Journal of Educational and Behavioral Statistics
2014, Vol. 39, No. 5, pp. 333–367
DOI: 10.3102/1076998614547576

Composition, Context, and Endogeneity in School and Teacher Comparisons

Katherine E. Castellano
Sophia Rabe-Hesketh
University of California, Berkeley

Anders Skrondal
Norwegian Institute of Public Health

Investigations of the effects of schools (or teachers) on student achievement focus on either (1) individual school effects, such as value-added analyses, or (2) school-type effects, such as comparisons of charter and public schools. Controlling for school composition by including student covariates is critical for valid estimation of either kind of school effect. Student covariates often have different effects between schools than within schools. Econometricians typically attribute such differences to a form of endogeneity*, specifically, "Level-2 endogeneity," or the confounding of student covariates with unobserved school characteristics, whereas education researchers primarily interpret the differences as* contextual effects *or the effects of collective peer attributes on individual student achievement. This article considers both and makes connections between the econometric and education research literatures. We show that the Hausman and Taylor approach from panel data econometrics can be used for valid estimation of individual school or school-type effects when there is only Level-2 endogeneity but can lead to bias when there are also contextual or peer effects. In contrast, contextual effects are typically estimated by including school means of student covariates in addition to the student-level covariates (equivalent to the Mundlak device), but this leads to biased school comparisons in the presence of Level-2 endogeneity. We interpret the estimates from these two competing estimators in terms of the "Type A" and "Type B" school effects defined by Raudenbush and Willms and show that both estimators are preferable to the common group-mean-centering approach.*

Keywords: *compositional effect; contextual effect; endogeneity; hierarchical linear model; multilevel model; school effects; value-added*

1. Introduction

Given the diversity of resources, practices, and student populations of schools and teachers, comparisons of the effectiveness of different types of schools and teachers are widespread in education research. Moreover, with recent educational initiatives such as the U.S. Race to the Top Grant program, more focus has been placed on using student achievement scores in teacher effectiveness and school accountability measures (U.S. Department of Education, 2010).

In this article, we consider the problem of comparing the effectiveness of schools or teachers based on student outcomes, particularly, educational achievement scores. Such comparisons may be between individual schools or teachers or between types of schools, such as Catholic versus public, or types of teachers, such as having a master's degree or not. For simplicity, we refer to these as "school" comparisons throughout the article, but all results and discussions hold for teacher comparisons as well with, perhaps, added concerns associated with teachers nested within schools and students being taught by multiple teachers. In regression models, individual school effects are typically represented by random or fixed school-specific intercepts, whereas school-type effects are represented by (fixed) regression coefficients of dummy variables for school type. Therefore, we consider estimation of school-specific intercepts for individual school comparisons, coefficients of school-level covariates for school-type comparisons, and the variance of the school intercepts for quantifying the variability of school effects. In particular, we focus on two challenges to the valid estimation of these parameters—Level-2 endogenous covariates and contextual effects—both of which are defined with a common education example.

To estimate school effects, it is crucial to control for student background using demographic variables or prior test scores (e.g., McCaffrey, Lockwood, Koretz, Louis, & Hamilton, 2004; National Research Council & National Academy of Education, 2010; Raudenbush & Willms, 1995; Wayne & Youngs, 2003). This helps "level the playing field," as schools can differ considerably in terms of the composition of the student body due to nonrandom assignment of students to schools (Aud et al., 2012; Palardy & Rumberger, 2008; Rivkin, Hanushek, & Kain, 2005). For instance, students with high socioeconomic status (SES), who tend to have good educational outcomes, are more likely to attend certain schools than others. Even if schools had no effect on student outcomes and students had no effects on each other, the schools with larger proportions of high-SES students would be expected to have better average test scores than those with low proportions. Following Duncan, Jones, and Moon (1998), we use the term *compositional effects* for such contributions to school differences that are purely due to differences in the prevalence of students with favorable background characteristics.

To eliminate compositional effects due to SES, we might be tempted to control for SES (either using student SES or using school mean SES). However, schools that attract more high-SES children also tend to have better teachers and more resources, making them more effective (Ballou, Sanders, & Wright, 2004; Kennedy & Mandeville, 2000; National Research Council & National Academy of Education, 2010; Nye, Konstantopoulos, & Hedges, 2004). Therefore, the coefficient of school mean SES absorbs some of the effect of the schools, leading to an underestimate of differences in effectiveness. In other words, some of the beneficial effects of unobserved school characteristics are falsely attributed to SES.

The problem is that SES is positively correlated with the school-level random intercepts (or errors/residuals) that represent the combined effect of unobserved school characteristics on student outcomes. This positive correlation results in positive omitted variable bias for the coefficient of SES, and this "overcontrolling" for SES leads to underestimation of school differences. Covariates that are correlated with error terms are called *endogenous* in econometrics (e.g., Wooldridge, 2010), and here we call covariates *Level-2 endogenous* when they are correlated with the school-level (Level 2) random intercepts.

Econometricians have addressed Level-2 endogeneity extensively in the context of longitudinal or panel data (e.g., Hausman & Taylor, 1981; Mundlak, 1978). As we show in this article, their approaches can also be applied to cross-sectional data on students clustered in schools, when Level-2 endogeneity is present (see also Wooldridge, 2010, section 20.2.1). For example, the popular *fixed-effects* (FE) approach can be used to obtain an unbiased estimate of the coefficient of student SES by controlling for schools via dummy variables, giving what is often called the *within-school effect* of SES.

In education research, the Level-2 endogeneity problem has also been acknowledged as confounding or selection effects (e.g., Bingenheimer & Raudenbush, 2004; Cronbach, 1976; Willms, 1986), and there is a large literature on the difference in within- and between-school effects of student-level covariates (e.g., Burstein, 1980; Cronbach, 1976; Wiley, 1975). In some cases, the discussion comes under the heading "group-mean centering" (GMC; Kreft, de Leeuw, & Aiken, 1995; Pacagnella, 2006; Raudenbush, 1989), which refers to subtracting the school mean from the student covariate, such as student SES minus school mean SES. The corresponding regression coefficient represents the within-school effect of SES. If in addition, school mean SES is added as a covariate, the coefficient of school mean SES represents the between-school effect of SES. Although this including-the-group-means (IGM) approach is equivalent to the Mundlak (1978) device, a version of the "correlated random-effects approach" (Wooldridge, 2010, pp. 286–287), the applicability of such panel data approaches to cross-sectional education data has been pointed out only recently by Kim and Frees (2006) and Hanchane and Mostafa (2012). These articles emphasize estimation of the within-school coefficient because it is not affected by Level-2 endogeneity.

Although the IGM approach, unlike the FE approach, can produce estimates of the coefficients of school-level covariates, such as school type (e.g., Catholic vs. public), it is often not recognized that these estimates are biased. With this approach, the coefficients of the group mean-centered student-level covariates, or the within-school effects, are consistently estimated, but the coefficients of the school mean student-level covariates, or the between-school effects, are not because they absorb some of the school effects due to the Level-2 endogeneity. Their bias, in turn, leads to an incorrect adjustment for student-level covariates in the estimation of the coefficients of school-level covariates. The Hausman and Taylor (1981) estimator from panel data econometrics can overcome this problem, but its use in a cross-sectional setting has not been discussed, to our knowledge, except for a brief comment by Kim and Frees (2007). We thus bridge this gap by describing the applicability of this estimator to cross-sectional education data in this article. In the first two steps of the Hausman–Taylor (HT) estimator, the essential idea is to eliminate the compositional effect of SES by first estimating the coefficient of SES with the unbiased FE estimator (Step 1), so that subsequent school comparisons (Step 2) are correctly adjusted for SES. Interestingly, Raudenbush and Willms (1995, pp. 321–322) independently developed a two-step estimator of individual school effects, which corresponds closely to the initial steps of HT. We extend their two-step approach to obtain consistent estimates of the coefficients for both student-level and school-level covariates. To our knowledge, such estimation of coefficients of school-level covariates has not been formally discussed. We recommend the HT estimator, as it improves on the simpler two-step estimators, not only because it produces valid standard errors and can be more efficient but also because it can handle the additional problem of school-level covariates that are Level-2 endogenous.

While differences in the within- and between-school effects of student-level covariates can be due to Level-2 endogeneity, a competing explanation is the presence of *contextual effects*, whereby student outcomes are affected by the covariates of their peers in the same school. For instance, school mean SES may have an effect on student outcomes, even after controlling for the student's own SES (i.e., compositional effects), because the SES of peers in the same school may promote or impede student learning (i.e., contextual effects). In fact, the coefficient of school mean SES in a model that also includes student SES has traditionally been interpreted as the contextual effect (Ballou et al., 2004; Burstein, 1980; Raudenbush, 1989; Raudenbush & Willms, 1995; Wiley, 1975). In this article, we follow this work by defining contextual effects as the (true) coefficients of school means of student-level covariates in a model that also contains the corresponding student-level SES.

We discuss the problem that contextual effects cannot generally be disentangled from endogeneity bias. Unfortunately, the IGM approach for estimating contextual effects is biased in the presence of Level-2 endogeneity, and the HT estimator for handling Level-2 endogeneity produces biased school comparisons in the presence of contextual effects. Fortunately, HT-based estimates of individual school effects can be viewed as valid for what Raudenbush and Willms (1995) call "Type A" school effects, that is, the impact individual schools have on student achievement both through mechanisms they control ("school practice") and those they do not control ("school context"). In contrast, Raudenbush and Willms define "Type B" school effects as effects of school practice alone, interpretable as "school value-added" effects—the value individual schools add to student achievement through malleable practices that they can control.

We extend these concepts pertaining to the effects of *individual* schools to the problem of school-*type* comparisons by defining Type A and Type B *coefficients*. Unfortunately, estimation of Type B effects or coefficients remains unresolved, but we show the circumstances under which the IGM (or Mundlak) approach and the HT (or two-step) approach provide lower and upper bounds, respectively, for Type B coefficients, Type B effects of individual schools, and their variance. Whether the HT or IGM estimator is preferred depends on the relative magnitudes of the endogeneity bias and the contextual effect.

To disentangle the effects of endogeneity and contextual effects, we first focus on the problem of Level-2 endogeneity alone and assume that there are no contextual effects (Section 2). Subsequently, we consider *both* Level-2 endogeneity and contextual effects (Section 3). Specifically, in Section 2, we formally define our model and Level-2 endogeneity. We then describe all the estimators considered in this article, including the FE, two-step, and HT estimators, as well as two variants of the traditional random-effects (RE) approaches, GMC and including the group means (IGM). We show how the RE approaches result in biased estimators for coefficients of school-level covariates, individual school effects, and their variance if there is Level-2 endogeneity. In Section 3, we extend our investigation to the case where both Level-2 endogeneity and contextual effects exist. After reviewing the literature on school effects, we derive bias of the estimators and interpret the results for the HT and IGM estimators in terms of Type A and Type B effects and coefficients. We also show that the GMC approach is biased for school comparisons whether or not Level-2 endogeneity and/or contextual effects are present, implying that the HT and IGM estimators are preferable. In Section 4, we conclude with a discussion of our recommended approach to school comparisons and review other issues that may complicate valid school comparisons.

2. Endogenous Covariates and No Contextual Effects

Here we consider school-level (Level-2) endogeneity of student-level covariates but assume, until Section 3, that there are no contextual effects. Before discussing methods for handling covariates that are Level-2 endogenous, we first introduce the multilevel model that is assumed to be the true model throughout this section and define endogeneity.

