798) (0.721)	(1.000) (0.881) (0.848)
Taiwan	Math > Reading ≈ Science	Math > Reading > Science
	(1.000) (0.729) (0.712)	(1.000) (0.765) (0.730)
Western countries		
Czech Republic	Math > Reading ≈ Science	Math ≈ Science > Reading
	(1.000) (0.908) (0.892)	(1.000) (0.993) (0.844)
Germany	Math ≈ Reading ≈ Science	Science ≈ Math > Reading
	(1.000) (0.988) (0.961)	(1.020) (1.000) (0.831)
United States	Science ≈ Math ≈ Reading	Science > Math ≈ Reading
	(1.024) (1.000) (0.968)	(1.088) (1.000) (0.992)

Note: Estimates are weighted by both student and school sampling probabilities. Numbers in parentheses are metric coefficients estimated by the model. The > sign indicates statistically significant differences at the $p < .05$ level; ≈ indicates insignificant differences.

background and school context do not determine math, science, and reading performance to the same degree. Rather, these effects have substantial cross-national differences and patterns, which also vary across model assumptions. It is thus erroneous to presume that models of performance inequality in one subject can be taken as proxies for other subjects; rather, performance inequality can be subject specific, with larger family effects in one subject than another.

Relationship between Equality and Excellence

How can we explain the patterns of performance inequality reflected in the loadings for reading, science, and math? To answer that, we compare the between-level intercepts for each subject and country (reported in Table 4), which are the adjusted mean scores within the framework of Models A and B. These SES-adjusted means do not vary substantially between the two models, but in some countries they vary a great deal in comparison to the unadjusted sample-weighted means in performance (which are also reported in Table 1). The adjusted means are more appropriate measures of academic performance, because they take into account the average values of family and school context and other model parameters for each country.

By comparing the loadings in Table 5 with the intercepts for each subject and country, different patterns emerge between the East Asian and Western countries. In East Asia, performance inequality tends to be largest in the subjects with the highest mean performance, reflecting a *modest* trade-off between inequality and efficiency. For example, in Japan, performance is modestly stronger in science (followed by reading and math), and science performance is influenced to a much stronger degree by family and school context compared to reading and math. In Taiwan, where adjusted math performance is substantially stronger than the other two subjects (590 points in math compared to 545 points in reading and 542 in science, under Model B), performance inequality is also the largest in math by a wide margin, confirming the "excellence-with-inequality" hypothesis for that country. On the other hand, performance inequality in Korea is largest in math, followed by reading and science (in that order), while adjusted mean performance according to Models A and B are roughly similar across the three subjects and substantially lower than the unadjusted means. Overall, in none of the East Asian countries do we find a low factor loading coupled with high achievement, but we do find evidence of a large factor loading coupled with high achievement (math in Taiwan) as well as large factor loadings coupled with modest academic performance (math and science in Korea and Japan, respectively).

The pattern is somewhat different in the Western countries. In Germany, a country with consistently high academic achievement in all three subjects, performance inequality does not statistically differ between the three subjects (the exception being the lower performance inequality in reading according to Model B but not Model A). That is, small differences in inequality between subjects are associated with small differences in achievement between subjects. In the Czech Republic, which has a similar educational system to Germany's, adjusted mean performance actually improves after taking into account model parameters. There, factor loadings between subjects are also quite small (especially compared to the differences in factor loadings estimated in the East Asian countries), and differences in estimated academic achievement between subjects are also relatively small. Similarly, although the United States has low academic achievement compared to other developed nations, it also has relatively small differences in the weight of subjects on latent academic achievement.

Why are the differences in the factor loadings of subjects much larger in East Asian countries (especially in math in Taiwan and Korea, and science in Japan) compared to Western countries? One plausible (and unobserved) explanation is the role of cultural forces: East Asian families likely give much more cultural importance to math and science as an expression of achievement, which, when coupled with the presence of shadow education and related mechanisms, also boosts performance in these culturally most prestigious academic fields. Western cultures, by contrast, likely grant equal importance to these subjects—along the lines of the U.S. emphasis on the "three *R*s" (reading, writing, and arithmetic) as equally important foundations in a standards-based curriculum—which may contribute to smaller differences in inequality and achievement between subjects but not necessarily high or low achievement per se.

