

# Measuring family socioeconomic status: An illustration using data from PIRLS 2006

**Daniel H. Caro and Diego Cortés**

*IEA Data Processing and Research Center, Hamburg, Germany*

Many analyses of educational outcomes include a single socioeconomic status (SES) index as a predictor or for statistical control. In large-scale assessment research, these analyses have primarily examined the influence of SES on academic achievement in schools and the influence of other individual, family, and school variables when controlling for SES. The findings have helped elucidate the mechanisms underlying the influence of SES and identify possible avenues for reducing it. The measurement of SES is critically important with respect to the findings and implications of these analyses. This paper demonstrates a step-by-step procedure for calculating a family SES index using data from the Progress in International Reading Literacy Study (PIRLS) 2006. Variables reflecting parental education, parental occupational status, and family wealth were reduced into a single SES index using principal component analysis (PCA). The conceptual definition of SES, its reliability and crossnational comparability are discussed, and recommendations for further research and survey developers are offered. The presented procedure can be extended to other large-scale studies of educational achievement. Its illustration is the main contribution of this paper.

## INTRODUCTION

Researchers have studied the influence of family background on academic performance mainly in school settings (Sirin, 2005; White, 1982). Studies indicate that the performance of students from socioeconomically disadvantaged backgrounds tends to be worse than that of their peers from more affluent families. Gaps between and across students from varying socioeconomic backgrounds tend to widen as students get older (Caro, McDonald, & Willms, 2009; Condrón, 2007; Kerckhoff, 1993; Oakes, 1985) and have lasting consequences on the educational attainment and labor force outcomes of students as adults (Alexander, Entwisle, & Olson, 2007; Kerckhoff, Raudenbush, & Glennie, 2001; Raudenbush & Kasim, 1998; Rumberger, 2010).

National and international student assessment studies offer a significant opportunity to study socioeconomic gaps in academic achievement between and within countries, and thereby to increase our understanding of how socioeconomic inequality reproduces across generations. Although some analyses have included separate variables representing concepts such as human capital, social capital, cultural capital, and economic capital, each postulated by a well-established theoretical model (e.g., Bourdieu, 1977, 1986; Coleman, 1988), others have employed a single variable to represent family socioeconomic status (SES).

Family SES in this latter group of studies is typically defined as the relative position of an individual or family within a hierarchical social structure, based on their access to, or control over, wealth, prestige, and power (Mueller & Parcel, 1981). SES indexes are traditionally operationalized through measures characterizing parental educational levels, parental occupational prestige, and family wealth (Buchmann, 2002; Gottfried, 1985; Hauser, 1994; Schulz, 2005; Yang & Gustafsson, 2004). There is, however, no strong consensus on the conceptual meaning of SES, which limits its use for testing theories and making policy recommendations (Bornstein & Bradley, 2003; Deaton, 2002). Despite these limitations, many studies using the SES approach have made valuable contributions to educational research.

### Practical Contributions of SES Studies

The many studies that have attested to the usefulness of including an SES index in analyses of student learning outcomes generally include an SES index for two main purposes (Buchmann, 2002):

1. To gain an understanding of the extent to which and the mechanisms by which family SES is associated with academic achievement; and
2. To evaluate the influence of individual, family, school, and community aspects while controlling for SES.

In regard to the first goal, Willms (2006a) proposed using socioeconomic gradients to examine the association between SES and academic achievement. Socioeconomic gradients display, in single gradient lines, the strength of the association between SES and academic achievement, the association within schools, and the association across schools. Also, the functional form of socioeconomic gradients (i.e., linear or

curvilinear) indicates whether the SES gap widens, narrows, or remains stable as SES levels increase.

Analyzing socioeconomic gradients helps us characterize the following:

- Achievement inequalities related to SES;
- How these inequalities relate to differences in family SES;
- Which aspect of the inequalities can be attributed to the SES composition of different schools; and
- The influence that the socioeconomic intake of schools has on achievement.

Analysis findings can also allow us to provide guidelines on the potential influence of different policy interventions (Willms, 2006a).

Many studies, including the Programme for International Student Assessment (PISA), conducted by the Organisation for Economic Co-operation and Development (OECD), have adopted the socioeconomic gradient framework for studying educational inequality within and across countries (e.g., Caro & Mirazchiyski, 2012; OECD, 2003, 2004, 2007, 2010; Willms, 2002, 2003, 2006b; Willms, Smith, Zhang, & Tramonte, 2006; Willms & Somers, 2001). Also, in a recent article, Caro and Lenkeit (2012) showed how researchers can use international assessment data to extend this framework to include the analysis of sociological theories. Using data from the Progress in International Reading Literacy Study (PIRLS) 2006, carried out by the International Association for the Evaluation of Educational Achievement (IEA), the authors used two-level and three-level hierarchical linear models to evaluate 10 hypotheses about educational inequality within and across countries.

Researchers have examined other relevant aspects of the association between SES and academic achievement. These include studies of whether and how:

- This association changes over time (Heath & Clifford, 1990; Willms & Raudenbush, 1989);
- Is mediated and moderated by risk and protective factors (Chao & Willms, 2002; Guo & Harris, 2000; Yeung, Linver, & Brooks-Gunn, 2002);
- Is consistent across subject areas (Ma, 2000);
- Changes over the course of schooling (Caro & Lehmann, 2009; Caro, McDonald et al., 2009); and
- Helps to explain school grouping or tracking decisions (Caro, Lenkeit, Lehmann, & Schwippert, 2009; Condrón, 2007; Maaz, Trautwein, Lüdtke, & Baumert, 2008; Schnabel, Alfeld, Eccles, Köller, & Baumert, 2002).

Overall, these studies have provided us with a better understanding of the mechanisms behind the influence of SES on academic achievement.

The second goal of analyses including an SES index has been to control for SES differences when evaluating the influence of other variables on academic achievement. For example, school effectiveness research controls for school SES in order to capture

the effect of schools on academic achievement. Findings from school effectiveness research have helped identify school practices that affect student achievement positively irrespective of the socioeconomic composition of schools (Rutter & Maughan, 2002; Rutter, Maughan, Mortimore, & Ouston, 1979).

Other applications include studying associations with migration background, social capital, cultural capital, and other family variables net of the effect of SES. Also, PISA (OECD, 2003, 2004, 2007) sets out league tables that rank countries according to overall student performance while controlling for SES. Such rankings take into account the varied socioeconomic conditions in which education systems operate and thus provide a fairer picture of relative performance.

### **Research Purpose**

Our main aim in this paper is to demonstrate the calculation and validation of an SES index, using data from PIRLS 2006. Unlike in PISA, an SES index is not available in PIRLS and other studies managed by IEA. Another aim is to stimulate discussion on the possibilities for and the importance of calculating and providing an SES index in IEA datasets. Apart from the implications that our work has for IEA studies, we hope it will provide researchers with a detailed procedure that they can extend to other national and international assessment studies.

Although we apply the traditional procedure for calculating SES, our approach is somewhat unusual on several counts. First, we employ coefficients other than the traditional Cronbach's alpha so as to offer a broader assessment of reliability. Second, we transform the job class classification scheme in PIRLS into an ordinal and internationally comparable occupational status scale. Third, we provide guidelines for SES assessment by placing the measurement of SES within literature on reflective and formative models. And, fourth, we offer researchers and survey developers recommendations relating to SES validation and improvement.

## **REFLECTIVE AND FORMATIVE MODELS**

The literature distinguishes between formative and reflective models for construct measurement (Bollen, 1989; Bollen & Lennox, 1991). These models differ mainly in their underlying assumptions of the causal relationship between the construct and its manifest indicators. In particular, three broad theoretical considerations are indicative of this type of model (Bollen & Lennox, 1991; Borsboom, Mellenbergh, & van Heerden, 2003; Edwards & Bagozzi, 2000; Jarvis, MacKenzie, & Podsakoff, 2003). These are:

1. The nature of the construct (i.e., reflected or formed);
2. The direction of causality between the indicators and the construct (i.e., from the construct to the indicators, or vice versa); and
3. The characteristics of the indicators used to measure the construct (i.e., interchangeable or not).