2.1. Two-Level Random-Intercept Model and Random-Effects Estimators

Hierarchical linear models (HLMs) or linear multilevel models are useful for analyzing educational data, given the natural nesting of students within classrooms/teachers and schools. As is standard in school comparisons, we consider a two-level random-intercept model for the response y_{ij}, typically a student achievement score on a standardized test for student i $(i = 1, 2, \ldots, n_j)$ in school j $(j = 1, 2, \ldots, J)$. We first write this model in multiple stages:

$$\begin{aligned} \text{Level 1}: \quad & y_{ij} = \beta_{0j} + \beta_{1j} x_{ij} + \varepsilon_{ij}, \\ \text{Level 2}: \quad & \beta_{0j} = \alpha + \gamma w_j + u_j, \\ & \beta_{1j} = \beta_w. \end{aligned}$$

In the Level-1 model, β_{0j} is a school-specific intercept, x_{ij} is a student-level (Level-1) covariate, such as student SES, with school-specific coefficient β_{1j}, and ε_{ij} is a student-level error term with mean 0 and variance σ_ε^2. In the Level-2 model for the school-specific intercept, α is the overall intercept, w_j is a school-level (Level-2) covariate, such as school type (e.g., Catholic versus public), with coefficient γ (e.g., the school-type effect), and u_j is a school-level random intercept with mean 0 and variance σ_u^2. In a model without w_j (or $\gamma = 0$), the u_j represent the effects of individual schools and σ_u^2 the variance of these school effects. The Level-2 model for the school-specific coefficient of x_{ij} is simply an intercept β_w, so that the coefficient of x_{ij} does not vary between schools. The error terms ε_{ij} and u_j are uncorrelated with each other, u_j is uncorrelated across schools, and ε_{ij} is uncorrelated across schools and students.

The model can be written more succinctly by substituting the equations for β_{0j} and β_{1j} into the Level-1 equation,

$$y_{ij} = \alpha + \beta_w x_{ij} + \gamma w_j + u_j + \varepsilon_{ij}, \tag{1}$$

and such reduced or "mixed" forms will be used throughout the article. The model could of course include more covariates (see the Online Appendix B available at http://jeb.sagepub.com/supplemental).

Under exogeneity assumptions discussed in Section 2.2, the parameters of the HLM can be consistently estimated using maximum likelihood (ML), restricted ML (REML), or feasible generalized least squares (FGLS). For instance, in Stata (StataCorp, 2013b), one could use the commands "mixed" or "xtreg" (Rabe-Hesketh & Skrondal, 2012), and in R (R Development Core Team, 2012), the "lmer" command in the "lme4" package (Bates, Maechler, & Bolker, 2011) or the "plm" command in the "plm" package (Croissant & Millo, 2008). Using econometric terminology, we will refer to these estimators as RE estimators.

2.2. Exogeneity and Endogeneity

We view the multilevel model described previously as a "structural model," in the sense that the regression coefficients are causal parameters, and the error terms represent the effects of omitted covariates. If there are omitted confounders that are correlated with included covariates, then the error terms are correlated with the included covariates, leading to bias for standard estimators. This perspective enables us to derive omitted variable bias for different estimators. In contrast, if we interpreted the multilevel model as a "statistical model," the regression coefficients would merely represent associations or linear projections, and the error terms would be uncorrelated with all covariates by definition (e.g., Spanos, 2006; see also Lehmann, 1990).

Standard exogeneity assumptions for valid estimation of the structural model in Equation 1 using RE estimators are Level-2 exogeneity, $E(u_j|x_{1j},\ldots,x_{n_jj},w_j)=0$, and Level-1 exogeneity, $E(\varepsilon_{ij}|x_{1j},\ldots,x_{n_jj},w_j,u_j)=0$ (Wooldridge, 2010, p. 292). These mean-independence assumptions are violated if u_j and ε_{ij} are correlated with the covariates x_{ij} and w_j. We define covariates as "Level-2 endogenous" if they are correlated with the Level-2 random error term u_j and as "Level-1 endogenous" if they are correlated with the Level-1 random error term ε_{ij}.

In this article, we focus on Level-2 endogeneity of a student-level covariate x_{ij}, where x_{ij} is correlated with u_j. This is an important problem in education that occurs because students are not randomly assigned to schools. For instance, if x_{ij} is SES, high-SES parents tend to be more successful than low-SES parents at placing their children in schools expected to have positive effects u_j on future achievement, leading to a positive correlation between x_{ij} and u_j. Here student SES is neither affected by u_j nor generated jointly with u_j, as the term "endogeneity" might imply, but is determined before u_j. The correlation with u_j arises because selection of schools depends on both student SES and omitted school characteristics, such as reputation, that are correlated with u_j. When we derive the resulting biases, we will express the extent of Level-2 endogeneity in terms of the correlation $\rho_{\bar{x}u}$ between school mean SES $\bar{x}_{\cdot j}$ and u_j because it is the between-school component of SES that is correlated with u_j.

In the bulk of this article, we assume that the school-level covariate w_j (e.g., school type) is not correlated with u_j (i.e., it is Level-2 exogenous). If this assumption is violated due to school-level confounding, the inconsistent estimate of the coefficient γ of w_j can be viewed as representing an association. If the structural parameter β_w is estimated consistently, that is, if the compositional effect is taken out, then the causal effects of individual schools, $\gamma w_j + u_j$, can be estimated consistently, making the association meaningful. For instance, if w_j is a dummy variable for Catholic schools, $\widehat{\gamma}$ represents the estimated difference in the mean causal school effect between Catholic and public schools, but this difference cannot be interpreted as being caused necessarily by the schools being Catholic versus public. Note that RE estimation of γ is inconsistent even if w_j is exogenous because w_j is correlated with the Level-2 endogenous x_{ij} and thus the bias for β_w spills over to γ. For instance, as we will show, a positive correlation between x_{ij} and u_j leads to overestimation of β_w, and, if x_{ij} is positively correlated with w_j, this overadjustment for x_{ij} leads to an underestimation of γ.

Throughout the article, we maintain the standard assumption that the student and school-level covariates are Level-1 exogenous, that is, uncorrelated with ε_{ij}.

This assumption would be violated if student characteristics were omitted that affect achievement, and thus become part of the Level-1 error term ε_{ij}, and are correlated with the covariates included in the model. In other words, we assume that there are no omitted student-level confounders.

2.3. Fixed-Effects Estimator

The most common approach for handling Level-2 endogeneity of Level-1 covariates in panel data econometrics is the FE estimator. One way to obtain this estimator is to specify an estimation model that includes dummy variables for schools instead of random school intercepts u_j and use ordinary least squares (OLS; i.e., analysis of covariance with school as a "factor"). This approach controls for schools and any school-level variables that could be the cause of endogeneity.

We use the term "estimation model" in this article to distinguish the model that is estimated from the assumed structural model. The estimation model often differs deliberately from the assumed model to obtain consistent estimators. For instance, adopting a FE approach does not necessarily mean that the investigator believes the school effects are fixed.

We illustrate the FE estimation model in Figure 1a, and use Figure 1 throughout the article to illustrate each estimation model in turn. For the SES and achievement example, Figure 1a shows the parallel school-specific regression lines (gray dashed lines) for a hypothetical set of 10 schools, where the slope of x_{ij} is interpretable as the within-school effect of x_{ij}. The implied relationship between $\bar{y}_{\cdot j}$ and $\bar{x}_{\cdot j}$, with slope β_w (black solid line), represents the *compositional* effect of SES, that is, the differences in school achievement that are purely due to their composition when schools themselves have no effect.

We can also implement the FE estimator by subtracting the school mean of the random-intercept model, $\bar{y}_{\cdot j} = \alpha + \beta_w \bar{x}_{\cdot j} + \gamma w_j + u_j + \bar{\varepsilon}_{\cdot j}$, from the random-intercept model (Equation 1) to obtain the estimation model

$$y_{ij} - \bar{y}_{\cdot j} \;=\; \beta_w (x_{ij} - \bar{x}_{\cdot j}) + \varepsilon^*_{ij}, \tag{2}$$

where $\bar{y}_{\cdot j} = \frac{1}{n_j}\sum_{i=1}^{n_j} y_{ij}$ (analogously for $\bar{x}_{\cdot j}$ and $\bar{\varepsilon}_{\cdot j}$) and $\varepsilon^*_{ij} = \varepsilon_{ij} - \bar{\varepsilon}_{\cdot j}$. Now β_w can be consistently and unbiasedly estimated by OLS because $E(\varepsilon^*_{ij} | x_{ij} - \bar{x}_{\cdot j}) = 0$ follows from the Level-1 exogeneity assumption. Note that an alternative estimation model with x_{ij} in the structural model replaced by $x_{ij} - \bar{x}_{\cdot j}$ yields an identical estimator for β_w.

Under normality of ε_{ij}, this estimator corresponds to the conditional ML estimator (Chamberlain, 1980). Unfortunately, w_j cancels out of the estimation model, making it impossible to estimate γ. In the school dummy variable approach, it is also obvious that γ cannot be estimated because w_j is perfectly collinear with the dummy variables.

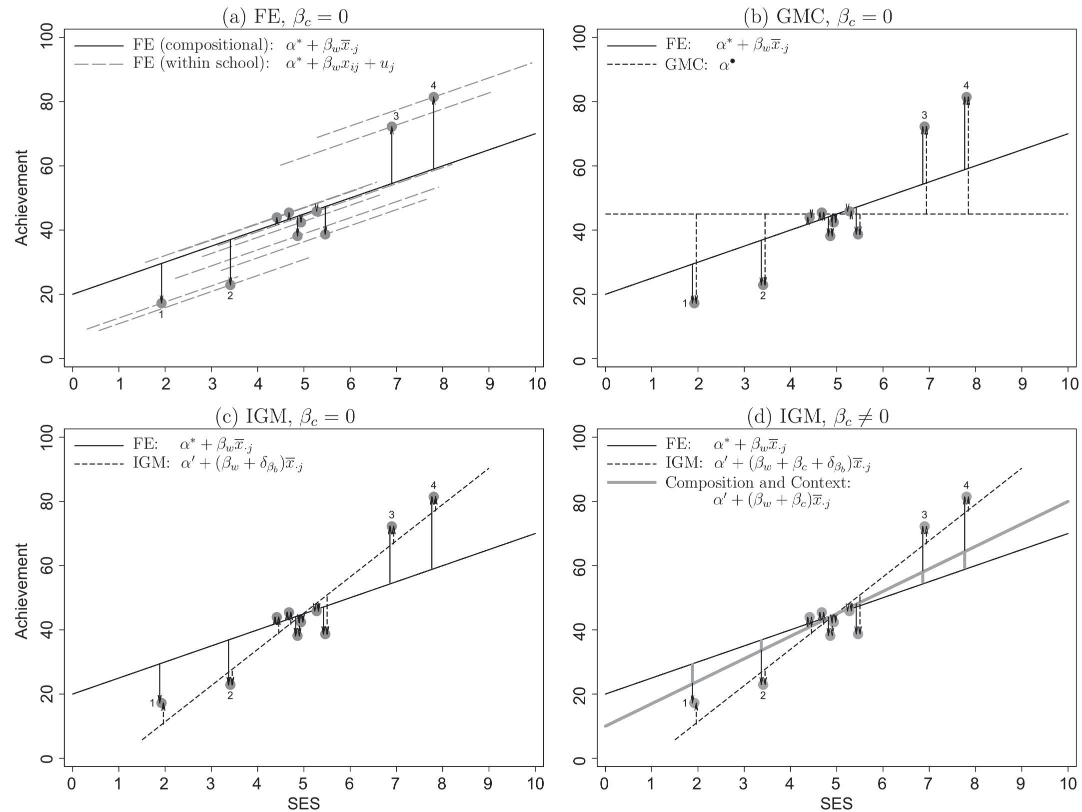

FIGURE 1. *Illustration of FE (or HT/2S), GMC, and IGM estimators for the relationship between school mean achievement $\bar{y}_{\cdot j}$ and school mean SES $\bar{x}_{\cdot j}$, as well as the corresponding school effects. Panel (d) is for the case with contextual effects $\beta_c \neq 0$, and δ_{β_b} is different for panels (c) and (d).*
Note. FE = fixed-effects; HT = Hausman–Taylor; 2S = two-step; GMC = group-mean centering; IGM = including the group means.

As pointed out by Baltagi (2013) and Kim and Swoboda (2011), researchers typically debate the choice of FE or RE estimators, which correspond to "all or nothing" approaches to treating student covariates as Level-2 endogenous. FE estimators allow all student covariates to be Level-2 endogenous, but at the expense of making coefficients for any school-level covariates inestimable. In contrast, RE estimators rely on the often unrealistic assumption that none of the covariates are endogenous, leading to biased estimation if any Level-2 endogeneity is present. As group sizes n_j increase, RE estimators (ML, REML, and FGLS) of β_w converge to the FE estimator (Maddala, 1971), but often group sizes are not large enough in school and teacher comparisons. Researchers therefore sometimes specify estimation models that differ from the two-level random-intercept model given in Equation 1 either by group-mean centering student-level covariates or by including the group means of the student covariates. However, we show that neither of these approaches produces unbiased or consistent estimates of the regression coefficients for school-level covariates (e.g., a school-type effect γ).