Relationship between Equality and Stratification

We also hypothesized that latent academic achievement should be influenced by structural forces to a larger extent in more diversified and stratified school systems, because in such systems, lower-SES students will likely be channeled into vocational tracks where mean achievement is also likely lower. If this is the case, then the explained within-school variance (the share of variance in achievement accounted for by variance in the family-based factors within schools) should be larger in the more stratified school systems of Germany and the Czech Republic as well as possibly in Taiwan, which has more school types at the upper-secondary level than do either Korea or Japan. The explained between-school variance on latent academic achievement, which can be estimated in Model B, should also be larger in the more stratified school systems.

Table 6 reports the R^2 coefficients under Models A and B for the six countries examined in terms of the within- and between-school explained variance. The explained variances in the three subjects under Model B are generally much larger than in Model A, due to the way the two models account for the covariances between error terms. As Table 6 reveals, explained within-school variance under Model A is largest in the United States and smallest in Japan, which does not confirm Hypothesis 2. However, the table does confirm that explained variance between schools is in fact larger in the countries with the more stratified school systems: Germany, Czech Republic, and Taiwan. These results are also confirmed by the explained variance in latent achievement between schools (which is possible to estimate only under Model B), which are several orders of magnitude larger in those three countries compared to the more standardized school systems in Korea, Japan, and the United States. In addition, Table 6 indicates that the explained variance between schools is much larger than the within-school factors under Models A and B for countries with highly stratified school systems, whereas the opposite pattern can be observed in the United States.

CONCLUSION

By applying a country-specific, two-level MIMIC model of family background and school effects on latent academic performance across six nations participating in the 2012 PISA survey, we confirmed that the relative importance of these family effects have country-specific patterns—patterns that would be masked had we used the ESCS variable used by a number of other studies. We also found that family and school effects matter

Table 6. Summary of Explained Variance (R^2), by Country.

	Model A		Model B	
Country	Within school	Between school	Within school	Between school
Japan				
Math	0.032	0.016	0.848	0.917
Science	0.042	0.031	0.927	0.978
Reading	0.025	0.015	0.844	0.893
Latent factor	—	—	0.043	0.080
Korea				
Math	0.082	0.215	0.874	0.933
Science	0.061	0.184	0.907	0.975
Reading	0.066	0.226	0.864	0.890
Latent factor	—	—	0.075	0.270
Taiwan				
Math	0.085	0.811	0.869	0.989
Science	0.089	0.783	0.957	0.978
Reading	0.073	0.712	0.828	0.942
Latent factor	—	—	0.094	0.801
Czech Republic				
Math	0.090	0.804	0.835	0.940
Science	0.078	0.788	0.923	0.933
Reading	0.090	0.807	0.729	0.842
Latent factor	—	—	0.118	0.844
Germany				
Math	0.085	0.771	0.835	0.935
Science	0.087	0.683	0.961	0.903
Reading	0.109	0.606	0.702	0.723
Latent factor	—	—	0.116	0.806
United States				
Math	0.147	0.013	0.909	0.896
Science	0.141	0.012	0.983	0.820
Reading	0.137	0.009	0.887	0.615
Latent factor	—	—	0.150	0.007

substantively more for some subjects than for others—that is, we find substantive differences in performance inequality depending on the subject examined, and thus performance inequality in one subject cannot be taken as a proxy for overall inequalities if an analysis does not actually warrant such an assumption. Compared to the common approach in the social sciences, in which scholars run separate regressions for each subject, the MIMIC model enables testing equality constraints imposed on the factor loadings for each subject, thereby establishing whether those loadings are statistically different from each other, which is certainly more robust than just eyeballing coefficients.

Our empirical results have broad policy relevance. The finding, for example, that achievement inequality in math is substantially larger than in reading or science in Korea and Taiwan suggests that unobserved subject- and country-specific mechanisms are at work, such as differential access to math tutoring or differences in the quality of math instruction between schools. Public resources invested in reducing educational inequality could be more efficiently spent on subject-specific mechanisms, such as better resources for math instructors in schools that underperform in math, rather than tinkering with the educational system as a whole.