The reflective view dominates psychological research, while the formative view is common in economics and sociology.

The construct in a reflective model reflects existing theories and exists independently of its manifest indicators. Causality flows from the construct into the observed indicators, but not vice versa, and the nature of the model suggests that adding or dropping one constituent indicator does not alter the definition of the construct. Personality and attitude scales are typical examples of reflective models.

In contrast, a formative scale is viewed as a weighted combination of the observed items, and the construct depends entirely on its operationalization. Causality flows from the observed indicators to the construct. The nature of this approach is such that the constituent components are not interchangeable. In other words, the definition of the construct is affected by the number and type of items.

A well-known example of a formative indicator is the Human Development Index (HDI). This index does not exist as an independent entity but is a composite measure directly determined by health, education, and income variables (United Nations Development Program, 2006). The measurement of SES is also more appropriate under a formative model than under a reflective model (Diamantopoulos, Riefler, & Roth, 2008). Rather than being grounded in solid theory, the SES construct is formed and depends largely on its operationalization. The manifest indicators are given in the literature (i.e., parental educational levels, parental occupational status, and family wealth) and are not interchangeable.

The rules for assessing formative indicators are not well established in the literature, and traditional methods are not appropriate. Reliability analysis assumes high intercorrelations among constituent components, but this is not necessarily the case for formative indicators, where the resulting indicator can be a composite of uncorrelated variables. Validity analysis assumes an underlying theoretical model, but the formative model is fully contingent on the operationalization procedure. A common practice for assessing formative indicators is to examine their effect on a benchmark variable (Bollen & Lennox, 1991).

## **SES MEASUREMENT**

### **Data**

We sourced our data from the PIRLS 2006 survey conducted by IEA (see Mullis, Martin, Kennedy, & Foy, 2007). PIRLS 2006, the second iteration of this survey (the third is presently underway), assessed the reading literacy of fourth graders in 45 education systems. Socioeconomic data were collected through student and home questionnaires (Foy & Kennedy, 2008). The home questionnaire surveyed parents about their educational attainment, jobs, and self-perceived financial situation. The student questionnaire included questions about home possessions. These questions were not available in the home questionnaire. The data relating to home possessions and self-perceived financial situation reflect family wealth. We calculated the parental education and parental occupation indicators of SES by recoding the original data, as follows.

## SES Items

### Parental education (momed/daded)

Parents were asked to state their highest level of education completed. Table 1 presents the original response options and the recoding scheme for the derived mother's and father's education variables. Original response options followed the International Standard Classification of Education (ISCED) (UNESCO, 1999) and were collapsed into five comparable categories for the mother (momed) and the father (daded).

**Table 1: Parental education: coding of original and derived variable**

Original variable		Derived variable	
<i>Categories</i>		<i>Code categories</i>	<i>Code original</i>
Some <ISCED Level 1 or 2> or did not go to school	1	Finished some primary or lower-secondary or did not go to school	1 1
<ISCED Level 2>	2	Finished lower-secondary	2 2
<ISCED Level 3>	3	Finished upper-secondary	3 3
<ISCED Level 4>	4	Finished post-secondary but not university	4 4 and 5
<ISCED Level 5B>	5	Finished university or higher	5 6 and 7
<ISCED Level 5A, first degree>	6	Missing	. 8 and .
Beyond <ISCED Level 5A, first degree>	7		
Not applicable	8		

### Parental occupational status (momsei/dadsei)

The question on occupation in PIRLS aligns with a class-based or categorical rationale to social stratification wherein occupations/members of society are divided into a limited number of discrete categories/classes (see, for example, Goldthorpe, 1980; Goldthorpe, Payne, & Llewellyn, 1978). Accordingly, parents were asked to mark, from a set of options, the kind of work that constituted their main job (see Table 2). The categorical approach assumes that stratification and mobility processes are multidimensional in nature.

In contrast, the SES concept presupposes a hierarchical order that allows an unlimited number of SES groups to be captured in a single status dimension with a continuous scale (Ganzeboom, de Graaf, & Treiman, 1992; Mueller & Parcel, 1981). The continuous approach captures variability within occupation categories, is more amenable to multivariate analysis than the categorical approach, and does not neglect significant information from other occupational dimensions (Ganzeboom et al., 1992). Due to these advantages, and in order to maintain consistency with the SES hierarchical nature, we transformed the class-type job categories in PIRLS into a continuous scale that reflected occupational status.

Ganzeboom et al. (1992) proposed using a scoring scheme to construct their International Socioeconomic Index of Occupational Status (ISEI). This index has been used extensively in studies of social mobility. Table 2 converts job categories from the home questionnaire into the ISEI scores. We took the original job types and descriptions,

as stated in the questionnaires, and used them to find matching categories with the index developed by Ganzeboom and his colleagues. When multiple categories matched the descriptions, we assigned the average score to the job category, thus creating an ISEI variable for mothers (momsei) and fathers (dadsei).

**Table 2: Parental occupational status (ISE scores)**

Original categories	Score
1 <i>Has never worked outside the home for pay</i>	22
2 <i>Small business owner (&lt; 25 employees)</i> Includes owners of small business such as retail shops, services, restaurants	57
3 <i>Clerk</i> Includes office clerks, secretaries, typists, data entry operators, customer service clerks	49
4 <i>Service or sales worker</i> Includes travel attendants, restaurant service workers, personal care workers, protective service workers, salespersons	45
5 <i>Skilled agricultural or fishery worker</i> Includes farmers, forestry workers, fishery workers, hunters and trappers	31
6 <i>Craft or trade worker</i> Includes builders, carpenters, plumbers, electricians, etc.; also metal workers, machine mechanics, handicraft workers	37
7 <i>Plant or machine operator</i> Includes plant and machine operators, assembly-line operators, motor-vehicle drivers	33
8 <i>General laborers</i> Includes domestic helpers and cleaners; building caretakers; messengers, porters and doorkeepers; farm, fishery, agricultural, and construction workers	24
9 <i>Corporate manager or senior official</i> Includes corporate managers such as managers of large companies (25 or more employees) or managers of departments within large companies; legislators or senior government officials; senior officials of special-interest organizations; military officers	67
10 <i>Professional</i> Includes scientists, mathematicians, computer scientists, architects, engineers, life science and health professionals, teachers, legal professionals, social scientists, writers and artists, religious professionals	73
11 <i>Technician or associate professional</i> Includes science, engineering, and computer associates and technicians; life science and health technicians and assistants; teacher aides; finance and sales associate professionals; business service agents; administrative assistants	52

### Home possessions (hompos)

Students in all countries answered questions about the possessions they had available at home, using binary responses (yes/no) to do so. Possessions listed in the survey were a computer, a study desk, a daily newspaper, their own room, and own mobile phone. We fitted the one-parameter Rasch (1960) model, specifically suited for the analysis of binary responses, to the home possessions items data. We then used the conditional maximum likelihood (CML) method to estimate item parameters, and the expected a posteriori (EAP) method (Bock & Aitken, 1981) to estimate home possessions individual scores.

### **Financial status (finan)**

In PIRLS, parents rated the financial wellbeing of their family according to five categories: (1) not at all well-off, (2) not very well-off, (3) average, (4) somewhat well-off, and (5) very well-off. Responses provide a subjective measure of family wealth.