The HT approach addresses this problem and provides consistent estimates of all coefficients as long as Level-2 exogeneity assumptions are valid for some of the covariates and all covariates are Level-1 exogenous. Before describing this approach, we explain a simpler and more intuitive two-step (2S) approach from educational statistics that is extended in the HT approach.

2.4. Two-Step Estimator

The 2S approach is based on rewriting Equation 1 as

$$y_{ij} - \beta_w x_{ij} = \alpha + \gamma w_j + u_j + \varepsilon_{ij}. \tag{3}$$

Because we assume that w_j is Level-2 exogenous, $\mathrm{Cor}(w_j, u_j) = 0$, the coefficient γ of w_j and the variance of u_j can be estimated consistently by estimating the model above.

In practice, we do not know β_w. In the first step, we therefore estimate β_w consistently using the FE estimator as explained in the previous subsection. For the second step, we substitute this estimate $\widehat{\beta}_w$ in Equation 3, so the response variable becomes a "quasi-residual" in the terminology of Ballou et al. (2004). Consistent estimates of γ and the random-intercept variance σ_u^2 are then obtained by fitting Equation 3 with $\widehat{\beta}_w$ substituted for β_w using ML, REML, or FGLS. Consistency follows from the theory of quasi-generalized extremum estimation (Gourieroux & Monfort, 1995) called pseudo ML estimation (Gong & Samaniego, 1981) if ML is used in the second step.

Although the 2S method provides consistent estimators of γ and σ_u^2, the standard error of $\widehat{\gamma}$ is underestimated because $\widehat{\beta}_w$ is treated as known in the second step. The HT approach described in Section 2.5 rectifies this problem and can also be more efficient when there are additional exogenous variables.

Historically, Wiley (1975, pp. 238–239) first proposed the 2S approach as part of a three-step procedure, and this work was discussed by Burstein (1980). Raudenbush and Willms (1995, pp. 321–322) propose the 2S method for estimating β_w and u_j in Equation 1 when there is no school-level covariate w_j. Ballou et al. (2004) use the Raudenbush and Willms (1995) approach in a teacher effectiveness study to estimate coefficients for teacher-level covariates as well. In a related two-stage approach (e.g., Goldhaber & Brewer, 1997; Buddin & Zamarro, 2009), OLS estimates of school fixed effects are obtained in Step 1 by fitting an estimation model with student covariates and dummy variables for schools, and these estimates are then regressed on school covariates in Step 2. As the estimated fixed effects are identical to the school means $\bar{y}_{\cdot j} - \widehat{\beta}_w \bar{x}_{\cdot j}$ of the quasi-residuals, this approach yields numerically identical estimates of the coefficients of teacher covariates as the 2S approach if the schools are weighted by the number of students n_j in Step 2. In an attempt to obtain correct standard errors, Plümper and Troeger (2007) add a third step that has been shown to be flawed (Breusch, Ward, Nguyen, & Kompos, 2011; Greene, 2011).

The central idea of the 2S estimator is that the *compositional effect* of x_{ij} is subtracted from $\bar{y}_{\cdot j}$ in the school-mean quasi-residuals to obtain consistent estimates of school effects. Figure 1a illustrates this using hypothetical data on achievement and SES for 10 schools. The black solid line is the FE regression line from Step 1 of the 2S approach. It represents the estimate of the school-mean achievement we would expect based only on the composition of the schools in terms of SES, if the schools themselves had no effects (i.e., as if $\gamma = 0$, $u_j = 0$). Because student SES affects student achievement (with coefficient β_w), we expect schools with higher mean SES to have higher mean achievement (with the same coefficient β_w). The deviations of the actual school means $\bar{y}_{\cdot j}$ (gray dots) from the estimates of these expectations are shown as black arrows and represent the school mean quasi-residuals. If there were no sampling variability (infinitely many students in each school), these quasi-residuals would be the true school effects $\gamma w_j + u_j$, apart from an arbitrary constant.

2.5. Hausman–Taylor Estimator

The HT approach is described in detail in Online Appendix A (available at http://jeb.sagepub.com/supplemental) for the general case with many covariates, but we describe the ideas as they apply to the model considered here. The first two steps (A.1 and A.2) are equivalent to the 2S method, except that OLS is used in the second step to estimate the regression coefficients, and moment estimators are used for the variance components σ_u^2 and σ_ε^2. For our model, the advantage of two further steps (A.3 and A.4) is that the coefficients β_w and γ are reestimated simultaneously, so that uncertainty in the estimate of β_w is not ignored when estimating the standard error for $\widehat{\gamma}$ as in the 2S approach.

Instrumental variable estimation is used to handle Level-2 endogeneity. However, the usual two-stage least squares instrumental variable estimator cannot be used because the total errors $v_{ij} = u_j + \varepsilon_{ij}$ are correlated within schools. Step A.3 eliminates these correlations by partial school mean centering all variables in the model, which is known as the GLS or Fuller–Battese transformation (Fuller & Battese, 1973). OLS estimation of the estimation model with the transformed variables would correspond to GLS (with covariance matrix treated as known), and two-stage least squares can now be used. In Step A.4, the deviation score $\left(x_{ij} - \bar{x}_{.j}\right)$ is used as an instrument for x_{ij} because it satisfies the following two requirements for an instrumental variable: (1) it is correlated with x_{ij} and (2) it is Level-2 exogenous because, by construction, its group mean is zero and therefore uncorrelated with u_j (see also Wooldridge, 2010, p. 360). The school covariate w_j also satisfies both requirements and serves as an additional instrument.

For models that include additional exogenous and Level-2 endogenous covariates, all school-level exogenous covariates, deviation scores of all student-level covariates, and group means of all exogenous student-level covariates serve as instruments for all Level-2 endogenous covariates, providing more efficient estimates than the 2S estimator. An additional advantage of HT is that it can also handle Level-2 endogenous *school-level* covariates as long as there are at least as many exogenous student-level covariates as endogenous school-level covariates.

Interestingly, the connection between the 2S and HT approaches does not appear to have been pointed out before. In an article about school effects published in an economics journal, Goldhaber and Brewer (1997) use the 2S approach and mention in a footnote that the HT approach is *another* possible solution (p. 510), but without connecting the methods.

The HT approach is more complex than the 2S approach but is fortunately implemented in the "xthtaylor" command in Stata (StataCorp, 2013a) and the "pht" function in the "plm" R package (Croissant & Millo, 2008). Unlike the "pht" R command, the "xthtaylor" Stata command requires that there is at least one exogenous student-level covariate.

2.6. Biased Estimators

We now consider two approaches that are popular in education, GMC and IGM, that produce biased estimators of γ and σ_u^2 in the structural model in Equation 1. We allow for Level-2 endogeneity of the student covariate x_{ij}, Cor(x_{ij}, u_j) $\neq 0$, which also implies that $\mathrm{Cor}(\bar{x}_{.j}, u_j) \equiv \rho_{\bar{x}u} \neq 0$. As before, the school covariate w_j is assumed to be Level-2 exogenous, $\mathrm{Cor}(w_j, u_j) \equiv \rho_{wu} = 0$.

First, it is helpful to rewrite the structural model in Equation 1 as

$$y_{ij} \;=\; \alpha + \beta_w(x_{ij} - \bar{x}_{.j}) + \beta_w\bar{x}_{.j} + \gamma w_j + u_j + \varepsilon_{ij}. \tag{4}$$

TABLE 1

Inconsistency Terms for γ, σ_u^2, and u_j for Different Methods When (1) Only Level-2 Endogeneity Is Present ($\rho_{xu} \neq 0$, $\beta_c = 0$) and (2) Both Level-2 Endogeneity and Contextual Effects are Present ($\rho_{xu} \neq 0$, $\beta_c \neq 0$)

Condition	Estimand	Group-Mean Centering (GMC)	Including the Group Means (IGM)	Two-Step (2S) or Hausman–Taylor (HT)
	γ	$\frac{\beta_w \rho_{\bar{x}w} \sigma_{\bar{x}}}{\sigma_w}$	$\frac{-\rho_{\bar{x}u}\rho_{\bar{x}w}\sigma_u}{\sigma_w\left(1-\rho_{\bar{x}w}^2\right)}$	0
$\beta_c = 0$	σ_u^2	$\beta_w^2 \sigma_{\bar{x}}^2 \left(1 - \rho_{\bar{x}w}^2\right) + 2\beta_w \rho_{\bar{x}u} \sigma_{\bar{x}} \sigma_u$	$-\frac{\rho_{\bar{x}u}^2 \sigma_u^2}{1-\rho_{\bar{x}w}^2}$	0
	u_j	$\beta_w\left(\bar{x}_{.j} - \bar{x}_{..}\right)$	$-\frac{\rho_{\bar{x}u}\sigma_u}{\sigma_{\bar{x}}}\left(\bar{x}_{.j} - \bar{x}_{..}\right)$	0
	γ	$\frac{(\beta_w+\beta_c)\rho_{\bar{x}w}\sigma_{\bar{x}}}{\sigma_w}$	$\frac{-\rho_{\bar{x}u}\rho_{\bar{x}w}\sigma_u}{\sigma_w\left(1-\rho_{\bar{x}w}^2\right)}$	$\frac{\beta_c \rho_{\bar{x}w}\sigma_{\bar{x}}}{\sigma_w}$
$\beta_c \neq 0$	σ_u^2	$(\beta_w + \beta_c)^2 \sigma_{\bar{x}}^2 \left(1 - \rho_{\bar{x}w}^2\right) + 2(\beta_w + \beta_c)\rho_{\bar{x}u}\sigma_{\bar{x}}\sigma_u$	$-\frac{\rho_{\bar{x}u}^2 \sigma_u^2}{1-\rho_{\bar{x}w}^2}$	$\beta_c^2 \sigma_{\bar{x}}^2 \left(1 - \rho_{\bar{x}w}^2\right) + 2\beta_c \rho_{\bar{x}u}\sigma_u \sigma_{\bar{x}}$
	u_j	$(\beta_w + \beta_c)\left(\bar{x}_{.j} - \bar{x}_{..}\right)$	$-\frac{\rho_{\bar{x}u}\sigma_u}{\sigma_{\bar{x}}}\left(\bar{x}_{.j} - \bar{x}_{..}\right)$	$\beta_c\left(\bar{x}_{.j} - \bar{x}_{..}\right)$

Note. The assumed structural model is $y_{ij} = \alpha + \beta_w x_{ij} + \beta_c \bar{x}_{.j} + \gamma w_j + u_j + \varepsilon_{ij}$, with Level-2 endogeneity if $\rho_{xu} > 0$, and a contextual effect if $\beta_c \neq 0$. Results for u_j are for the case where $\gamma = 0$ and represent the conditional bias $E\left(\widehat{u}_j | u_j, \ \boldsymbol{x}\right) - u_j$. Note that only the IGM method estimates β_c, and this estimate is biased, with $\delta_{\beta_c} = \frac{(\rho_{\bar{x}u} - \rho_{wu}\rho_{\bar{x}w})\sigma_u}{\sigma_{\bar{x}}\left(1-\rho_{\bar{x}w}^2\right)}$.

Note that the coefficient of $(x_{ij} - \bar{x}_{.j})$ is identical to the coefficient of $\bar{x}_{.j}$, whereas the GMC estimation model sets the coefficient of $\bar{x}_{.j}$ to zero and IGM specifies two separate coefficients for $(x_{ij} - \bar{x}_{.j})$ and $\bar{x}_{.j}$.

We discuss expressions for the inconsistency and bias of ML estimators for the regression coefficients β_w and γ for the GMC and IGM approaches that follow from the results derived in the Appendix and Online Appendix B (available at http://jeb.sagepub.com/supplemental). Because identical results are obtained for inconsistency and bias, we just refer to bias for simplicity. The bias terms discussed in this section are for the case where each school has the same number of students in the sample, $n_j = n$, but the qualitative findings also hold when n_j varies between schools. All expressions given in this section are also shown in the upper part of Table 1 and can be obtained by substituting $\beta_c = 0$ and $\rho_{wu} = 0$ in the equations in the Appendix.