We found only modest support for the thesis of a trade-off between performance equality and academic achievement. In Taiwan, larger inequalities in math performance were accompanied by greater math achievement, and we found a similar pattern in Japan in terms of science performance. However, in Japan and Korea, the differences in academic achievement between subjects were small, even though between-subject differences in inequality could be large. Germany and the Czech Republic had small differences in between-subject inequality as well as small differences in achievement between subjects. The United States was the only country we examined in which adjusted scores in reading were actually larger than in math and science and where inequality in reading performance was statistically smaller (than in science, and possibly also math, depending on the model), but the estimated differences in inequality were very small. Overall, we did not find evidence of a "virtuous" association between efficiency and equality, and evidence of a "vicious" association was modest at best. The application of our model to more countries and survey years would likely yield more variation in the equity–efficiency relationship, and thus it is premature to draw definitive conclusions from these six countries.

We also found support for the hypothesis that variance in academic achievement can be explained by school factors to a larger extent in more diversified and stratified school systems. In Taiwan, which has more secondary-school tracking than either Japan or Korea, the explained variance between schools was several times larger than in the latter countries. Germany and the Czech Republic have very similar educational systems as well as similar degrees of achievement dispersion explained at the between-school level. These results suggest that policy interventions in those countries aimed at reducing performance inequality must first and foremost address the low level of achievement at vocational and technical schools, as Table 4 indicated.

Our analysis does face a number of limitations. To maximize cross-national comparability, we used only a small set of background variables. Single-country analyses could take advantage of our multilevel MIMIC model approach but include a wider range of variables, such as gender, ethnicity, regional dummies, or school factors specific to a country, such as indicators of within-school tracking, teacher quality, or school resources. Taking into account within-school tracking would be particularly important for countries with comprehensive school systems, like the United States, where there are few institutional differences between schools in the PISA data.

An alternative approach might specify a three-level MIMIC model that would have standardized school-level variables across countries as well as a set of country-level factors, which could then be applied to a larger set of countries participating in PISA. Specifying such a model is easier said than done, but it is a promising path for future research. In this article, we focused on country-specific school-level factors, which are important for scholars and policy makers concerned with specific school systems, a benefit that would be lost using harmonized data over a larger set of countries. We also sought to demonstrate that the impact of family background and school factors on student performance can be different for different dimensions of performance within each country, for which the two-level MIMIC model is most appropriate.

Finally, the motivation for this article was not only to bring to light methodological problems in the way many sociologists analyze performance inequality but also to explain the merits of an alternative approach that can yield new insights, particularly in the study of educational inequality between academic subjects. The model we specified here can be applied to any number of countries and yield results that are robust, comparable, and policy relevant. The same model could be used, for example, to analyze gender gaps in family and school effects on different subjects within the same country, or to measure change in performance inequality over time.

APPENDIX

Table A. Testing for Equality in Factor Loadings (λ) across Subjects: Sample Size–adjusted BIC.

Model and Contrast	Japan	Korea	Taiwan	Czech Republic	Germany	United States
I. Model with intercorrelations between outcome variables						
(1) Model A	278407.271	230246.746[a]	188308.469	98767.185	60804.623	164599.223
(2) (1) + math = science	278432.965	230343.750	188520.145	98781.430	60802.533	164595.717
(3) (1) + math = reading	278402.109[a]	230289.340	188447.875	98771.086	60800.882	164596.030
(4) (1) + science = reading	278437.251	230250.337	188304.602[a]	98763.123[a]	60801.347	164600.053
(5) (1) + math = science = reading	278447.472	230339.899	188519.640	98778.105	60798.856[a]	164595.288[a]
II. Model without intercorrelations between outcome variables						
(6) Model B	278504.410[a]	230282.380[a]	188302.533[a]	98794.288	60840.186	164593.186
(7) (6) + math = science	278742.836	230643.046	189353.060	98790.187[a]	60837.773[a]	164743.913
(8) (6) + math = reading	278586.140	230471.632	188954.894	98881.940	60895.327	164589.189[a]
(9) (6) + science = reading	278523.863	230298.050	188325.467	98887.872	60923.170	164742.793
(10) (6) + math = science = reading	278743.176	230639.644	189348.122	98908.885	60924.689	164814.309

Note: BIC = Bayesian information criterion.
[a]Model has the lowest BIC within panels.

ACKNOWLEDGMENTS

We would like to thank Cyrus Chu, Gwo-Shyong Shieh, and Raymond Wong for their comments and Ivy Wu for her research assistance.