#### ***Missing Data***

As is common in survey studies, our estimation of SES was hampered by a high rate of missing data (May, 2006). Data on parental education, parental occupational status, and home possessions were not collected in the United States, while France and South Africa lacked financial status variables. We therefore excluded these education systems from the estimation of SES. We also excluded education systems with more than 30% of missing student data in three or more SES items: Scotland (51%), England (50%), Israel (45%), Spain (41%), New Zealand (40%), Qatar (39%), Kuwait (38%), the Netherlands (34%), and Iceland (Grade 5) (33%). The remaining 35 education systems comprised the analytic sample.

However, missing data were still present in the SES items of the analytic sample. If the data were missing completely at random (MCAR), analysis relying on complete cases would have yielded unbiased estimates. But unreported analyses suggested that the data were not MCAR. Instead, it appeared that students with missing data on SES items differed systematically from the rest of the students on observed characteristics. For example, we observed that students with missing data tended not to perform as well as the rest of the student cohort on the PIRLS reading test. Therefore, performing complete-cases analysis or listwise deletion would have led to biased estimates; for example, the average reading performance would have been overestimated.

To account for this source of bias, we decided to replace the missing values in SES items with simulated or plausible values calculated with multiple imputation. Importantly, the purpose of multiple imputation is not to predict missing values as closely as possible to the true values, but to account for missing data uncertainty in SES items and thereby yield statistically valid inferences.

Multiple imputation methods produce unbiased estimates if the data are missing at random (MAR) and even when they are not MCAR. The MAR assumption means that the propensity for missing values on SES items is related to other observed variables, but not to unobserved variables or values of the SES items themselves (Little & Rubin, 2002).

In our case, the MAR assumption implied two possibilities. The first was that the propensity for missing values in each of the SES items could be explained through reference to other variables in the PIRLS dataset, for example, family and student characteristics. The second was that the missing SES data would not be higher for specific values of SES items. Thus, for example, the probability of missing parental education information would not be greater for parents with educational attainment at or above the university degree level after we had controlled for observables. If the

MAR assumption held, multiple imputation techniques would produce valid inferences at the analysis stage. (In most analyses, MAR is regarded as a reasonable assumption, especially if a rich set of covariates is available and can be included in the imputation model, as was the situation in our case.)

We employed data augmentation (DA) for the multiple imputation model. DA is an iterative Markov chain Monte Carlo (MCMC) procedure that assumes a multivariate normal distribution for the data (Gelman, Carlin, Stern, & Rubin, 2004; StataCorp, 2009). The imputation model produced five complete versions of a dataset that included the SES items, reading performance, gender, age, reading attitudes, self-concept, early literacy skills, parental attitudes toward reading, migration background, the sampling weight, and the school cluster indicator, among other variables. The sampling variables and potential reading performance outcome helped us preserve the structural characteristics of the data and so obtain valid inferences in the analysis stage (Rubin, 1996). We then applied Rubin's (1987) rule to the imputed datasets to produce point estimates of item weights, reliability coefficients, and correlation coefficients, as reported in Tables 3 and 4 on pages 20 and 22, respectively.

### Principal Component Analysis

We calculated the SES index by applying principal component analysis (PCA) to the six SES constituent items: mother's education, father's education, mother's occupational status, father's occupational status, home possessions, and financial status. The PCA model is consistent with a formative measurement in that the direction of causality goes from the items to the SES index and not vice versa (Diamantopoulos et al., 2008). Put more formally, the SES score for each student  $i$  is a weighted average of the SES items:

$$SES_i = \alpha_1 momed_i + \alpha_2 daded_i + \alpha_3 momsei_i + \alpha_4 dadsei_i + \alpha_5 hompos_i + \alpha_6 finan_i \quad (1)$$

We calculated item weights in each imputed dataset by applying PCA to the complete sample of countries, and then averaged the resulting weights across imputed datasets to produce point estimates of each item weight. We then used the average weights,  $\alpha_s$ , and the item-imputed datasets in equation (1) to calculate five SES scores for each student. As a result, five complete datasets of SES scores were produced. We recommend that model-based estimates using SES should use these five imputed datasets and Rubin's (1987) rule to obtain standard errors that take into account missing data uncertainty.

Figure 1 shows the distribution of the first SES score by education system and for the total international sample. The figure also includes a normal distribution for comparative purposes.

## SES VALIDATION

Criteria for assessing the reliability and validity of formative indicators are not well established in the literature, and traditional methods are not appropriate (Diamantopoulos et al., 2008). As we explained earlier, reliability analysis assumes internal consistency among constituent components, but this is not necessarily the case for formative indicators where the derived variable can be a composite of uncorrelated variables. Validity analysis assumes an underlying theoretical model, but the formative model is not necessarily grounded in theory; it could be fully contingent on the operationalization procedure.

Despite the SES concept not being comparable to the established theories of cultural capital or social capital postulated by Bourdieu (1977) and Coleman (1988), it has proved very useful for understanding educational inequalities. From this perspective, it could be said to refer to a less developed stage of theorization. As such, we considered it likely that the SES model would satisfy certain conditions anticipated by the theory and previous empirical evidence:

1. The SES items of parental education, parental occupational status, home possessions, and financial status would show some degree of intercorrelation or internal consistency if they all indicated the relative position of a family within a hierarchical social structure (Mueller & Parcel, 1981);
2. The item weights would not vary substantially across countries if the SES items consistently reflected SES across cultures; and
3. SES would correlate positively with student achievement, as has been extensively reported in the literature (Sirin, 2005; White, 1982).

We used these criteria for our validation of SES (Bollen & Lennox, 1991; Diamantopoulos et al., 2008).

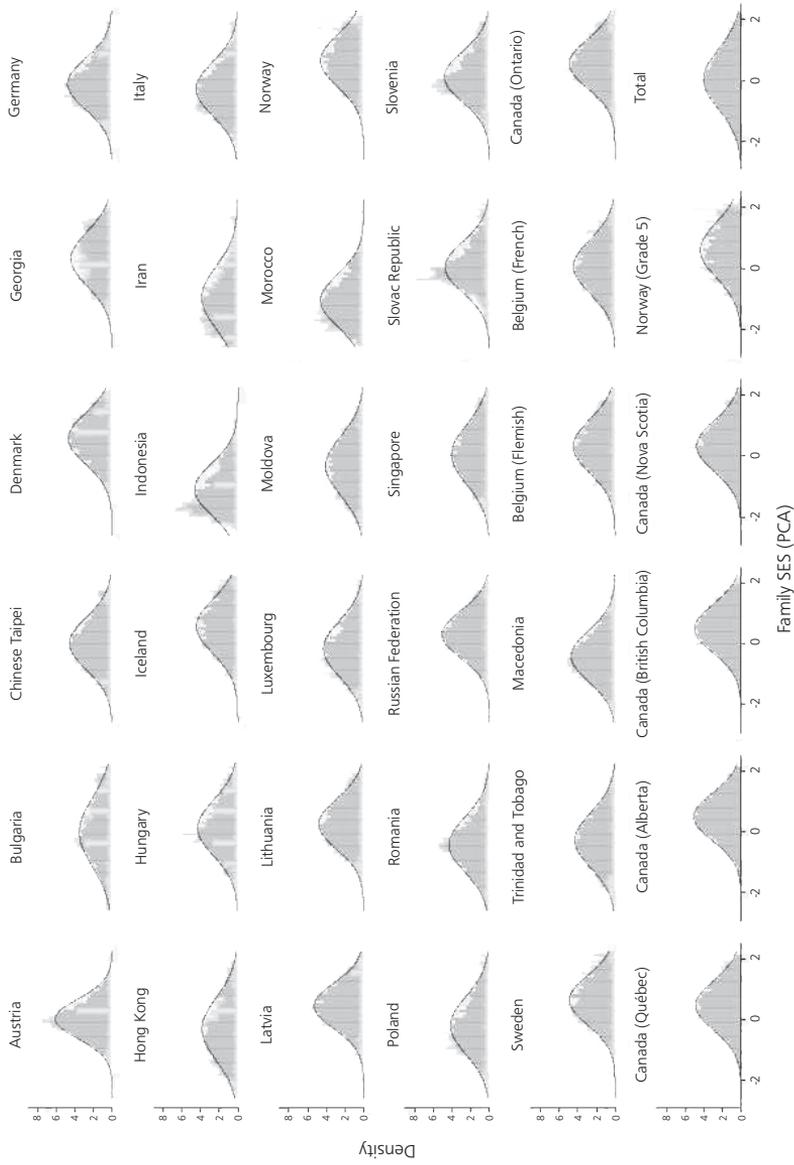
### Reliability

We used Cronbach's alpha coefficient, the beta coefficient (minimum lower bound), and the greater lower bound (GLB) coefficient to examine the consistency of the SES items. Most reliability analyses rely exclusively on alpha, probably because it is easy to estimate with available statistical software. However, alpha does not comply with certain theoretical conditions. It tends, for example, to underestimate true reliability or it provides a lower bound estimate of reliability.