We also consider the setting in which we want to estimate effects of individual schools but not control for w_j—the scenarios in which the interest is in determining the effectiveness of individual schools, such as value-added analyses, instead of comparing schools by their characteristics w_j. In this case, OLS estimates $\widehat{u}_j$ of the school effects can be obtained as the school means of the total residuals $y_{ij} - \widehat{\mu}_{ij}$, where $\widehat{\mu}_{ij}$ is the estimated fixed part of the model with covariate $(x_{ij} - \bar{x}_{.j})$ for GMC and with an additional covariate $\bar{x}_{.j}$ for IGM. The Appendix provides expressions for the conditional bias $\delta_u = E(\widehat{u}_j | u_j, \boldsymbol{x}) - u_j$, given the true random intercept u_j and the covariate values $\boldsymbol{x}$ for the entire sample. Empirical Bayes predictions are just shrunken versions of OLS estimates (e.g., Raudenbush & Bryk, 2002, p. 90; Skrondal & Rabe-Hesketh, 2009) and will not be considered here. For the random-intercept variance, we derive inconsistencies, which are different from finite-sample bias, in the Appendix. Results for the model without w_j can be obtained by setting any terms involving w_j to 0.

2.6.1. Centering student covariates. Researchers often substitute the GMC covariate $x_{ij} - \bar{x}_{.j}$ for the student-level covariate x_{ij} in order to estimate the within-school effect of x_{ij}. Such centering has the advantage that the corresponding school-specific intercept β_{0j} in the Level-1 model described in Section 2.1 becomes interpretable as the adjusted school mean, instead of the school intercept (Raudenbush & Bryk, 2002, p. 33) and that it produces a consistent estimator of β_w. The GMC approach has been used to estimate the coefficients of teacher-level variables (e.g., Von secker & Lissitz, 1999) and in school effectiveness research (e.g., Kennedy & Mandeville, 2000, p. 198). Unfortunately, omitting the group mean $\bar{x}_{.j}$ from the estimation model, which amounts to assuming that there are no between-school effects, means that student covariates are not controlled for when making school comparisons, a point also recently made by Steedle (2012).

Coefficient of the school-level covariate. As shown in the Appendix, the bias term for the coefficient γ of the school-level covariate for the GMC approach is:

$$\delta_{\gamma} = \frac{\beta_w \rho_{\bar{x}w} \sigma_{\bar{x}}}{\sigma_w}. \tag{5}$$

This bias can be recognized as the usual "omitted variable bias," here due to omitting $\bar{x}_{.j}$ from the estimation model. As expected, the bias increases as the correlation between w_j and $\bar{x}_{.j}$ increases. Importantly, the bias occurs regardless of whether Level-2 endogeneity is present as $\rho_{\bar{x}u}$ does not appear in the bias term. Note that β_w is estimated without bias because the school mean of $x_{ij} - \bar{x}_{.j}$ is uncorrelated with w_j, u_j, and ε_{ij}.

School-level random-intercept variance. The inconsistency term for the school-level random-intercept variance is

$$\delta_{\sigma_u^2} = \beta_w^2 \sigma_{\bar{x}}^2 \left(1 - \rho_{\bar{x}w}^2\right) + 2\beta_w \rho_{\bar{x}u} \sigma_{\bar{x}} \sigma_u. \tag{6}$$

The first term is due to failing to control for $\bar{x}_{.j}$, even in the absence of Level-2 endogeneity, and the second term is due to endogeneity (i.e., disappears if $\rho_{\bar{x}u} = 0$). The first term decreases when $|\rho_{\bar{x}w}|$ increases because w_j increasingly behaves like a proxy for $\bar{x}_{.j}$.

School effects. If the model does not include the school-level covariate w_j, but we are interested in comparing individual schools by estimating u_j, it follows from the Appendix that the conditional bias of the OLS estimator $\widehat{u}_j$, given u_j and $\mathbf{x}$, is

$$\delta_{u_j} = \beta_w \left(\bar{x}_{.j} - \bar{x}_{..}\right), \tag{7}$$

where $\bar{x}_{..}$ is the grand mean of x_{ij}. The bias term δ_{u_j} is just the compositional effect, as we show in Figure 1b, which is for the same hypothetical data as in Figure 1a. The dashed horizontal line represents the relationship between school mean achievement $\bar{y}_{.j}$ and school mean SES $\bar{x}_{.j}$ implied by the GMC approach. The slope is zero because GMC omits $\bar{x}_{.j}$ from the estimation model. Therefore, the deviations of school mean achievement $\bar{y}_{.j}$ from the GMC line, shown as dashed arrows, include the compositional effects, which are represented by the portions of the dashed arrows between the dashed GMC line and the solid FE line (reproduced from Figure 1a). As in Figure 1a, the corresponding unbiased FE estimates of the school effects are given by the black arrows. Accordingly, using the GMC approach, schools with mean SES above the grand mean (such as Schools 3 and 4 in the figure) will be ranked higher than they should be and vice versa for schools with mean SES below the grand mean (such as Schools 1 and 2).

This could result in an incorrect distribution of resources and accolades/sanctions to schools.

2.6.2. Including the group means of student covariates. Another approach for making school comparisons is the IGM approach, sometimes referred to as the "contextual-effects model" (Raudenbush, 1989) or the Mundlak device (Mundlak, 1978). For this approach, different coefficients β_w and β_b are estimated for the deviation scores, $x_{ij} - \bar{x}_{.j}$, and the school mean $\bar{x}_{.j}$ of the student-level covariate, respectively. As $\beta_w(x_{ij} - \bar{x}_{.j}) + \beta_b \bar{x}_{.j} = \beta_w x_{ij} + (\beta_b - \beta_w)\bar{x}_{.j}$, an equivalent estimation model is obtained by including x_{ij} and $\bar{x}_{.j}$ as covariates, with coefficients β_w and $\beta_c = \beta_b - \beta_w$, respectively. For example, Rockoff, Jacob, Kane, and Staiger (2011) include individual student characteristics and their classroom and school averages to estimate the effects of teacher characteristics on student outcomes. Palardy and Rumberger (2008) also include several student covariates and their classroom averages in their "classroom compositional" teacher effects model. Similarly, Hanchane and Mostafa (2012) include the school mean of each student-level covariate in their model to estimate effects of school-level covariates in the presence of Level-2 endogenous student covariates.

Like GMC, the IGM approach produces an unbiased estimator for β_w (Mundlak, 1978). However, it is often not recognized that IGM produces a biased estimator for γ. Hanchane and Mostafa (2012) acknowledge that their approach "may still suffer from endogeneity bias" (p. 1112), but they do not offer a solution. We now briefly discuss the bias terms for the IGM approach (see the Appendix for derivations).

Coefficient of school-level covariate. The bias term for the coefficient β_b of the school-mean covariate is

$$\delta_{\beta_b} = \frac{\rho_{\bar{x}u}\sigma_u}{\sigma_{\bar{x}}(1 - \rho^2_{\bar{x}w})}, \tag{8}$$

and the bias term for γ is

$$\delta_\gamma = \frac{-\rho_{\bar{x}u}\rho_{\bar{x}w}\sigma_u}{\sigma_w(1 - \rho^2_{\bar{x}w})}. \tag{9}$$

For positive $\rho_{\bar{x}u}$, the coefficient β_b of $\bar{x}_{.j}$ is overestimated, which results in *overcontrolling* for $\bar{x}_{.j}$ if $\beta_w > 0$. If $\rho_{\bar{x}w} > 0$, then overcontrolling for $\bar{x}_{.j}$ results in underestimating the coefficient of w_j, so $\delta_\gamma < 0$. This relationship between the bias for β_b and the bias for γ can also be understood by noting that $\delta_\gamma = -\delta_{\beta_b}\rho_{\bar{x}w}$. An example would be underestimating the difference in effectiveness between private and public schools ($\delta_\gamma < 0$) by overcontrolling for SES $\left(\delta_{\beta_b} > 0\right)$ because for each type of school, high-SES students are more likely to be placed in those schools that will produce better outcomes with positive u_j (so that $\rho_{\bar{x}u} > 0$).

School-level random-intercept variance. The inconsistency for the random-intercept variance is

$$\delta_{\sigma_u^2} = -\frac{\rho_{\bar{x}u}^2 \sigma_u^2}{1 - \rho_{\bar{x}w}^2}, \tag{10}$$

which is negative if $\rho_{\bar{x}u} \neq 0$, (and $\sigma_u^2 > 0, |\rho_{\bar{x}w}| < 1$). In this case, the unexplained between-school variance can only decrease when the coefficient β_b of $\bar{x}_{\cdot j}$ is unconstrained because more of the total between-school variance becomes explained by the covariates.

School effects. For individual school comparisons with no consideration of school characteristics (i.e., no w_j), the conditional bias of the OLS estimator $\widehat{u}_j$, given u_j and $\mathbf{x}$, is

$$\delta_{u_j} = -\frac{\rho_{\bar{x}u}\sigma_u}{\sigma_{\bar{x}}}\left(\bar{x}_{\cdot j} - \bar{x}_{\cdot\cdot}\right). \tag{11}$$

In contrast to the GMC approach, school effects are underestimated, in absolute value, if $\rho_{\bar{x}u} > 0$. This is illustrated in Figure 1c, where the relationship between school-mean SES and school-mean achievement for the IGM approach is shown as a dashed line. The slope of this dashed line is greater than β_w (slope of the solid FE line) because the school effects u_j are positively correlated with SES, leading to a positive bias, δ_{β_b}. Consequently, what is subtracted from $\bar{y}_{\cdot j}$ to estimate u_j exceeds the compositional effect in absolute value. This phenomenon is shown in the figure by the dashed arrows from the IGM line being shorter than the unbiased school effect estimates represented by the solid arrows from the FE line. Therefore, schools with large school mean SES (such as Schools 3 and 4) will receive lower rankings than they should and vice versa for schools with small means (such as Schools 1 and 2).

3. Endogenous Covariates and Contextual Effects

Here, we extend the structural model of Section 2 (Equation 1) to also include a contextual effect. To motivate this extension, we first introduce the idea of contextual effects and discuss the distinction between effects due to school composition, school context, and school practice.

3.1. School Composition, Context, and Practice

To define school composition, school context, and school practice effects, we will consider SES as the student-level covariate of interest, but note that other student covariates could be used, such as scores on previous assessments.

As previously noted, *compositional effects* are due to the effect of students' SES on their own outcomes, combined with differences in school mean SES (Bingenheimer & Raudenbush, 2004; Burstein, 1980; Duncan et al.,1998;

Raudenbush & Willms, 1995). It is important to control for compositional effects, so that schools are not unfairly disadvantaged or advantaged due to their composition, as shown for the GMC approach in Section 2.6.1.

In education (and public health), the school (or neighborhood) *contextual effect* of SES is the effect that high- versus low-SES environments/contexts have on individual educational (or health) outcomes (Bingenheimer & Raudenbush, 2004; Burstein, 1980; Duncan et al., 1998; Raudenbush & Willms, 1995). Specifically, the contextual effect (for SES) is often defined as the coefficient β_c of group-mean SES, $\bar{x}_{\cdot j}$, in a model that also includes individual SES, x_{ij},

$$y_{ij} = \alpha + \beta_w x_{ij} + \beta_c \bar{x}_{\cdot j} + u_j + \varepsilon_{ij}. \tag{12}$$

Burstein (1980) describes two processes—psychological and sociological—that might give rise to contextual effects in education. The psychological, or opportunity to learn, explanation, is that the average SES of a school may affect the style of instruction of its teachers which in turn affects individual achievement. The sociological, or normative climate, explanation is that the average SES level creates a climate that affects the individual student's motivation to learn and hence affects his or her individual achievement level.

We use the classical definition of contextual effect as the coefficient β_c of $\bar{x}_{\cdot j}$ in Equation 12, noting that context can also be represented by a different school-aggregate of x_{ij}, such as the standard deviation (Plewis, 1989), the latent school (super) population mean (Lütke et al., 2008; Shin & Raudenbush, 2010), the observed school (finite) population mean from administrative data, or by characteristics of the school neighborhood (e.g., crime rate). Manski (1993) emphasizes that the influence of peer characteristics is difficult to distinguish from the influence of peer outcomes. The influence of peers may also operate through the "frog-pond" effect (Burstein, 1980) or "big-fish–little-pond" effect (Marsh et al., 2008), where student's motivation to learn, and hence achievement, increases with the difference between the SES (or previous achievement) of the student and that of their peers, $x_{ij} - \bar{x}_{\cdot j}$. However, frog-pond effects cannot be distinguished from contextual effects because $x_{ij} - \bar{x}_{\cdot j}$, x_{ij}, and $\bar{x}_{\cdot j}$ are linearly dependent (Burstein, 1980). In this article, we do not address these challenges.