RESEARCH ETHICS

Research for this article has been carried out consistent with the American Sociological Association's Code of Ethics and related standards of research ethics. Programme for International Student Assessment data are collected by participating Organization for Economic Cooperation and Development and partner states on the basis of the principle of informed consent. This article is based on the analysis of already publicly available data and did not involve original data collection.

FUNDING

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This article was made possible due to financial support from the Institute of Sociology, Academia Sinica and the Ministry of Science and Technology of Taiwan (formerly National Science Council, Grant NSC 101-2923-H-001-001-MY3), for the first author, and the Czech Science Foundation (Grant No. P404/12/J006 for initial work, and Grant No. 14-36154G for subsequent research), for the second author.

REFERENCES

Allmendinger, Jutta. 1989. "Educational Systems and Labor Market Outcomes." *European Sociological Review* 5(3):231–50.

Arum, Richard, and Melissa Velez, eds. 2012. *Improving Learning Environments: School Discipline and Student Achievement in Comparative Perspective*. Stanford, CA: Stanford University Press.

Bishop, John H. 1998. "The Effect of Curriculum-based External Exit Exam Systems on Student Achievement." *Journal of Economic Education* 29(2): 171–82.

Bol, Thijs, Jacqueline Witschge, Herman G. Van de Werfhorst, and Jaap Dronkers. 2014. "Curricular Tracking and Central Examinations: Counterbalancing the Impact of Social Background on Student Achievement in 36 Countries." *Social Forces* 92(4):1545–72.

Brunello, Giorgio, and Daniele Checchi. 2007. "Does School Tracking Affect Equality of Opportunity? New International Evidence." *Economic Policy* 22(52):782–861.

Buchmann, Claudia, Dennis J. Condron, and Vincent J. Roscigno. "Shadow Education, American Style: Test Preparation, the SAT and College Enrollment." *Social Forces* 89:435–61.

Buchmann, Claudia, and Hyunjoon Park. 2009. "Stratification and the Formation of Expectations in Highly Differentiated Educational Systems." *Research in Social Stratification and Mobility* 27(4):245–67.

Carolan, Brian V., and Sara J. Wasserman. 2015. "Does Parenting Style Matter? Concerted Cultivation, Educational Expectations, and the Transmission of Educational Advantage." *Sociological Perspectives* 58(2):168–86.

DiMaggio, Paul. 1982. "Cultural Capital and School Success: The Impact of Status Culture Participation on the Grades of US High School Students." *American Sociological Review* 47(2):189–201.

Evans, M. D. R., Jonathan Kelley, and Joanna Sikora. 2014. “Scholarly Culture and Academic Performance in 42 nations.” *Social Forces* 92(4): 1573–605.

Evans, M. D. R., Jonathan Kelley, Joanna Sikora, and Donald J. Treiman. 2010. “Family Scholarly Culture and Educational Success: Evidence from 27 Nations.” *Research in Social Stratification and Mobility* 28: 171–97.

Gamoran, Adam. 1987. “The Stratification of High School Learning Opportunities.” *Sociology of Education* 60(3):135–55.

Gamoran, Adam, and Robert D. Mare. 1989. “Secondary School Tracking and Educational Inequality: Compensation, Reinforcement, or Neutrality?” *American Journal of Sociology* 94(5):1146–83.

Grodsky, Eric, John R. Warren, and Erika Felts. 2008. “Testing and Social Stratification in American Education.” *Annual Review of Sociology* 34:385–404.

Hanushek, Eric A., and Ludger Woessmann. 2008. “The Role of Cognitive Skills in Economic Development.” *Journal of Economic Literature* 46(3):607–68.

Hauser, Robert M. 1972. “Disaggregating a Social-psychological Model of Educational Attainment.” *Social Science Research* 1:159–88.

Hauser, Robert M. 2009. “On ‘Quality and Equity in the Performance of Students and Schools.’” Center for Demography and Ecology Working Paper No. 2009–06.

Hauser, Robert M., and Arthur S. Goldberger. 1971. “The Treatment of Unobservable Variables in Path Analysis.” *Sociological Methodology* 3:81–117.

Hauser, Robert M., William H. Sewell, and Duane F. Alwin. 1976. “High School Effects on Achievement.” Pp. 309–41 in *Schooling and Achievement in American Society*, edited by W. H. Sewell, R. M. Hauser, and D. L. Featherman. New York: Academic Press.

Hauser, Robert M., and Raymond Sin-Kwok Wong. 1989. “Sibling Resemblance and Intersibling Effects in Educational Attainment.” *Sociology of Education* 62(3):149–71.