Psychometricians have criticized alpha's overuse and have proposed complementary measures of reliability, such as the beta and GLB coefficients (Bentler, 2009; Sijtsma, 2009). These coefficients are also not exempt from statistical flaws, but they provide a broader assessment of reliability by giving a minimum and maximum value of reliability under lower bound conditions (Revelle, 2012).

The beta coefficient is based on split-half reliability and yields a minimum lower bound of the reliability estimate. We could intuitively expect the coefficient to be derived by finding the split halves least related to the SES item data and by using the inter-

Figure 1: Distribution of SES by education system and for the total international sample



Source: Graphs by 'COUNTRY.ID'.

Table 3: Reliability analysis of SES

Country	Beta	Alpha	Glb
Austria	0.17	0.62	0.73
Belgium (Flemish)	0.07	0.70	0.80
Belgium (French)	0.12	0.68	0.79
Bulgaria	0.65	0.86	0.91
Canada (Alberta)	0.22	0.64	0.74
Canada (British Columbia)	0.35	0.65	0.75
Canada (Novia Scotia)	0.20	0.67	0.76
Canada (Ontario)	0.10	0.65	0.75
Canada (Québec)	0.32	0.69	0.78
Chinese Taipei	0.31	0.73	0.80
Denmark	0.41	0.72	0.81
Georgia	0.39	0.72	0.81
Germany	0.20	0.71	0.81
Hong Kong SAR	0.36	0.79	0.85
Hungary	0.40	0.80	0.86
Iceland	0.04	0.68	0.79
Indonesia	0.41	0.74	0.82
Iran	0.50	0.81	0.86
Italy	0.31	0.75	0.83
Latvia	0.48	0.69	0.78
Lithuania	0.48	0.75	0.83
Luxembourg	0.19	0.72	0.81
Macedonia	0.30	0.70	0.82
Moldova	0.39	0.69	0.79
Morocco	0.50	0.67	0.75
Norway	0.14	0.68	0.79
Norway (Grade 5)	0.28	0.72	0.82
Poland	0.54	0.78	0.84
Romania	0.62	0.82	0.88
Russian Federation	0.46	0.72	0.81
Singapore	0.38	0.76	0.83
Slovak Republic	0.53	0.79	0.86
Slovenia	0.13	0.75	0.84
Sweden	0.23	0.69	0.79
Trinidad and Tobago	0.42	0.71	0.78
<b>Total</b>	0.48	0.76	0.84
Median	0.35	0.72	0.81
Min	0.04	0.62	0.73
Max	0.65	0.86	0.91
CV	0.48	0.08	0.05

group intercorrelation to estimate the total variance accounted for by a general factor (Revelle, 2012). Rather than being a particular estimator, the GLB coefficient is more of a theoretical concept that can be approximated with several methods (Sočan, 2000). We used an iterative procedure that allowed us to produce the highest lower-bound reliability measure still consistent with the data (see also R Development Core Team, 2011; Revelle, 2011; Sočan, 2000).

We used the functions *omega* and *guttman* developed by Revelle (2011) in the *Psych* package for the R project to calculate the beta and GLB coefficients. Table 3 reports the derived coefficients for each education system. As with alpha, beta, and GLB, the coefficients ranged from 0 to 1, with the higher values indicating greater reliability.

The median, minimum, and maximum values across education systems can be found at the bottom of the table, together with the coefficient of variation (CV), that is, the standard deviation divided by the mean.

The alpha coefficient of 0.76 for the total sample indicated satisfactory reliability (see Table 3). Alpha values ranged from 0.62 in Austria to 0.86 in Bulgaria, with a median of 0.72 across education systems and a coefficient of variation of 0.08. The beta and GLB coefficients offered complementary information pertaining to overall reliability and dispersion.

The beta coefficient for the total sample was 0.48, while the specific education system beta coefficients ranged from 0.04 in Iceland to 0.65 in Bulgaria, with a median value of 0.35 (see Table 3). The beta coefficients varied substantially more than the alpha, producing a coefficient of variation of 0.48. The greater variability of beta reflects the half-split nature of this coefficient and was driven mostly by the home possession weight. In fact, the correlation of the home possession weight and the beta coefficient across education systems was 0.95.

In education systems where the correlations between the home possessions item and the other SES items were stronger, the home possessions weight and beta coefficient were also higher. In systems where these correlations were weaker, the beta coefficient and the home possessions weight were lower, at times approaching zero. For example, in Bulgaria, the country with the highest beta coefficient, the correlations of home possessions with financial status and with mother's education were 0.34 and 0.43, respectively, and the home possession weight amounted to 0.32. In Iceland, however, the intercorrelations with home possessions and other items were almost zero, leading to a beta coefficient and home possession weight of about zero (see Table 3 above and also Table 4 below).

GLB coefficients ranged from 0.73 in Austria to 0.91 in Bulgaria, with a median of 0.81 and a coefficient of variation of 0.05 across education systems (see Table 3). The GLB coefficients exhibited the lowest variability according to the coefficient of variation, an outcome that probably reflected the greater stability and accuracy of this estimator (Sočan, 2000). We anticipate that other reliability coefficients in the literature will yield coefficient values that range between 0.48 (beta) and 0.84 (GLB) for the complete international sample (see Table 3).

## **Validation**

We carried out this process separately for the SES items and the total SES scores. We evaluated item validation by considering the variation of item weights across education systems, and evaluated score validation by examining the correlation with reading performance, here the benchmark variable (Bollen & Lennox, 1991; Diamantopoulos et al., 2008).

### Individual Items

Table 4 reports weights for the SES items by education system. The median for all education systems as well as the minimum, maximum, and coefficient of variation are also reported. Here, we could expect that weights would all be positive and relatively stable across education systems if they served to indicate SES equally in different societies.