In addition to compositional and contextual effects, there are what Raudenbush and Willms (1995) call "school practice effects" due to the specific leadership and instructional activities of the school administrators and teachers. These practice effects are distinct from school context effects that are outside the control of the school. They are typically not modeled directly because school practice variables are unknown or intangible and are therefore unobserved effects that become part of the school-level random intercept u_j given in Equation 12.

In the estimation of school effects, Raudenbush and Willms (1995) emphasize the importance of determining whether the effect of interest is the combination of school practice and context effects—the "Type A effect"—or the effect of school practice alone—the "Type B effect." Type A school effects are generally of interest to parents in "school choice" decisions as parents are not concerned whether the practices of the school or the peers in the school are responsible for the academic achievement outcomes of their own child. Specifically, Type A school effects are due to the composite of school context effects $\beta_c\bar{x}_{.j}$ and practice effects u_j and can be estimated by adding these effects together or more easily by subtracting the estimated school compositional effect $\widehat{\beta}_w\bar{x}_{.j}$ from the school-mean outcome,

$$\widehat{A_j} \ = \ \widehat{\beta}_c\bar{x}_{.j} + \widehat{u_j} = \bar{y}_{.j} - \widehat{\alpha} - \widehat{\beta}_w\,\bar{x}_{.j}.$$

Raudenbush and Willms (1995) demonstrate how to obtain a consistent estimate of A_j and its variance by consistently estimating β_w. This corresponds directly to the 2S approach we described in Section 2.4, as will be shown in Section 3.2.3. When the model also includes a school covariate w_j, the 2S or HT approaches can be viewed as estimating what we call the Type A *coefficient* of w_j.

Type B effects, due to school practice alone, are of interest to policy makers and school administrators who want to hold schools accountable only for practices directly under their control. Estimation of Type B school effects requires also subtracting the estimated contextual effects $\widehat{\beta}_c\bar{x}_{.j}$ from $\bar{y}_{.j}$. However, as we will show in Sections 3.2.2 and 3.2.3, contextual effects can generally not be consistently estimated in the presence of Level-2 endogeneity. When the model includes a school-level covariate w_j, the composite of the school covariate and school practice effect, $\gamma w_j + u_j$, can be interpreted as the Type B effect of school j. Moreover, we interpret the coefficient γ of the school-level covariate as the Type B coefficient, which is a Type B *school-type* effect, if w_j is an indicator variable for school type. When there is no w_j in the model, σ_u^2 is the variance of the Type B school effects. We show in Section 3.2.3 that the 2S or HT and the IGM approaches provide upper and lower bounds for these parameters, respectively.

3.2. Contextual-Effects Model and Data Example

We now extend the structural model considered in Section 2 by adding a contextual effect β_c:

$$\begin{aligned} y_{ij} &= \alpha + \beta_w x_{ij} + \beta_c\bar{x}_{.j} + \gamma w_j + u_j + \varepsilon_{ij} \\ &= \alpha + \beta_w(x_{ij} - \bar{x}_{.j}) + \beta_b\bar{x}_{.j} + \gamma w_j + u_j + \varepsilon_{ij}, \end{aligned} \tag{13}$$

where $\beta_b = \beta_w + \beta_c$. As before, we allow for x_{ij} to be Level-2 endogeneous, but assume that w_j is Level-2 exogenous and that both covariates are Level-1 exogenous. The structural model of Section 2 is obtained by setting $\beta_c = 0$.

TABLE 2
Comparing Methods[a] Using the High School and Beyond Data

	Standard Random-Effects (RE)		Group-Mean Centering (GMC)		Including the Group Means (IGM)		Two-Step (2S)		Hausman–Taylor (HT)	
	Est	(*SE*)	Est	(*SE*)	Est	(*SE*)	Est	(*SE*)	Est	(*SE*)
					Fixed part					
Intercept	11.72	(0.23)	11.39	(0.29)	12.13	(0.20)	11.69	(0.23)	11.69	(0.23)
SES (x_{ij})	2.37	(0.11)					2.19	(0.11)	2.19	(0.11)
DevSES $\left(x_{ij} - \bar{x}_{.j}\right)$			2.19	(0.11)	2.19	(0.11)				
MeanSES $\left(\bar{x}_{.j}\right)$					5.34	(0.37)				
Sector (w_j)	2.10	(0.34)	2.81	(0.44)	1.22	(0.31)	2.16	(0.35)	2.16	(0.34)
					Random part					
School (σ_u^2)	3.69		6.74		2.37		3.86		3.58	
Student (σ_ε^2)	37.04		37.01		37.02		37.02		37.00	

[a]The Hausman–Taylor estimates are computed using the R function "pht" (in the "plm" package) with the modification in the variance computations given in Step A.3 in Online Appendix A (available at http://jeb.sagepub.com/supplemental). All other estimators are estimated using REML with the "lmer" R function in the "lme4" package. See Online Appendix C for R code (available at http://jeb.sagepub.com/supplemental).

To ground our comparisons of the different approaches when estimating this model, we use the well-known Raudenbush and Bryk (2002) data from the 1982 High School and Beyond (HSB) survey. This subsample from the full HSB survey contains 7,185 students nested within 160 schools: 3,642 students in 90 public schools and 3,543 students in 70 Catholic schools with school sizes ranging from 14 to 67 students ($M = 45$, $SD = 12$). For this data set, the response variable y_{ij} is a mathematics standardized test achievement score ($M = 12.75$, $SD = 6.88$) for student i in school j, x_{ij} is a continuous student SES index—a composite of parental education, parental occupation, and parental income ($M = 0$, $SD = 0.78$)—and w_j indicates whether school j is Catholic ($w_j = 1$) or public ($w_j = 0$).

As mentioned previously, GMC, IGM, 2S, and HT are consistent estimators for β_w. When applied to the HSB data, it is shown in Table 2 that all these methods (as well as FE, not shown) produce an estimate of 2.19 with an estimated standard error of about 0.11 (coefficient of DevSES $\left(x_{ij} - \bar{x}_{.j}\right)$ for GMC and IGM and the coefficient of SES x_{ij} for 2S and HT). We also include the estimates from the standard RE approach in Table 2 to show how the estimates differ when neither Level-2 endogeneity nor contextual effects are taken into account. For the standard RE approach (first column), the estimate of 2.37 is a little larger because this estimate is a weighted mean of the within-effect estimate of 2.19 and the between-effect estimate of 5.34 obtained by the IGM approach (Maddala, 1971).

The bias for γ, inconsistency for σ_u^2, and conditional bias for u_j for each approach are derived in the Appendix and are found by substituting $\rho_{wu} = 0$ into those derivations. The results are also presented in the lower part of Table 1 (for $\beta_c \neq 0$). When discussing these expressions, we will consider the reasonable scenario that SES has a positive contextual effect, $\beta_c > 0$, and is positively correlated with the school random intercept, $\rho_{\bar{x}u} > 0$.

3.2.1. Centering student covariates. We introduced the GMC approach in Section 2 as a common method used in school comparison analyses. We showed that it produces bias for the coefficients of the school-level covariates and inconsistency for the school-level random-intercept variance even if the student-level covariates are exogenous. In the presence of contextual effects and Level-2 endogeneity of x_{ij}, the bias for γ becomes

$$\delta_\gamma = \frac{(\beta_w + \beta_c)\rho_{\bar{x}w}\sigma_{\bar{x}}}{\sigma_w}. \tag{14}$$

This bias term differs from that given in Equation 5 only by the addition of β_c, the contextual effect, in the numerator. Generally, when β_c has the same sign as β_w, the absolute bias of γ increases with increasing $|\beta_c|$. Moreover, γ is overestimated whenever $\beta_w + \beta_c > 0$. Indeed, for the HSB example, Table 2 shows that the GMC approach (second column) produces the highest estimate of γ of 2.81.

As just seen for the bias term for γ, the only change to the inconsistency term for the school-level random-intercept variance from Section 2 (Equation 6) is that β_w is replaced with $\beta_b = \beta_w + \beta_c$:

$$\delta_{\sigma_u^2} = (\beta_w + \beta_c)^2\sigma_{\bar{x}}^2\left(1 - \rho_{\bar{x}w}^2\right) + 2(\beta_w + \beta_c)\rho_{\bar{x}u}\sigma_{\bar{x}}\sigma_u. \tag{15}$$

This expression includes both β_c and $\rho_{\bar{x}u}$, so both contextual effects and Level-2 endogeneity contaminate the school-level random-intercept variance estimate. If $(\beta_w + \beta_c) > \frac{2\rho_{\bar{x}u}\sigma_u}{\sigma_{\bar{x}}\left(1-\rho_{\bar{x}w}^2\right)}$, the school-level random-intercept variance is overestimated. Table 2 shows that, for the HSB data example, the GMC approach produces the largest estimate of the school-level random-intercept variance (6.74) just as it did for γ. Generally, if β_w and β_c are both positive, both terms in Equation 15 increase with increasing β_c.

The bias term for the estimated school (value-added) effects when the true and estimated models do not include w_j is

$$\delta_{u_j} = (\beta_w + \beta_c)(\bar{x}_{.j} - \bar{x}_{..}). \tag{16}$$

Again, this term differs from that given in Section 2 only by the substitution of β_w with $\beta_b = \beta_w + \beta_c$, meaning that the school effect estimates are biased by the between-school effects of the student-level covariate.

In all three of these expressions, there is bias even if there is no endogeneity and no contextual effect, which provides strong evidence against the use of the GMC approach even though it is unbiased for β_w. The GMC approach is consistent for γ, $\sigma^2_{u_j}$, and u_j only in the unlikely case where the within-school effect β_w and the contextual effect β_c have equal magnitudes with opposite signs; that is, when $\beta_w = -\beta_c$.

3.2.2. Including the group means of student covariates. As introduced in Section 2, the IGM approach is another common method in school comparison analyses. We showed that it produces biased estimates of γ and the school-level random-intercept variance in the presence of Level-2 endogenous student-level covariates. However, by including the group means of the student-level covariates, it adequately handles contextual effects. Thus, the bias and inconsistency expressions given in Section 2.6.2 do not change when there are contextual effects in addition to Level-2 endogeneity.

The bias δ_γ for the coefficient of the school-level covariate is given in Equation 9 and is due only to Level-2 endogeneity and not contextual effects. If both $\rho_{\bar{x}u}$ and $\rho_{\bar{x}w}$ are positive, this bias term is negative. For the HSB data example, in which $r_{\bar{x}w} = 0.36 > 0$ and we expect mean SES to be positively correlated with omitted school-level covariates, this bias term is expected to be negative and γ underestimated. Indeed, Table 2 (third column) shows that the IGM approach produces the smallest estimate of γ at 1.22.

The bias of the contextual effect estimator due to Level-2 endogeneity is $\delta_{\beta_c} = \delta_{\beta_b}$, given in Equation 8. Although such bias has been discussed by Hauser (1970) and Bingenheimer and Raudenbush (2004) among others, it is often ignored and does not appear to have been quantified. Because $\beta_c = \beta_b - \beta_w$ and the estimator has expectation $\beta_c + \delta_{\beta_c}$, we see that some of the difference in between- and within-school estimates of x_{ij} is due to Level-2 endogeneity $\left(\delta_{\beta_c}\right)$, as econometricians would generally attribute it, and some is due to contextual effects (β_c), as education researchers would generally attribute it. If $\rho_{\bar{x}u} > 0$, then $\delta_{\beta_c} > 0$, and it follows that β_c is overestimated. In other words, the difference in the between-school and within-school effects is the upper bound for the contextual effect. For instance, in the HSB example for which we believe that $\rho_{\bar{x}u} > 0$ and $\beta_c > 0$, the upper bound estimate for β_c can be computed from the values in the third column of Table 2 as: $\widehat{\beta}_c = \widehat{\beta}_b - \widehat{\beta}_w = 5.34 - 2.19 = 3.15$.