Huang, Min-Hsiung. 2009. “Classroom Homogeneity and the Distribution of Student Math Performance: A Country-level Fixed-effects Analysis.” *Social Science Research* 38(4):781–91.

Huang, Min-Hsiung. 2013. “After-school Tutoring and the Distribution of Student Performance.” *Comparative Education Review* 57(4):689–710.

Ishida, Hiroshi, and Satoshi Miwa. 2012. “School Discipline and Academic Achievement in Japan.” Pp. 163–94 in *Improving Learning Environments: School Discipline and Student Achievement in Comparative Perspective*, edited by R. Arum and M. Velez. Stanford, CA: Stanford University Press.

Jackson, Michelle, ed. 2013. *Determined to Succeed? Performance versus Choice in Educational Attainment*. Stanford, CA: Stanford University Press.

Jackson, Michelle, Jan O. Jonsson, and Frida Rudolphi. 2012. “Ethnic Inequality in Choice Driven Education Systems: A Longitudinal Study of Performance and Choice in England and Sweden.” *Sociology of Education* 85(2):158–78.

Kerckhoff, Alan C. 1995. “Institutional Arrangements and Stratification Processes in Industrial Societies.” *Annual Review of Sociology* 21:323–47.

Kerckhoff, Alan C. 2001. “Education and Social Stratification Processes in Comparative Perspective.” *Sociology of Education* Extra Issue:3–18.

Kuan, Ping-Yin. 2011. “Effects of Cram Schooling on Mathematics Performance: Evidence from Junior High Students in Taiwan.” *Comparative Education Review* 55(3):342–68.

Lareau, Annette. 1987. “Social Class Differences in Family–School Relationships: The Importance of Cultural Capital.” *Sociology of Education* 60:73–85.

Lareau, Annette. 2002. “Invisible Inequality: Social Class and Childrearing in Black Families and White Families.” *American Sociological Review* 67 (5): 747–76.

Le Donné, Noémie. 2014. “European Variations in Socioeconomic Inequalities in Students’ Cognitive Achievement: The Role of Educational Policies.” *European Sociological Review* 30(3):329–43.

Lee, Soojeong, and Roger C. Shouse. 2011. “The Impact of Prestige Orientation on Shadow Education in South Korea.” *Sociology of Education* 84(3):212–24.

Lee, Valerie E., and Anthony S. Bryk. 1989. “A Multilevel Model of the Social Distribution of High School Achievement.” *Sociology of Education* 62(3):172–92.

Liu, Jeng. 2012. “Does Cram Schooling Matter? Who Goes to Cram Schools? Some Evidence from Taiwan.” *International Journal of Educational Development* 32(1):46–52.

Lucas, Samuel R. 2001. “Effectively Maintained Inequality: Education Transitions, Track Mobility, and Social Background Effects.” *American Journal of Sociology* 106(6):1642–90.

Marks, Gary N. 2006. “Are Between- and Within-school Differences in Student Performance Largely Due to Socio-economic Background? Evidence from 30 countries.” *Educational Research* 48(1):21–40.

Mijs, Jonathan J. B. 2016. “Stratified Failure: Educational Stratification and Students’ Attributions of Their Mathematics Performance in 24 Countries.” *Sociology of Education* 89(2):137–53.

Montt, Guillermo. 2011. “Cross-national Differences in Educational Achievement Inequality.” *Sociology of Education* 84(1):49–68.

Organization for Economic Cooperation and Development. 2007. “Quality and Equity in the Performance of Students and Schools.” Pp. 169–212 in *Science Competencies for Tomorrow’s World*. Paris: Author.

Organization for Economic Cooperation and Development. 2010. *PISA 2009 Results: What Makes a School*