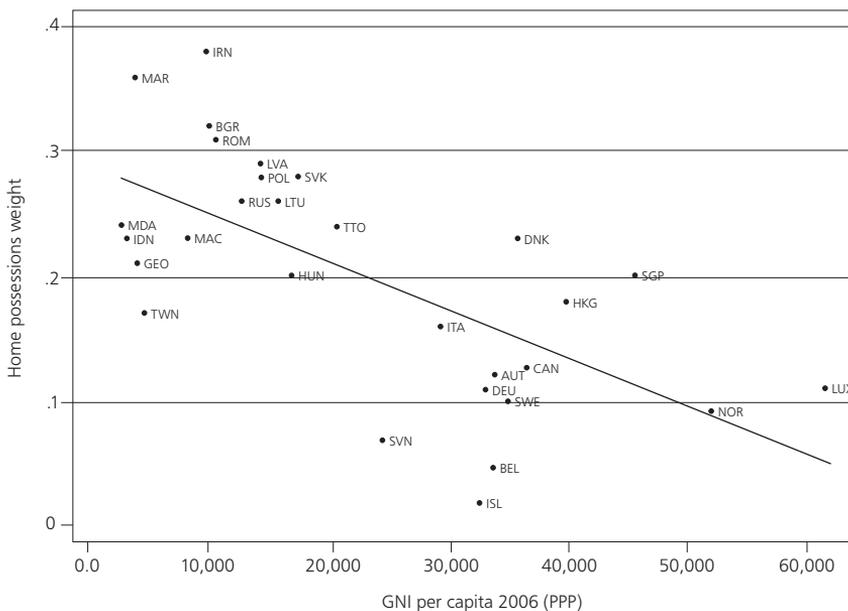
**Table 4: SES item weights and correlation with reading achievement**

Country	Item weights						Rho (SES, reading)
	Parental education		Parental occupational status		Home possessions	Financial status	
	Father	Mother	Father	Mother			
Austria	0.48	0.50	0.50	0.50	0.12	0.08	0.36
Belgium (Flemish)	0.50	0.50	0.47	0.46	0.03	0.24	0.41
Belgium (French)	0.48	0.50	0.48	0.48	0.07	0.22	0.42
Bulgaria	0.46	0.46	0.43	0.45	0.32	0.29	0.42
Canada (Alberta)	0.47	0.49	0.46	0.42	0.13	0.35	0.28
Canada (British Columbia)	0.49	0.49	0.47	0.44	0.14	0.29	0.26
Canada (Nova Scotia)	0.47	0.47	0.46	0.45	0.12	0.36	0.31
Canada (Ontario)	0.47	0.50	0.46	0.45	0.06	0.34	0.24
Canada (Québec)	0.48	0.47	0.45	0.45	0.19	0.32	0.33
Chinese Taipei	0.49	0.49	0.48	0.44	0.17	0.25	0.39
Denmark	0.46	0.47	0.46	0.46	0.23	0.29	0.35
Georgia	0.51	0.50	0.47	0.43	0.21	0.20	0.35
Germany	0.51	0.49	0.48	0.44	0.11	0.24	0.42
Hong Kong SAR	0.47	0.47	0.46	0.42	0.18	0.36	0.15
Hungary	0.47	0.48	0.46	0.47	0.20	0.27	0.49
Iceland	0.48	0.50	0.46	0.49	0.02	0.25	0.31
Indonesia	0.49	0.50	0.47	0.41	0.23	0.27	0.34
Iran	0.48	0.49	0.43	0.37	0.38	0.26	0.51
Italy	0.47	0.49	0.47	0.45	0.16	0.29	0.30
Latvia	0.43	0.47	0.46	0.46	0.28	0.31	0.32
Lithuania	0.47	0.48	0.45	0.47	0.26	0.24	0.40
Luxembourg	0.50	0.49	0.48	0.45	0.11	0.26	0.40
Macedonia	0.50	0.51	0.45	0.48	0.23	0.05	0.46
Moldova	0.50	0.51	0.42	0.46	0.24	0.19	0.30
Morocco	0.52	0.51	0.42	0.29	0.36	0.30	0.23
Norway	0.49	0.51	0.48	0.48	0.08	0.20	0.36
Norway (Grade 5)	0.48	0.50	0.48	0.48	0.17	0.16	0.39
Poland	0.47	0.47	0.45	0.43	0.28	0.31	0.40
Romania	0.46	0.46	0.45	0.44	0.31	0.28	0.48
Russian Federation	0.47	0.48	0.46	0.45	0.26	0.26	0.35
Singapore	0.49	0.49	0.47	0.44	0.20	0.27	0.44
Slovak Republic	0.46	0.46	0.47	0.46	0.28	0.27	0.49
Slovenia	0.46	0.47	0.47	0.47	0.07	0.35	0.40
Sweden	0.49	0.49	0.46	0.48	0.11	0.25	0.36
Trinidad and Tobago	0.48	0.48	0.46	0.44	0.24	0.28	0.41
<b>Total</b>	0.48	0.49	0.44	0.45	0.26	0.26	0.45
Median	0.48	0.49	0.46	0.45	0.19	0.27	0.36
Min	0.43	0.46	0.42	0.29	0.02	0.05	0.15
Max	0.52	0.51	0.50	0.50	0.38	0.36	0.51
CV	0.04	0.03	0.04	0.08	0.50	0.26	0.22

Parental education exerted the greatest load on the SES index, with median weights of 0.48 for father's education and 0.49 for mother's education, followed by parental occupational status, with median weights of 0.46 for father's occupational status and 0.45 for mother's occupational status. The lowest weights corresponded to financial status and home possessions, with median values of 0.27 and 0.19, respectively. The weights of these latter items also exhibited the greatest variation across education systems, with coefficients of variation of 0.26 and 0.50, respectively (see Table 4).

Variation of the home possessions weight was particularly high. We found that it related to the country's income per capita. Figure 2 depicts the association between the home possessions weights and the country's gross national income per capita (PPP) in 2006, as reported by the World Bank (2007). The correlation between income per capita and home possessions weights amounted to -0.66, indicating the greater importance that home possessions has with respect to SES in lower income per capita societies. Iran, for example, had a relatively low income per capita and a home possessions weight of 0.38, whereas Luxembourg had a much higher income per capita; there, the weight was 0.11 (see Figure 2).

Figure 2: Home possessions weights and national income per capita



**Note:**

Country abbreviations: AUT = Austria, BEL = Belgium, BGR = Bulgaria, CAN = Canada, DEU = Germany, DNK = Denmark, GEO = Georgia, HKG = Hong Kong SAR, HUN = Hungary, IDN = Indonesia, IRN = Iran, ISL = Iceland, ITA = Italy, LVA = Latvia, LTU = Lithuania, LUX = Luxembourg, MAC = Macedonia, MAR = Morocco, MDA = Moldova, NOR = Norway, POL = Poland, ROM = Romania, RUS = Russian Federation, SGP = Singapore, SVK = Slovak Republic, SVN = Slovenia, SWE = Sweden, TTO = Trinidad and Tobago, TWN = Chinese Taipei.

This result can also be partly explained by the lower variability in the home possessions item in wealthier societies. In unreported analyses, we found that the mean of the home possessions item increased and the variance reduced significantly as national income per capita levels increased. The home possessions item is thus a less important indicator of SES in wealthier societies because most families tend to have all the possessions surveyed, whereas in poorer countries these possessions still serve to distinguish between higher and lower SES families.

The correlations between the remaining SES items and gross national income per capita were lower and positive for all other SES items: 0.08 for financial status, 0.04 for both father's education and mother's education, 0.55 for father's occupational status, and 0.35 for mother's occupational status.

### ***Correlation with Reading Achievement***

We expected to find a positive correlation between SES and academic achievement. Meta-analyses in the United States indicate a correlation of around 0.30 (Sirin, 2005; White, 1982). Table 4 above shows the correlation between SES and reading achievement in PIRLS 2006 by education system. Correlations ranged from 0.15 in Hong Kong to 0.51 in Iran, with a median of 0.36 for the international sample. These numbers agree with data in studies from the United States, which show a medium level of association between SES and academic achievement. However, a direct comparison was not possible because of a lack of socioeconomic data in the United States.

As with other association statistics, international comparisons of correlation coefficients are restricted by crossnational comparability limitations and therefore need to be interpreted with considerable caution. For example, the correlation between SES and reading achievement yielded the same value of 0.42 in Germany and Bulgaria, but the weight of home possessions was stronger in Bulgaria than in Germany, with values of 0.32 and 0.11, respectively (see Table 4). Because the international SES index uses the complete sample home possessions weight of 0.26, it tended to underestimate the importance of home possessions in Bulgaria and overestimate it in Germany.