The inconsistency for the school-level random-intercept variance is given in Equation 10. Like the bias terms for the regression coefficients, it is only affected by Level-2 endogeneity $(\rho_{\bar{x}u})$ and not by contextual effects. As previously noted, this term is never positive. For the HSB data, Table 2 shows that this variance estimate of 2.37 is smaller than all but the RE estimates. If we set $\rho_{\bar{x}w} = 0$, we obtain the inconsistency for this variance when the model does not contain w_j ($\gamma = 0$). The resulting expression, $-\rho^2_{\bar{x}u}\sigma^2_u$, agrees with equation 19 of Raudenbush and Willms (1995) for the inconsistency of the Type B effects variance estimator.

When estimating individual school effects and $\gamma = 0$, the conditional bias is given in Equation 11. By recognizing that $\rho_{\bar{x}u}\sigma_u/\sigma_{\bar{x}}$ is the coefficient of u_j regressed on $\bar{x}_{\cdot j}$, this expression corresponds to equation 17 of Raudenbush and Willms (1995) for the estimates of Type B school effects. That is, the IGM approach (with $\gamma = 0$), which differentiates between school context effects β_c and school practice effects u_j, uses the same estimation model that Raudenbush and Willms (1995) discuss for estimating Type B school effects. They also acknowledge that the resulting Type B effect estimates will generally be biased. In the following section, we establish bounds for the Type B effects under certain conditions.

3.2.3. Hausman–Taylor/Two-Step Estimators and Bounds for Parameters. When we assumed in Section 2 that there were no contextual effects, the 2S or HT approaches were consistent for γ and the variance of the school-level random intercepts. However, when contextual effects exist, these approaches become inconsistent. One might be tempted to include $\bar{x}_{\cdot j}$ as a covariate, but if treated as exogenous in 2S or HT, the bias would be similar as for the IGM approach. If treated as Level-2 endogenous in HT, an additional exogenous covariate would be required. We therefore do not include $\bar{x}_{\cdot j}$ as a covariate. The Appendix gives derivations for the inconsistency terms, which cannot be interpreted as finite-sample bias, unlike the other estimators. As seen in the fourth and fifth columns of Table 2, the 2S and HT estimates of the regression coefficients for the HSB example are identical to two decimal places. The only noticeable difference is in the Level-2 random-intercept variance estimates, which is due to the different estimators used in Step 2 of 2S and Step A.2 of HT.

By ignoring contextual effects, the inconsistency in the coefficient of w_j is

$$\delta_\gamma = \frac{\beta_c \rho_{\bar{x}w} \sigma_{\bar{x}}}{\sigma_w}. \tag{17}$$

As would be expected, the absolute value of the inconsistency increases with increasing $|\beta_c|$ and $|\rho_{\bar{x}w}|$.

We can extend the idea of Type A school effects—the effects of the combination of school context and school practice—presented by Raudenbush and Willms (1995) to the Type A coefficient of w_j. In this case, δ_γ does not represent inconsistency but the contribution of school context to the overall effect. In the HSB example, w_j is an indicator for Catholic schools, and we may wish to estimate the difference in mean Type A school effects between Catholic and public schools. That is, a positive Type A Catholic school effect indicates the “bump”

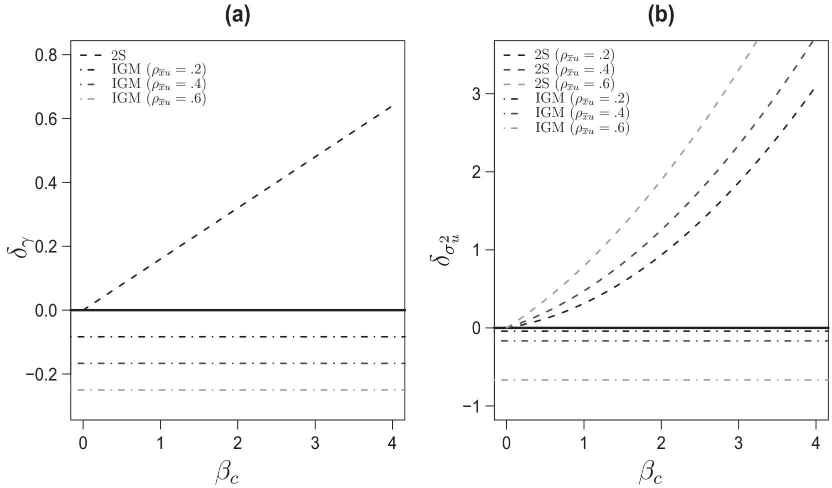

FIGURE 2. *Illustration of 2S and IGM estimators providing upper and lower bounds, respectively, for (a) the inconsistency* δ_γ *for* γ *and (b) the inconsistency* $\delta_{\sigma_u^2}$ *for* σ_u^2*. The values of* β_c *and* $\rho_{\bar{x}u}$ *are varied, whereas the following quantities are fixed:* $\rho_{\bar{x}w} = 0.2$, $\beta_w = 1$, $\sigma_{\bar{x}} = 0.4$, $\sigma_w = 0.5$, *and* $\sigma_u = 1$.
Note. 2S = two-step; IGM = including the group means.

students who attend Catholic schools receive due to the practices Catholic schools implement *and* the peer (or contextual) effects from students who attend Catholic schools. The inconsistency term above can equivalently be written as $\delta_\gamma = \beta_c[E(\bar{x}_{\cdot j}|w_j = 1) - E(\bar{x}_{\cdot j}|w_j = 0)]$, the contribution to the coefficient of w_j due to the contextual effect β_c and differences in mean school composition between the school types. In the HSB example, the GMC estimate of 2.80 is larger than the 2S estimate of 2.16 (given in Table 2), and the difference is $\widehat{\beta}_w$ times the difference in mean SES between Catholic and public schools (i.e., $2.80 - 2.16 = 0.64 = 2.19(0.150 - -0.146)$).

When an investigator is strictly interested in the effects that school types have on student achievement through the practices they implement, γ is viewed as a Type B coefficient, and δ_γ represents its inconsistency. The absolute value of this inconsistency $|\delta_\gamma|$ is less than that of the GMC approach in Equation 14 whenever β_c and β_w have the same sign.

We now compare the inconsistency for γ between the 2S approach (or HT) and the IGM approach. Whenever $\rho_{\bar{x}u}$ and β_c have the same sign, these inconsistencies have opposite signs that depend on the sign of $\rho_{\bar{x}w}$. For instance, when $\rho_{\bar{x}w} > 0$, we obtain

$$\frac{-\rho_{\bar{x}u}\rho_{\bar{x}w}\sigma_u}{\sigma_w(1-\rho_{\bar{x}w}^2)} < 0 < \frac{\beta_c\rho_{\bar{x}w}\sigma_{\bar{x}}}{\sigma_w}, \tag{18}$$

so that IGM gives a lower bound estimate and 2S gives an upper bound estimate for the Type B coefficient γ. As shown in Table 2, in the HSB example, where $r_{\bar{x}w} > 0$, the estimates of these bounds are 1.22 (IGM) and 2.16 (2S or HT).

We also show these bounds in Figure 2a as a function of β_c (from 0 to 4) for different values of $\rho_{\bar{x}u}$ (0.2, 0.4, and 0.6) with all other parameters in the bias terms for the 2S and IGM approaches (i.e., Equations 17 and 9) fixed and roughly based on the values of the corresponding statistics from the HSB data example ($\rho_{\bar{x}w} = 0.2$, $\beta_w = 1$, $\sigma_{\bar{x}} = 0.4$, $\sigma_w = 0.5$, and $\sigma_u = 1$). Note that this figure is similar to those often plotted in simulation studies, but, in this case, a simulation study is unnecessary as we derived analytical expressions for these bias terms. In Figure 2a, three horizontal lines of varying shades of gray represent the δ_γ bias term for the IGM approach at each $\rho_{\bar{x}u}$ value. They are horizontal because they are not affected by the value of the contextual effect, β_c, but they become increasingly negative as the correlation between $\bar{x}_{\cdot j}$ and u_j increases. Figure 2a thus clearly shows that the IGM approach handles the contextual effect but not Level-2 endogeneity.

In contrast, there is only a single line for the 2S approach because its δ_γ inconsistency term is not affected by Level-2 endogeneity; that is, it does not depend on $\rho_{\bar{x}u}$. However, this line has a positive slope because, as seen in Equation 17, the 2S approach does not adequately handle a nonzero contextual effect, that is, as β_c increases, δ_γ increases. At $\beta_c = 0$, the 2S line is at $\delta_\gamma = 0$, showing that when the student-level covariate is Level-2 endogenous and there is no contextual effect, the 2S approach produces a consistent estimate of γ, as discussed in Section 2.4. The 2S and IGM lines in Figure 2a clearly bound the (darkened) $\delta_\gamma = 0$ line, illustrating that these approaches produce estimates of the bounds for γ when $\rho_{\bar{x}u}$ and β_c have the same sign.

The inconsistency for the school-level random-intercept variance for the 2S or HT approaches differs from that for the GMC approach only in $\beta_w + \beta_c$ being replaced with β_c:

$$\delta_{\sigma_u^2} = \beta_c^2 \sigma_{\bar{x}}^2 (1 - \rho_{\bar{x}w}^2) + 2\beta_c \rho_{\bar{x}u} \sigma_u \sigma_{\bar{x}}. \qquad (19)$$

This expression is also the contribution of contextual effects to the variance of Type A effects, conditional on w_j (or within school types). If we set $\rho_{\bar{x}w} = 0$, we obtain the inconsistency for this variance when $\gamma = 0$.

Raudenbush and Willms (1995) in their equation 18 also write down an estimator for the Type A school-effects variance, but it is incorrect because it ignores the covariance between $\bar{x}_{\cdot j}$ and u_j. Specifically, their expression is missing the second term of the $\delta_{\sigma_u^2}$ expression. They also do not state explicitly that this variance would in practice be estimated by analyzing the residuals given in equation 15 of their article; that is, in Step 2 of the 2S approach.

Just as the 2S and IGM approaches provide bounds for γ when β_c and $\rho_{\bar{x}u}$ have the same sign, under these same conditions, these approaches provide estimated bounds for the school-level random-intercept variance:

$$-\frac{\rho^2_{\bar{x}u}\sigma^2_u}{1-\rho^2_{\bar{x}w}} < 0 < \beta^2_c\sigma^2_{\bar{x}}\left(1-\rho^2_{\bar{x}w}\right) + 2\beta_c\rho_{\bar{x}u}\sigma_u\sigma_{\bar{x}}. \tag{20}$$

For the HSB data, these bounds are estimated as 2.37 (IGM) and 3.58 (HT). Raudenbush and Willms (1995; Equation 9) also derive this inequality for the special case when the model does not include w_j. Their equation is obtained by setting $\rho_{\bar{x}w} = 0$ in the above inequality.

We illustrate these bounds for the school-level random-intercept variance in Figure 2b just as we did for the bounds on γ in Figure 2a. Again, we fix all parameter values given in the inconsistency terms (i.e., Equations 10 and 19) except for β_c and $\rho_{\bar{x}u}$. Just as in Figure 2a, the lines for the IGM approach are horizontal as they are not affected by β_c. They also become more negative as $\bar{x}_{\cdot j}$ becomes more correlated with u_j. Unlike in Figure 2a, where there was only a single line for the 2S approach as its δ_γ was not affected by $\rho_{\bar{x}u}$, in Figure 2b, there are three curves for the 2S approach—one for each $\rho_{\bar{x}u}$ value. The squared β_c term in Equation 19 makes these "lines" curved, and the degree of curvature decreases as $\rho_{\bar{x}u}$ increases. The figure clearly shows that 2S and IGM provide estimates of upper and lower bounds, respectively, for σ^2_u when β_c and $\rho_{\bar{x}u}$ share the same sign.

In the special case where there is no school-level covariate ($\gamma = 0$), the 2S approach is equivalent to the approach used by Raudenbush and Willms (1995) to estimate Type A school effects. Accordingly, if we are interested in comparing individual schools, the bias term for the school effects is

$$\delta_{u_j} = \beta_c(\bar{x}_{\cdot j} - \bar{x}_{\cdot\cdot}). \tag{21}$$

What is estimated is therefore the sum of school practice effects u_j and school context effects $\beta_c(\bar{x}_{\cdot j} - \bar{x}_{\cdot\cdot})$; see also equation 13 of Raudenbush and Willms (1995). Thus, if $\beta_c > 0$, effects estimated for advantaged schools with high mean SES will be inflated, ranking the schools higher than they should be and vice versa for disadvantaged schools with low mean SES. Note that δ_{u_j} is also the bias when school effects are estimated as the coefficients of school dummy variables in an estimation model that includes the student covariate x_{ij}, which is the FE approach used for school value added by Koedel and Betts (2010) and for teacher and school value added in the controversial Buddin (2011) study.