Successful? Resources, Policies and Practice, vol. IV. Paris: Author.
Organization for Economic Cooperation and Development. 2013. *PISA 2012 Results: Excellence Through Equity*, vol. II. Paris: Author.
Park, Hyunjoon. 2010. "Japanese and Korean High Schools and Students in Comparative Perspective." Pp. 255–73 in *Quality and Inequality of Education: Cross National Perspectives*, edited by J. Dronkers. New York: Springer Press.
Park, Hyunjoon. 2012. "School Disciplinary Climate and Consequences for Student Achievement in Korea." Pp. 251–76 in *Improving Learning Environments: School Discipline and Student Achievement in Comparative Perspective*, edited by R. Arum and M. Velez. Stanford, CA: Stanford University Press.
Park, Hyunjoon. 2013. *Re-evaluating Education in Japan and Korea*. New York: Routledge.
Park, Hyunjoon, Soo-yong Byun, and Kyung-Keun Kim. 2011. "Parental Involvement and Students' Cognitive Outcomes in Korea: Focusing on Private Tutoring." *Sociology of Education* 84(1):3–22.
Park, Hyunjoon, and Pearl Kyei. 2011. "Literacy Gaps by Educational Attainment: A Cross-national Analysis." *Social Forces* 89(3):879–904.
Park, Hyunjoon, and Gary D. Sandefur. 2006. "Families, Schools, and Reading in Asia and Latin America." *Research in Sociology of Education* 15:133–62.
Raudenbush, Stephen W., and Anthony S. Bryk. 1986. "A Hierarchical Model for Studying School Effects." *Sociology of Education* 59(1):1–17.
Raudenbush, Stephen W., and Anthony S. Bryk. 1992. *Hierarchical Linear Models in Social and Behavioral Research: Applications and Data Analysis Methods*. Newbury Park, CA: Sage.
Roksa, Josipa, and Daniel Potter. 2011. "Parenting and Academic Achievement: Intergenerational Transition of Education Advantage." *Sociology of Education* 84(4):299–321.
Shavit, Yossi, Richard Arum, and Adam Gamoran. 2007. *Stratification in Higher Education: A Comparative Study*. Stanford, CA: Stanford University Press.
Stevenson, David L., and David P. Baker. 1992. "Shadow Education and Allocation in Formal Schooling: Transition to University in Japan." *American Journal of Sociology* 97(6):1639–57.
Tsai, Shu-Ling, Hill Gates, and Hei-Yuan Chiu. 1994. "Schooling Taiwan's Women: Educational Attainment in the Mid-20th Century." *Sociology of Education* 67(4):243–63.
University of California–Los Angeles. 2013. "How Can amd Compute a Chi-square Test for Nested Models with the MLR or MLM Estimators?" UCLA Institute for Digital Research and Education. Accessed December 1, 2015 (http://statistics.ats.ucla.edu/stat/mplus/faq/s_b_chi2.htm).
Van de Werfhorst, Herman G., and Jonathan J. B. Mijs. 2010. "Achievement Inequality and the Institutional Systems: A Comparative Perspective." *Annual Review of Sociology* 36:407–28.
Warren, John Robert, and Robert M. Hauser. 1997. "Social Stratification across Three Generations: New Evidence from the Wisconsin Longitudinal Study." *American Sociological Review* 62:561–72.
Woessmann, Ludger. 2004. "How Equal are Educational Opportunities? Family Background and Student Achievement in Europe and the US." Working paper, SSRN.
Woessmann, Ludger. 2007. "Fundamental Determinants of School Efficiency and Equity: German States as a Microcosm for OECD Countries." Working paper, IZA.

Author Biographies

Shu-Ling Tsai, PhD, is research fellow of the Institute of Sociology, Academia Sinica, Taipei, Taiwan. Her main fields of interest are sociology of education, social stratification and social mobility, and income inequality. Her current research is on inequality in student academic achievement and selection in heterogeneous returns to college education.

Michael L. Smith, PhD, is senior researcher at CERGE-EI (a joint workplace of the Center for Economic Research and Graduate Education of Charles University and the Economics Institute of the Czech Academy of Sciences), and at the Institute for Social and Economic Analyses in Prague. His research interests include inequalities in academic performance and attainment, returns to education, tracking mechanisms in education, and social stratification and well-being.

Robert M. Hauser, PhD, is Vilas Research Professor and Samuel Stouffer Professor of Sociology, Emeritus, at the University of Wisconsin-Madison. He also served as executive director of the Division of Behavioral and Social Sciences and Education at the National Academies of Sciences, Engineering, and Medicine. While at the UW-Madison, he directed the Center for Demography of Health and Aging, the Institute for Research on Poverty, and the Center for Demography and Ecology. Hauser's research interests include statistical methodology, trends in educational progression and achievement, the uses of educational assessment as a policy tool, and changes in socioeconomic standing, cognition, health, and well-being across the life course.