Comparability of SES across nations can also be affected by unobserved quality differences in SES. For example, parental education levels capture quantitative differences in education but may not capture differences in the quality of education. And two students in two different countries with comparable SES levels may actually have very different SES if the quality of education attained by their parents differs by country.

## **DISCUSSION**

The procedure that we employed to measure and validate SES used PIRLS data for illustrative purposes. We consider that the procedure can serve as a practical guide for educational researchers seeking to construct single SES indexes and study educational inequalities related to family background when using data from national and international student assessment studies.

The proposed SES index is a composite of variables reflecting parental education, parental occupational status, and family wealth, and it follows traditional operationalizations of the SES concept (Buchmann, 2002; Gottfried, 1985; Hauser, 1994; Mueller & Parcel, 1981). Specifically, the SES index is a composite of six SES items: mother's education, father's education, mother's occupational status, father's occupational status, home possessions, and financial status, with the last two items expected to capture the family wealth component.

Other operationalizations of SES use the maximum educational level and occupational status of either parent to reduce the amount of missing information in SES (e.g., OECD, 2009). However, we used multiple imputation techniques and contextual data to produce five complete versions of the SES item data. The SES model based on six separate items yielded greater reliability than the one with four items using the maximum strategy ( $\alpha = 0.76$  vs.  $0.63$ ), produced a greater number of possible item combinations and therefore also of SES scores, and achieved a greater balance between concepts and variables, with each SES concept (i.e., education, occupation, and wealth) being measured by two items.

The presented reliability analysis provided a broader evaluation of internal consistency than the traditional alpha coefficient. The alpha coefficient of  $0.76$  indicated the overall satisfactory internal consistency for the total sample, and the beta and GLB coefficients indicated that the reliability of the lower bound coefficients lay between  $0.48$  and  $0.84$ . The reliability analysis also suggested that relying on a single coefficient can be misleading. For example, if we had exclusively based our conclusions on alpha, we would have found no substantial difference in terms of SES data consistency between Québec's education system ( $\alpha = 0.69$ ) and Iceland's ( $\alpha = 0.68$ ). However, we would have neglected critical differences captured by the beta coefficient ( $\beta = 0.32$  and  $0.04$  in Québec and in Iceland, respectively). Differences in reliability captured by beta, but not by alpha coefficients are likely explained by the split-half nature of the beta coefficient.

The beta coefficient was severely affected by the home possessions weight or the intercorrelation of home possessions and other SES items. In education systems where intercorrelations are low, the beta coefficient will also tend to be low, and vice versa, which is why the correlation of the home possessions weight and the beta coefficient across the education systems was almost one. In Iceland, intercorrelations with the home possessions items were almost zero, which meant the beta coefficient was almost zero as well, whereas in Québec, home possessions still indicated SES with a weight of  $0.19$ . Thus, in addition to the alpha coefficient, the beta coefficient suggests that SES reliability in Iceland was lower than in Québec, and particularly so if we assume that items should weigh more or less equally in the calculation of SES.

In general, the lower bound reliability coefficients indicated overall satisfactory internal consistency for the total sample ( $\beta = 0.48$ ,  $\alpha = 0.76$ ,  $GLB = 0.84$ ). However, the SES items were not equally indicative of SES across nations. For example, we found a substantial positive correlation between the parental occupational status items

and national income per capita. That the weight of the mother's and the father's occupational status items was higher in wealthier nations likely reflects the greater income and more varied occupations of fathers and mothers in wealthier societies. Mothers, especially, tend to be more educated and employed in different sectors of the economy in wealthier than in poorer societies.

Although parental occupational status was a better indicator of SES in the wealthier societies, it still served to indicate SES in poorer societies. In contrast, the home possessions component did not seem to reflect SES in all nations. The home possessions weight was the least stable across nations and was strongly and negatively correlated with the national income per capita. It seems that the home possessions items surveyed in the study were important indicators of social mobility in the poorer societies, but that their importance decreased as income per capita increased, eventually becoming negligible for the wealthiest countries participating in PIRLS.

### **Limitations and Implications for Future Research and Survey Development**

Probably the most important limitation is the theoretical basis of the proposed SES index. There is no consensus on the conceptual meaning of SES, which prevents researchers from understanding the mechanisms underlying the association between SES and student outcomes. And although most studies regard SES as a formative indicator, they often also use different variables to measure it, thereby violating the operationalization condition of formative measurement and implicitly suggesting that SES does reflect a well-established concept.

We think it likely that researchers intuitively invoke well-developed concepts of human capital, cultural capital, and economic capital to justify the calculation of a single SES index, but a comprehensive SES theory has yet to be elaborated. It seems, therefore, that further work is needed on the theoretical foundations of SES. A related conjecture is that once a formal definition of SES is developed, construct measurement and validity could switch from the current formative model to a reflective model. Confirmatory factor analysis and crosscultural invariance could then be applied for validity assessment, for example. In this current study, we validated the SES index by analyzing internal consistency, the correlation with reading achievement, and item weights across education systems, but these analyses were insufficient for proper validity assessment.

Another limitation for SES validation is the substantial variability of the home possessions weight across education systems. Further research should study the possible reasons as to why the lowest weight was that for home possessions in wealthier societies. One plausible explanation supported by the data is that students in wealthier societies have greater access to and low variability in the home possessions surveyed (i.e., computer, study desk, daily newspaper, own room, and own mobile phone), whereas these items still distinguish higher-SES and lower-SES families in poorer countries. But another plausible explanation is that, irrespective of item selection, home possessions play a less important role in SES for wealthier societies.

Research identifying home possessions items that reflect SES in both poorer and wealthier societies would contribute to survey development and improve the crossnational validity of the SES index. In this regard, May (2006) provides an interesting methodological proposal using national-specific items in TIMSS 1995 and also available for PIRLS. Brese and Mirazchyski (2010), however, found a negligible association between the national-specific home possessions items and academic achievement. Also, a review of the empirical evidence and theories in the international literature could help us determine if this pattern can be explained mostly by data restrictions or by substantive reasons. Whatever the case, research in this area could contribute importantly to survey development and the study of educational inequalities.

Still another limitation is the presence of missing data in the SES items. We employed multiple imputation to counteract this potential source of bias, but the quality of the derived SES index might still be affected by the substantial amount of missing data. It is also possible that multiple imputation would not yield unbiased estimates if the MAR assumption were not satisfied, that is, if the data were not missing at random (NMAR). This situation would occur if the missing data mechanism depended not only on the observed information but also on unobserved variables. And although the MAR assumption is usually realistic when a large set of variables is available, as with PIRLS, in some cases it can be violated. This situation is evident, for example, in longitudinal studies where lower-SES participants are more likely to leave the study earlier. Another example relates to collecting annual income data, wherein families are less likely to report their income once a certain income level is surpassed. In general, though, it is almost impossible to determine the missing data mechanism, the MAR assumption is not testable, and the possibility of a NMAR mechanism can never be completely ruled out.

These situations are why we recommend further research that uses sensitivity analysis to address the possibility that the data are NMAR—where the “missingness” depends on unobservables (Resseguier, Giorgi, & Paoletti, 2011). Sensitivity analysis identifies groups according to missing-data patterns and then evaluates whether critical estimates vary across groups. The results help establish a range of conditions under which derived estimates are unbiased as well as conditions under which estimates are biased. Another alternative is to use techniques that directly allow using data without the need to impute missing responses, or directly incorporate the mechanisms for missing data into the model (Rose, von Davier, & Xu, 2010).

Yet another limitation relates to how parental occupational data were collected during PIRLS. A multiple-choice question was used to ask parents about their occupations. However, research shows that the categorical approach to social stratification has several disadvantages in relation to a continuous or hierarchical scheme achieved with an open-ended question (Ganzeboom et al., 1992). We referenced the ISEI developed by Ganzeboom et al. (1992) to assign ISE scores to the 11 job categories in PIRLS. However, the reduced number of categories in PIRLS limited the variability and validity that the ISEI offers. A broader set of options grounded in theory would therefore be

needed to establish a more accurate scoring of occupations. Survey developers should therefore consider existing theories in this area, their proposed survey questions, and the missing data rates in previous studies (Buchmann, 2002).