This bias is illustrated in Figure 1d, which is for the same set of hypothetical data as Figures 1a–1c but now considers a structural model that includes a contextual effect (i.e., $\beta_c \neq 0$). The black line with slope β_w again represents the FE regression line—the estimated expected school mean achievement as a function of school mean SES due to composition only if there were no contextual effects β_c or school practice (Type B) effects u_j. The thick gray line with slope $\beta_w + \beta_c$ represents the combined effect of composition and context. Hence, the black arrows from the gray line to school mean achievement $\bar{y}_{\cdot j}$ (gray dots) represent the hypothetical, unbiased estimates of school practice effects u_j if we could estimate β_c without bias. The FE school-effect estimates, represented by the arrows from the FE line, that start off as gray and then become black, incorporate the contextual effects (the gray portions of the arrows) and are hence overestimates of the Type B effects in absolute value. As mentioned previously, IGM school-effect estimates, represented by the dashed arrows in the figure, are too small, in absolute value.

4. Discussion

In this article, we have shown that the 2S or HT approaches are preferable over common estimation approaches, namely GMC and IGM, when the student-level covariates are Level-2 endogenous, but there are no contextual effects. The beauty of the 2S and HT estimators is that they relax Level-2 endogeneity assumptions without requiring any additional exogeneity assumptions or external instrumental variables. In the presence of both contextual effects and Level-2 endogenous student-level covariates, the coefficient of the school-level covariate w_j estimated by the 2S and HT approaches can be interpreted as the Type A coefficient. In addition, if β_c and $\rho_{\bar{x}w}$ have the same sign, the 2S or HT and the IGM approaches provide estimates of bounds for the Type B coefficient and the variance of the individual Type B school effects. In summary, we recommend HT for estimating school-type coefficients and HT or FE for estimating school effects when the estimates are not used for accountability purposes but are interpreted in terms of Type A effects. When used for accountability purposes, such estimates are unfair to schools with unfavorable student composition, in the presence of contextual effects. For this reason, we recommend reporting corresponding IGM estimates as a sensitivity analysis or robustness check. Although HT and IGM may not always provide strict bounds for the target quantities (because β_c and $\rho_{\bar{x}u}$ may not have the same sign for all student-level covariates, the bounds are estimated, and models could be misspecified), finding large differences in the estimates still signals a problem, whereas small differences provide some reassurance. We have also shown that GMC fails to produce unbiased school comparisons whether or not Level-2 endogeneity and/or contextual effects are present, providing strong evidence against its use.

For all of these findings, we assumed that the school-level covariate was Level-2 exogenous ($\rho_{wu} = 0$), but this may not be the case in practice. If w_j is Level-2 endogenous ($\rho_{wu} \neq 0$), we could obtain a consistent estimate of its coefficient using HT if there were Level-2 exogenous covariates available as well. However, in our example, we only have a Level-2 endogenous student-level covariate. Thus, if w_j is also Level-2 endogenous, we cannot interpret the coefficient estimated by the HT or 2S approaches as causal at the school level. Instead, it merely represents the difference in conditional expectations of Type A effects associated with a unit change in w_j. The effects of student-level covariates are causal because Level-2 endogeneity of the student-level variables has been handled, but the effects of school-level covariates are only "descriptive." This can be seen in the expression for δ_γ for HT when $\rho_{wu} \neq 0$ given in the Appendix, where the first term is the contribution of the contextual effect (making it a Type A coefficient) and the second term is due to the Level-2 endogeneity of w_j and takes the recognizable form of omitted variable bias. When $\rho_{wu} \neq 0$, the IGM estimator can be less biased than HT even if $\beta_c = 0$ (compare δ_γ in the Appendix). In the extreme case where the correlation ρ_{wu} is entirely due to w_j and u_j having $\bar{x}_{\cdot j}$ as a "common cause," the bias for IGM is zero. In less extreme situations, controlling for $\bar{x}_{\cdot j}$ may reduce the correlation ρ_{wu} sufficiently to make the bias of IGM smaller than that of HT.

We have assumed that all covariates are Level-1 exogenous, or in other words, that covariates are uncorrelated with the Level-1 error term. This assumption is violated if there are omitted student-level covariates or unobserved confounders that contribute to the Level-1 error term and are correlated with the covariates of interest. Relaxing the Level-1 exogeneity assumption would require external instrumental variables, but finding credible instrumental variables that are sufficiently correlated with the covariates of interest is a challenge. Moreover, even if such instruments were available, new methods would have to be developed to exploit them. We refer to Ebbes, Böckenholt, and Wedel (2004) for a useful review of methods for handling Level-1 as well as Level-2 endogeneity.

The methods we have discussed cannot disentangle contextual effects from Level-2 endogeneity. However, contextual effects can be estimated consistently if students are randomly assigned to schools or classes (so that there is no Level-2 endogeneity). Additionally, under appropriate assumptions and with richer data, as discussed by Boucher, Bramoullé, Djebbari, and Fortin (2012), consistent estimators can be obtained by exploiting demographic variation across grades but within schools (cross-sectional data) or across cohorts but within school grades (longitudinal data) as in Hanushek, Kain, Markman, and Rivkin (2003). See also Sacerdote (2011) for a recent survey of these and other approaches.

More elaborate models than the one considered in this article are sometimes used, particularly when longitudinal data are available, both teacher and school effects are of interests, and the persistence of the effects of previous schools and teachers must be taken into account (e.g., McCaffrey et al., 2004). While further issues arise in this case, the results derived here are important for deciding how to control for student composition in the presence of Level-2 endogeneity and contextual effects.

Appendix

Derivations of Bias and Inconsistency Terms

Here, we provide derivations of the bias and inconsistency terms $\delta_\gamma, \delta_{\sigma_u^2},$ and δ_{u_j} for the GMC, IGM and 2S/HT estimators for the structural model in Equation 13 with endogenous $x_{ij}, \mathrm{Cor}(\bar{x}_{\cdot j}, u_j) = \rho_{\bar{x}u},$ and endogenous $w, \mathrm{Cor}(w_j, u_j) = \rho_{wu}$. (See Online Appendix B for the inconsistency and bias terms for the regression coefficients of a more general structural model with multiple student- and school-level covariates at http://jeb.sagepub.com/supplemental). For simplicity, the following results are derived under the assumption of constant group size (i.e., $n_j = n$ for all j). We use the notation $\mu_{\bar{x}}$ and $\sigma_{\bar{x}}$ for the mean and standard deviation of $\bar{x}_{\cdot j}$ and σ_w for the standard deviation of w_j.

Group-Mean Centering. The true model can be rewritten as

$$y_{ij} = \alpha^* + \beta_w\left(x_{ij} - \bar{x}_{\cdot j}\right) + \gamma w_j + \zeta_j + \varepsilon_{ij},$$

where $\zeta_j = \beta_b \bar{x}_{\cdot j} + u_j$, $\beta_b = \beta_w + \beta_c$, and $\alpha^* = \alpha - \beta_b \mu_{\bar{x}}$.

However, the specified model assumes that the random intercept is uncorrelated with w_j. We therefore derive the linear projection of ζ_j on w_j,

$$\zeta_j = \delta_0 + \delta_\gamma w_j + v_j,$$

where v_j is uncorrelated with w_j. The estimator of the coefficient of w_j will converge to $\gamma + \delta_\gamma$, where δ_γ is the covariance between ζ_j and w_j divided by the variance of w_j, given by

$$\delta_\gamma = \frac{\mathrm{Cov}\left(\beta_b \bar{x}_{\cdot j} + u_j, w_j\right)}{\mathrm{Var}\left(w_j\right)} = \frac{(\beta_w + \beta_c)\rho_{\bar{x}w}\sigma_{\bar{x}} + \rho_{wu}\sigma_u}{\sigma_w}.$$

The estimator of the random intercept variance converges to $\sigma_u^2 + \delta_{\sigma_u^2}$, where

$$\begin{aligned}\delta_{\sigma_u^2} &= \mathrm{Var}\left(\beta_b \bar{x}_{\cdot j} + u_j - \delta_\gamma w_j\right) - \sigma_u^2 \\ &= \beta_b^2 \sigma_{\bar{x}}^2\left(1 - \rho_{\bar{x}w}^2\right) - \rho_{uw}^2 \sigma_u^2 + 2\beta_b \sigma_u \sigma_{\bar{x}}\left(\rho_{\bar{x}u} - \rho_{\bar{x}u}\rho_{wu}\right).\end{aligned}$$

The conditional bias for the individual school effects when there is no school-level covariate ($\gamma = 0$) is

$$\delta_{u_j} \equiv E\left[\widehat{u}_j | u_j, \ \mathbf{x}\right] - u_j = (\beta_w + \beta_c)\left(\bar{x}_{\cdot j} - \bar{x}_{\cdot\cdot}\right),$$

where $\bar{x}_{..}$ is the grand mean of x_{ij}.

Including the Group Means. We consider the linear projection of u_j in the original model (Equation 13) on $\bar{x}_{.j}$ and w_j

$$u_j \;=\; \delta_0 + \delta_{\beta_c}\bar{x}_{.j} + \delta_\gamma w_j + v_j,$$

where

$$\delta_{\beta_c} \;=\; \delta_{\beta_b} = \frac{(\rho_{\bar{x}u} - \rho_{wu}\rho_{\bar{x}w})\sigma_u}{\sigma_{\bar{x}}(1-\rho^2_{\bar{x}w})},$$

is the inconsistency for β_c and β_b and

$$\delta_\gamma = \frac{(\rho_{wu} - \rho_{\bar{x}u}\rho_{\bar{x}w})\sigma_u}{\sigma_w(1-\rho^2_{\bar{x}w})}$$

is the inconsistency for γ. Here, we have derived the partial regression coefficients from the covariance matrix of u_j, $\bar{x}_{.j}$, and w_j.

The inconsistency for the random-intercept variance is

$$\delta_{\sigma_u^2} = \;\mathrm{Var}\left(u_j - \delta_{\beta_c}\bar{x}_{.j} - \delta_\gamma w_j\right) - \sigma_u^2 \;= \frac{\sigma_u^2\left(\rho^2_{\bar{x}u} + \rho^2_{uw} - 2\rho_{\bar{x}u}\rho_{\bar{x}w}\rho_{wu}\right)}{1-\rho^2_{\bar{x}w}}.$$

The conditional bias for the individual school effects when there is no school-level covariate is given by:

$$\delta_{u_j} = -\frac{\rho_{\bar{x}u}\sigma_u}{\sigma_{\bar{x}}}\left(\bar{x}_{.j} - \bar{x}_{..}\right).$$

Two-Step and Hausman–Taylor Approaches. The true model can be written as

$$y_{ij} - \beta_w x_{ij} = \alpha^\bullet + \gamma w_j + \zeta_j + \varepsilon_{ij},$$

where $\zeta_j = \beta_c\bar{x}_{.j} + u_j$ and $\alpha^\bullet = \alpha - \beta_c\mu_{\bar{x}}$. Accordingly, we derive the linear projection of ζ_j on w_j to determine the inconsistency for γ,

$$\zeta_j = \;\delta_0 + \delta_\gamma w_j + v_j,$$

where v_j is uncorrelated with w_j and

$$\delta_\gamma = \frac{\mathrm{Cov}\left(\beta_c\bar{x} + u, w\right)}{\sigma_w^2} \;= \frac{\beta_c\rho_{\bar{x}w}\sigma_{\bar{x}} + \rho_{wu}\sigma_u}{\sigma_w}.$$

The inconsistency for the random-intercept variance is given by:

$$\begin{aligned}\delta_{\sigma_u^2} &= \mathrm{Var}\left(\beta_c\bar{x}_{.j} + u_j - \delta_\gamma w_j\right) - \sigma_u^2 \\ &= \beta_c^2\sigma_{\bar{x}}^2\left(1-\rho^2_{\bar{x}w}\right) - \rho^2_{uw}\sigma_u^2 + 2\beta_c\sigma_u\sigma_{\bar{x}}(\rho_{\bar{x}u} - \rho_{\bar{x}w}\rho_{wu}).\end{aligned}$$

The conditional bias for the individual school effects when there is no school-level covariate is given by:

$$\delta_{u_j} = \beta_c\left(\bar{x}_{\cdot j} - \bar{x}_{\cdot\cdot}\right).$$

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through grant R305B110017 to University of California, Berkeley. The opinions expressed are those of the authors and do not represent the views of the Institute or the U.S. Department of Education.

Supplementary Material

The online appendices are available at http://jeb.sagepub.com/supplemental.

References

Aud, S., Hussar, W., Johnson, F., Kena, G., Roth, E., Manning, E., . . . Zhang, J. (2012). *The condition of education 2012 (NCES 2012-045)*. Washington, DC: U.S. Department of Education, National Center for Education Statistics. Retrieved from http://nces.ed.gov/pubsearch

Ballou, D., Sanders, W., & Wright, P. (2004). Controlling for student background variables in value-added assessment of teachers. *Journal of Educational and Behavioral Statistics*, *29*, 37–65.

Baltagi, B. H. (2013). *The econometrics of panel data* (4th ed.). London, England: John Wiley.

Bates, D., Maechler, M., & Bolker, B. (2011). lme4: Linear mixed-effects models using S4 classes [Computer software manual]. Retrieved from http://CRAN.R-project.org/package=lme4

Bingenheimer, J. B., & Raudenbush, S. W. (2004). Statistical and substantive inferences in public health: Issues in the application of multilevel models. *Annual Review of Public Health*, *25*, 53–77.

Boucher, V., Bramoullé, Y., Djebbari, H., & Fortin, B. (2012). Do peers affect student achievement? Evidence from Canada using group size variation. *Journal of Applied Econometrics*, *29*, 91–109.

Breusch, T., Ward, M. B., Nguyen, H. T. M., & Kompos, T. (2011). On the fixed-effects vector decomposition. *Political Analysis*, *19*, 123–134.

Buddin, R. (2011). *Measuring teacher and school effectiveness at improving student achievement in Los Angeles elementary schools*. Retrieved from http://documents.latimes.com/buddin-white-paper-20110507/

Buddin, R., & Zamarro, G. (2009). Teacher qualifications and student achievement in urban elementary schools. *Journal of Urban Economics*, *66*, 103–115.

Burstein, L. (1980). The analysis of multilevel data in educational research and evaluation. *Review of Research in Education*, *8*, 158–233.

Chamberlain, G. (1980). Analysis of covariance with qualitative data. *Review of Economic Studies*, *47*, 225–238.

Croissant, Y., & Millo, G. (2008). Panel data econometrics in R: The plm package. *Journal of Statistical Software*, *27*, 1–43.

Cronbach, L. J. (1976). *Research on classrooms and schools: Formulation of questions, design, and analysis* (Occasional Paper). Stanford Evaluation Consortium, Stanford, CA.

Duncan, C., Jones, K., & Moon, G. (1998). Context, composition, and heterogeneity: Using multilevel models in health research. *Social Science & Medicine*, *46*, 97–117.

Ebbes, P., Böckenholt, U., & Wedel, M. (2004). Regressor and random-effects dependencies in multilevel models. *Statistica Neerlandica*, *58*, 161–178.

Fuller, W. A., & Battese, G. E. (1973). Transformations for estimation of linear models with nested-error structure. *Journal of the American Statistical Association*, *68*, 626–632.

Goldhaber, D. D., & Brewer, D. J. (1997). Why don't schools and teachers seem to matter? Assessing the impact of unobservables on educational productivity. *The Journal of Human Resources*, *32*, 505–523.

Gong, G., & Samaniego, F. J. (1981). Pseudo maximum likelihood estimation: Theory and applications. *The Annals of Statistics*, *9*, 861–869.

Gourieroux, C., & Monfort, A. (1995). *Statistics and econometric models* (Vol. 2). Cambridge, England: Cambridge University Press.

Greene, W. (2011). Fixed-effects vector decomposition: A magical solution to the problem of time-invariant variables in fixed effects models? *Political Analysis*, *19*, 135–146.

Hanchane, S., & Mostafa, T. (2012). Solving endogeneity problems in multilevel estimation: An example using education production functions. *Journal of Applied Statistics*, *39*, 1101–1114.

Hanushek, E. A., Kain, J. F., Markman, J. M., & Rivkin, S. G. (2003). Does peer ability affect student achievement? *Journal of Applied Econometrics*, *18*, 527–544.

Hauser, R. M. (1970). Context and consex: A cautionary tale. *American Journal of Sociology*, *75*, 645–664.

Hausman, J. A., & Taylor, W. E. (1981). Panel data and unobservable individual effects. *Econometrica*, *49*, 1377–1398.

Kennedy, E., & Mandeville, G. (2000). Some methodological issues in school effectiveness research. In C. Teddie & D. Reynolds (Eds.), *International handbook of school effectiveness research* (pp. 189–205). New York, NY: Routledge.

Kim, J. S., & Frees, E. W. (2006). Omitted variables in multilevel models. *Psychometrika*, *71*, 659–690.

Kim, J. S., & Frees, E. W. (2007). Multilevel modeling with correlated effects. *Psychometrika*, *72*, 505–533.

Kim, J. S., & Swoboda, C. M. (2011). Handling omitted variable bias in multilevel models: Model specification tests and robust estimation. In J. J. Hox & J. K. Roberts (Eds.), *Handbook of advanced multilevel analysis* (p. 197–217). New York, NY: Routledge.

Koedel, C., & Betts, J. (2010). Value added to what? How a ceiling in the testing instrument influences value-added estimation. *Education Finance and Policy*, *5*, 54–81.

Kreft, I. G. G., de Leeuw, J., & Aiken, L. S. (1995). The effect of different forms of centering in hierarchical linear models. *Multivariate Behavioral Research*, *30*, 1–21.

Lehmann, E. L. (1990). Model specification: The views of Fisher and Neyman, and later developments. *Statistical Science*, *5*, 160–168.

Lütke, O., Marsh, H. W., Robitzsch, A., Trautwein, U., Asparouhov, T., & Muthén, B. (2008). The multilevel latent covariate model: A new, more reliable approach to group-level effects in contextual studies. *Psychological Methods*, *13*, 203–229.

Maddala, G. S. (1971). The use of variance components models in pooling cross-section and time series data. *Econometrica*, *39*, 341–358.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *Review of Economic Studies*, *60*, 531–542.

Marsh, H. W., Seaton, M., Trautwein, U., Lütke, O., Hau, K. T., O'Mara, A. J., & Craven, R. G. (2008). The big-fish-little-pond-effect stands up to critical scrutiny: Implications for theory, methodology, and future research. *Educational Psychology Review*, *20*, 319–350.

McCaffrey, D. F., Lockwood, J., Koretz, D., Louis, T. A., & Hamilton, L. (2004). Models for value-added modeling of teacher effects. *Journal of Educational and Behavioral Statistics*, *29*, 67–101.

Mundlak, Y. (1978). On the pooling of time series and cross-sectional data. *Econometrica*, *46*, 69–86.

National Research Council & National Academy of Education. (2010). *Getting value out of value-added: Report of a workshop* (Committee on Value-Added Methodology for Instructional Improvement, Program Evaluation, and Educational Accountability, H. Braun, N. Chudowsky, & J. Koenig, Eds.). Washington, DC: National Academic Press.

Nye, B., Konstantopoulos, S., & Hedges, L. V. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis*, *26*, 237–257.

Pacagnella, O. (2006). Centering or not centering in multilevel models? *Evaluation Review*, *30*, 66–85.

Palardy, G. J., & Rumberger, R. (2008). Differential school effects among low, middle, and high social class composition schools: A multiple group, multilevel latent growth curve analysis. *School Effectiveness and School Improvement*, *19*, 21–49.

Plewis, I. (1989). Comment on "centering" predictors. *Multilevel Modelling Newsletter*, *1*, 6–11.

Plümper, T., & Troeger, V. E. (2007). Efficient estimation of time-invariant and rarely changing variables in finite sample panel analyses with unit fixed effects. *Political Analysis*, *15*, 124–139.

R Development Core Team. (2012). R: A Language and Environment for Statistical Computing [computer software]. Vienna, Austria. Retrieved from http://www.R-project.org/

Rabe-Hesketh, S., & Skrondal, A. (2012). *Multilevel and longitudinal modeling using Stata* (3rd ed., Vol. 1: Continuous responses). College Station, TX: Stata Press.

Raudenbush, S. W. (1989). Centering predictors in multilevel analysis: Choices and consequences. *Multilevel Modelling Newsletter* , *1*, 10–12.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models*: *Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Raudenbush, S. W., & Willms, J. D. (1995). The estimation of school effects. *Journal of Educational and Behavioral Statistics*, *20*, 307–335.

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, *73*, 417–458.

Rockoff, J. E., Jacob, B. A., Kane, J., & Staiger, D. O. (2011). Can you recognize an effective teacher when you recruit one? *Association for Education Finance and Policy*, *6*, 43–74.

Sacerdote, B. (2011). Peer effects in education: How might they work, how big are they and how much do we know thus far? In E. A. Hanushek, S. Machin, & L. Woessmann (Eds.), *Handbook of the economics of education* (Vol. 3, pp. 250–277). Amsterdam, The Netherlands: North-Holland.

Shin, Y., & Raudenbush, S. W. (2010). A latent cluster-mean approach to the contextual effects model with missing data. *Journal of Educational and Behavioral Statistics*, *35*, 26–53.

Skrondal, A., & Rabe-Hesketh, S. (2009). Prediction in multilevel generalized linear models. *Journal of the Royal Statistical Society, Series A*, *172*, 659–687.

Spanos, A. (2006). Where do statistical models come from? Revisiting the problem of specification. *IMS Lecture Notes—Monograph Series 2nd Lehmann Symposium—Optimality*, *49*, 98–119.

StataCorp. (2013a). Stata 13 Longitudinal-data/panel-data Reference Manual [Computer software manual]. College Station, TX: Stata Press.

StataCorp. (2013b). Stata statistical software: Release 13 [Computer software]. College Station, TX: Stata Press.

Steedle, J. T. (2012). Selecting value-added models for postsecondary institutional assessment. *Assessment & Evaluation in Higher Education*, *37*, 637–652.

U.S. Department of Education. (2010). *Overview information: Race to the top fund; notice inviting applications for new awards for fiscal year (FY) 2010* (Federal Register Vol. 75, No. 71).

Von Secker, C. E., & Lissitz, R. W. (1999). Estimating the impact of instructional practices on student achievement in science. *Journal of Research in Science Teaching*, *36*, 1110–1126.

Wayne, A. J., & Youngs, P. (2003). Teacher characteristics and student achievement gains: A review. *Review of Educational Research*, *73*, 89–122.

Wiley, D. E. (1975). Another hour, another day: Quantity for schooling, a potent path for policy. In R. M. Hauser, W. H. Sewell, & D. Alwin (Eds.), *Schooling and achievement in American society* (pp. 225–265). New York, NY: Academic Press.

Willms, J. D. (1986). Social class segregation and its relationship to pupil's examination results in Scotland. *American Sociological Review*, *51*, 224–241.

Wooldridge, J. M. (2010). *Econometric analysis of cross-section and panel data*. Cambridge, MA: The MIT Press.

Authors

KATHERINE E. CASTELLANO was an Institute of Education Sciences postdoctoral fellow at the University of California, Berkeley at the time of the writing and submitting of this manuscript. She is currently an Associate Psychometrician at ETS; e-mail: KEcastellano@ets.org. Her research interests involve the application of statistical models to educational policy issues, such as the use of the growth models in accountability systems.

SOPHIA RABE-HESKETH is a professor at the Graduate School of Education and Graduate Group in Biostatistics, University of California, Berkeley, CA 94720; e-mail: sophiarh@berkeley.edu. Her primary research interests include multilevel and latent variable modeling.

ANDERS SKRONDAL is a senior scientist at the Division of Epidemiology, Norwegian Institute of Public Health, Oslo, Norway; e-mail: anders.skrondal@fhi.no. His research interests are in statistics, psychometrics, econometrics, and biostatistics, including latent variable, multilevel, and longitudinal modeling.

Manuscript received July 2, 2013
First revision received November 5, 2013
Second revision received February 27, 2014
Third revision received July 3, 2014
Accepted July 10, 2014