Researchers and survey developers should also carefully consider the consequences of collecting socioeconomic data from parents and students. Theoretically, parents are the ultimate authority for reporting SES, while students, especially in the early grades, tend to lack a precise knowledge of socioeconomic home resources and are more likely to provide less reliable SES information (Kreuter, Eckman, Maaz, & Watermann, 2010; Sirin, 2005). In practice, however, the implementation of parent questionnaires or the application of specific items is not always possible (Schulz, 2005). Therefore, it is important to know how unreliable student responses are with respect to the different SES concepts.

Research shows that unreliability of student responses is lower for home possessions, as utilized in PIRLS 2006, than for parental occupations and for parental educational levels, in that order (Brese & Mirazchiyski, 2010; Schulz, 2005; Yang & Gustafsson, 2004). We therefore expect that our SES index will be less affected by unreliable responses than other measures that draw on student reports of parental education or parental occupations. Research also shows that the reliability of socioeconomic information provided by students varies by country and increases with age, family SES, and academic achievement levels (Brese & Mirazchiyski, 2010; Kreuter et al., 2010). Socioeconomic information provided by parents can also be somewhat biased by social desirability and high levels of non-response (Schulz, 2005). In any case, survey developers and researchers should be aware of these caveats when collecting and analyzing SES information.

A final limitation regarding the application of this approach to other IEA studies is that studies unevenly collect socioeconomic data. For example, PIRLS and the International Civic and Citizenship Education Study (ICCS) gather more information on socioeconomic background than does TIMSS, where no parental occupation or financial status data are available (Buchmann, 2002; Foy & Kennedy, 2008). Therefore, our presented SES estimation procedure cannot be directly generalized to other IEA studies. If the goal is to adopt the SES approach, survey developers should consider defining a set of socioeconomic variables that can be collected evenly across studies. However, further research and discussion are required before advocating this approach. Our aim in this paper has been to contribute to this direction.

**Acknowledgement:** The authors are very grateful to two anonymous reviewers as well as Dirk Hastedt, Robert Whitwell, and Aurora Cortés for helpful comments and suggestions on previous drafts of this paper.

## References

- Alexander, K. L., Entwisle, D. R., & Olson, L. S. (2007). Lasting consequences of the summer learning gap. *American Sociological Review*, *72*, 167–180.
- Bentler, P. M. (2009). Alpha, dimension-free, and model-based internal consistency reliability. *Psychometrika*, *74*(1), 137–143.
- Bock, R. D., & Aitken, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika*, *46*(4), 443–459.
- Bollen, K. A. (1989). *Structural equations with latent variables* (Wiley Series in Probability and Mathematical Statistics). New York, NY: Wiley.
- Bollen, K. A., & Lennox, R. (1991). Conventional wisdom on measurement: A structural equation perspective. *Psychological Bulletin*, *110*(2), 305–314.
- Bornstein, M. H., & Bradley, R. H. (Eds.). (2003). *Socioeconomic status, parenting, and child development*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Borsboom, D., Mellenbergh, G., & van Heerden, J. (2003). The theoretical status of latent variables. *Psychological Review*, *110*(2), 203–219.
- Bourdieu, P. (1977). Cultural reproduction and social reproduction. In J. Karabel & A. H. Halsey (Eds.), *Power and ideology in education* (pp. 487–511). New York, NY: Oxford University Press.
- Bourdieu, P. (1986). The forms of capital. In J. G. Richardson (Ed.), *Handbook of theory and research for the sociology of education* (pp. 241–258). New York, NY: Greenwood Press.
- Brese, B., & Mirazchiyski, P. (2010, July). *Measuring students' family background in large-scale education studies*. Paper presented at the fourth IEA International Research Conference in Gothenburg, Sweden.
- Buchmann, C. (2002). Measuring family background in international studies of education: Conceptual issues and methodological challenges. In A. C. Porter & A. Gamoran (Eds.), *Methodological advances in cross-national surveys of educational achievement* (pp. 150–197). Washington, DC: National Academy Press.
- Caro, D. H., & Lehmann, R. (2009). Achievement inequalities in Hamburg schools: How do they change as students get older? *School Effectiveness and School Improvement*, *20*(4), 407–431.
- Caro, D. H., & Lenkeit, J. (2012). An analytical approach to study educational inequalities: Ten hypothesis tests in PIRLS 2006. *International Journal of Research and Method in Education*, *35*(1), 3–30.
- Caro, D. H., Lenkeit, J., Lehmann, R., & Schwippert, K. (2009). The role of academic achievement growth in school track recommendations. *Studies in Educational Evaluation*, *35*(4), 183–192.
- Caro, D. H., McDonald, T., & Willms, J. D. (2009). Socio-economic status and academic achievement trajectories from childhood to adolescence. *Canadian Journal of Education*, *32*(3), 558–590.

- Caro, D. H., & Mirazchiyski, P. (2012). Socioeconomic gradients in Eastern European countries: Evidence from PIRLS 2006. *European Educational Research Journal*, 11(1), 96–110.
- Chao, R. K., & Willms, J. D. (2002). The effects of parenting practices on children's outcomes. In J. D. Willms (Ed.), *Vulnerable children: Findings from Canada's national longitudinal survey of children and youth* (pp. 149–166). Edmonton, Alberta, Canada: University of Alberta Press.
- Coleman, J. S. (1988). Social capital in the creation of human capital. *American Journal of Sociology*, 94(Supplement), S95–S120.
- Condron, J. (2007). Stratification and educational sorting: Explaining ascriptive inequalities in early childhood reading group placement. *Social Problems*, 54(1), 139–160.
- Deaton, A. (2002). Policy implications of the gradient of health and wealth. *Health Affairs*, 21(2), 13–30.
- Diamantopoulos, A., Riefler, R., & Roth, K. P. (2008). Advancing formative measurement models. *Journal of Business Research*, 61(12), 1203–1218.
- Edwards, J. R., & Bagozzi, R. P. (2000). On the nature and direction of relationships between constructs and measures. *Psychological Methods*, 5(2), 155–174.
- Foy, P., & Kennedy, A. M. (Eds.). (2008). *PIRLS 2006 user guide Supplement 3 for the international database*. Chestnut Hill, MA: Boston College.
- Ganzeboom, H. B. G., de Graaf, P. M., & Treiman, D. J. (1992). A standard international socio-economic index of occupational status. *Social Science Research*, 21, 1–56.
- Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2004). *Bayesian data analysis* (2nd ed.). London, UK: Chapman & Hall.
- Goldthorpe, J. H. (1980). *Social mobility and class structure in modern Britain*. Oxford, UK: Clarendon Press.
- Goldthorpe, J. H., Payne, C., & Llewellyn, C. (1978). *Trends in class mobility*. *Sociology*, 12(3), 441–468.
- Gottfried, A. (1985). Measures of socioeconomic status in child development research: Data and recommendations. *Merrill-Palmer Quarterly*, 31(1), 85–92.
- Guo, G., & Harris, K. (2000). The mechanisms mediating the effects of poverty on children's intellectual development. *Demography*, 37(4), 431–447.
- Hauser, R. M. (1994). Measuring socioeconomic status in studies of child development. *Child Development*, 65(6), 1541–1545.
- Heath, A. F., & Clifford, P. (1990). Class inequalities in education in the twentieth century. *Journal of the Royal Statistical Society: Series A*, 153(1), 1–16.
- Jarvis C. B., Mackenzie, S. B., & Podsakoff, P. M. (2003). A critical review of construct indicators and measurement model misspecification in marketing and consumer research. *Journal of Consumer Research*, 30(3), 199–218.
- Kerckhoff, A. C. (1993). *Diverging pathways: Social structure and career deflections*. Cambridge, UK/New York, NY: Cambridge University Press.

- Kerckhoff, A., Raudenbush, S., & Glennie, E. (2001). Education, cognitive skill, and labor force outcomes. *Sociology of Education*, 74(1), 1–24.
- Kreuter, F., Eckman, S., Maaz, K., & Watermann, R. (2010). Children's reports of parents' education level: Does it matter whom you ask and what you ask about? *Survey Research Methods*, 4(3), 127–138.
- Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data* (2nd ed.). Hoboken, NJ: Wiley.
- Ma, X. (2000). Socioeconomic gaps in academic achievement within schools: Are they consistent across subject areas? *Educational Research and Evaluation*, 6(4), 337–355.
- Maaz, K., Trautwein, U., Lüdtke, O., & Baumert, J. (2008). Educational transitions and differential learning environments: How explicit between-school tracking contributes to social inequality in educational outcomes. *Child Development Perspectives*, 2(2), 99–106.
- May, H. (2006). A multilevel Bayesian item response theory method for scaling. *Journal of Educational and Behavioral Statistics*, 31(1), 63–79.
- Mueller, C. W., & Parcel, T. L. (1981). Measures of socioeconomic status: Alternatives and recommendations. *Child Development*, 52(1), 13–30.
- Mullis, I. V. S., Martin, M. O., Kennedy, A. M., & Foy, P. (2007). *PIRLS 2006 international report: IEA's Progress in International Reading Literacy Study in primary schools in 40 countries*. Chestnut Hill, MA: Boston College.
- Oakes, J. (1985). *Keeping track: How schools structure inequality*. New Haven, CT: Yale University Press.
- Organisation for Economic Co-operation and Development (OECD). (2003). *Literacy skills for the world of tomorrow: Further results from PISA 2000*. Paris, France: Author.
- Organisation for Economic Co-operation and Development (OECD). (2004). *Learning for tomorrow's world: First results from PISA 2003*. Paris, France: Author.
- Organisation for Economic Co-operation and Development (OECD). (2007). *PISA 2006: Science competencies for tomorrow's world*. Paris, France: Author.
- Organisation for Economic Co-operation and Development (OECD). (2009). *PISA 2006 technical report*. Paris, France: Author.
- Organisation for Economic Co-operation and Development (OECD). (2010). *PISA 2009 results: Overcoming social background. Equity in learning opportunities and outcomes* (Vol. II). Paris, France: Author. Available online at <http://dx.doi.org/10.1787/9789264091504-en>
- R Development Core Team. (2011). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Available online at <http://www.R-project.org/>
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen, Denmark: Nielsen and Lydiche.
- Raudenbush, S. W., & Kasim, R. M. (1998). Cognitive skill and economic inequality: Findings from the National Adult Literacy Study. *Harvard Educational Review*, 68(1), 33–79.

- Resseguier, N., Giorgi R., & Paoletti, X. (2011). Sensitivity analysis when data are missing not-at-random. *Epidemiology*, *22*(2), 282.
- Revelle, W. (2011). *Procedures for personality and psychological research*. Evanston, IL: Northwestern University. Available online at <http://personality-project.org/r/psych.manual.pdf.1.0-97>
- Revelle, W. (2012, Spring). *An introduction to psychometric theory with applications in R*. Evanston, IL: Northwestern University. Available online at <http://personality-project.org/r/book/>
- Rose N., von Davier, M., & Xu, X. (2010). *Modeling nonignorable missing data with item response theory (IRT) (ETS Research Report No. RR-10-11)*. Princeton, NJ: Educational Testing Service.
- Rubin, D. B. (1987). *Multiple imputation for non-response in surveys*. New York, NY: John Wiley & Sons.
- Rubin, D. B. (1996). Multiple imputation after 18+ years. *Journal of the American Statistical Association*, *91*(434), 473–489.
- Rumberger, R. W. (2010). Education and the reproduction of economic inequality in the United States: An empirical investigation. *Economics of Education Review*, *29*(2), 246–254.
- Rutter, M., & Maughan, B. (2002). School effectiveness findings 1979–2002. *Journal of School Psychology*, *40*(6), 451–475.
- Rutter, M., Maughan, B., Mortimore, P., & Ouston, J. (1979). *Fifteen thousand hours: Secondary schools and their effects on children*. Cambridge, MA: Harvard University Press.
- Schnabel, K., Alfeld, C., Eccles, J., Köller, O., & Baumert, J. (2002). Parental influence on students' educational choices in the United States and Germany: Different ramifications—same effect? *Journal of Vocational Behavior*, *60*(2), 178–198.
- Schulz, W. (2005, April). *Measuring the socio-economic background of students and its effect on achievement in PISA 2000 and PISA 2003*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.
- Sijtsma, K. (2009). On the use, misuse, and the very limited usefulness of Cronbach's Alpha. *Psychometrika*, *74*(1), 107–120.
- Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research*, *75*(3), 417–453.
- Sočan, G. (2000). Assessment of reliability when test items are not essentially  $\tau$ -equivalent. *Developments in Survey Methodology*, *15*, 23–35. Available online at <http://mrvar.fdv.uni-lj.si/pub/mz/mz15/socan.pdf>
- StataCorp. (2009). *Stata 11 multiple-imputation reference manual*. College Station, TX: Stata Press.
- The World Bank. (2007). *GNI per capita, PPP (current international \$)*. Washington, DC: Author. Retrieved from <http://data.worldbank.org/indicator/NY.GNP.PCAP.PP.CD?page=1>

- United Nations Development Program (UNDP). (2006). *Human development report*. New York, NY: Palgrave.
- United Nations Educational, Scientific and Cultural Organization (UNESCO). (1999). *International Standard Classification of Education–ISCED*. Paris, France: Author.
- White, K. R. (1982). The relation between socioeconomic status and academic achievement. *Psychological Bulletin*, *91*(3), 461–481.
- Willms, J. D. (Ed.). (2002). *Vulnerable children: Findings from Canada’s National Longitudinal Survey of Children and Youth*. Edmonton, Alberta, Canada: University of Alberta Press.
- Willms, J. D. (2003). *Ten hypotheses about socioeconomic gradients and community differences in children’s developmental outcomes*. Ottawa, Ontario, Canada: Applied Research Branch of Human Resources Development Canada.
- Willms, J. D. (2006a). *Learning divides: Ten policy questions about the performance and equity of schools and schooling systems*. Montreal, Québec, Canada: UNESCO Institute for Statistics.
- Willms, J. D. (2006b). Variation in socioeconomic gradients among cantons in French- and Italian-speaking Switzerland: Findings from the OECD PISA. *Educational Research and Evaluation*, *12*(2), 129–154.
- Willms, J. D., & Raudenbush, S. W. (1989). A longitudinal hierarchical linear model for estimating school effects and their stability. *Journal of Educational Measurement*, *26*(3), 209–232.
- Willms, J. D., Smith, T. M., Zhang, Y., & Tramonte, L. (2006). Raising and levelling the learning bar in Central and Eastern Europe. *Prospects*, *36*(4), 411–418.
- Willms, J. D., & Somers, M.-A. (2001). Family, classroom, and school effects on children’s educational outcomes in Latin America. *Journal of School Effectiveness and School Improvement*, *12*(4), 409–445.
- Yang, Y., & Gustafsson, J.-E. (2004). Measuring socioeconomic status at individual and collective levels. *Educational Research and Evaluation*, *10*(3), 259–288.
- Yeung, W. J., Linver, M. R., & Brooks-Gunn, J. (2002). How money matters for young children’s development: Parental investment and family processes. *Child Development*, *73*, 1861–1879.