The danger of conflating level-specific effects of control variables when primary interest lies in level-2 effects

Jason D. Rights1*, Kristopher J. Preacher2 and David A. Cole2
1University of British Columbia, Vancouver, British Columbia, Canada
2Vanderbilt University, Nashville, Tennessee, USA

In the multilevel modelling literature, methodologists widely acknowledge that a level-1 variable can have distinct within-cluster and between-cluster effects, and that failing to disaggregate these can yield a slope estimate that is an uninterpretable, conflated blend of the two. Methodologists have stated, however, that including conflated slopes of level-1 variables in a model is not problematic if substantive interest lies only in effects of level-2 predictors. Researchers commonly follow this advice and use methods that do not disaggregate effects of level-1 control variables (e.g., grand mean centering) when examining effects of level-2 predictors. The primary purpose of this paper is to show that this is a dangerous practice. When level-specific effects of level-1 variables differ, failing to disaggregate them can severely bias estimation of level-2 predictor slopes. We show mathematically why this is the case and highlight factors that can exacerbate such bias. We corroborate these findings with simulations and present an empirical example, showing how such distortions can severely alter substantive conclusions. We ultimately recommend that simply including the cluster mean of the level-1 variable as a control will alleviate the problem.

1. Introduction

Multilevel modelling (MLM) is a popular framework for analysing hierarchical data structures, such as students nested within schools, employees nested within companies, or repeated observations nested within persons. In such analyses, researchers are often interested in the potential effects of both level-1 (observation-level) as well as level-2 (cluster-level) variables.

In the MLM literature, methodologists have long recognized that a level-1 variable can have distinct within-cluster and between-cluster effects on an outcome of interest. Cronbach (1976) first highlighted this phenomenon in the social sciences, pointing out that the ‘overall’ relation between a level-1 variable and some outcome is, implicitly, an ‘uninterpretable blend’ of the within-cluster and between-cluster relations (Cronbach, 1976, p. 219). To ensure appropriate interpretation and inferences for slopes of level-1 predictors, methodologists widely recommend disaggregating these level-specific effects (Algina & Swaminathan, 2011; Bell, Jones, & Fairbrother, 2018; Cronbach, 1976; Curran &

*Correspondence should be addressed to Jason Rights, Department of Psychology, University of British Columbia, 2136 West Mall, Vancouver, BC, Canada V6T 1Z1 (email: jrights@psych.ubc.ca).

DOI:10.1111/bmsp.12194
As an illustration of disparate level-specific effects, Enders and Tofighi (2007) analysed data consisting of employees nested within organizations and examined the influence of workload (the number of hours worked per week) on psychological well-being. Of note is that, implicit in the level-1 variable of workload, there is a between-cluster component (the aggregate level of working hours for each organization) as well as a within-cluster component (an individual employee’s workload relative to their organization’s aggregate level). Enders and Tofighi (2007) showed that models that disaggregate these two level-specific components (e.g., by including both cluster-mean-centered workload and cluster-mean workload as separate predictors) yielded significant and negative slopes associated with both, suggesting that employees who work more relative to their organization’s aggregate level have worse well-being, and that companies with higher overall levels of working hours have employees with worse well-being. Importantly, the slope associated with the latter was larger than that associated with the former. Models that failed to disaggregate these two level-specific effects (e.g., those in which workload was grand-mean-centered) yielded a slope estimate that was somewhere in between the two; such a slope is said to be conflated (Preacher, Zyphur, & Zhang, 2010). Enders and Tofighi (2007) recommended that, for researchers interested in the influence of workload, both level-specific effects should be included in the model.

Conversely, methodologists have also stated that including conflated slopes of level-1 variables in a model is not problematic if substantive interest lies in effects of only level-2 predictors. For instance, Enders and Tofighi (2007) recommended that level-1 predictors be grand-mean-centered (a method that, by itself, does not disaggregate level-specific effects) when they are used merely as control variables for level-2 predictors; other methodologists echo these recommendations (e.g., Dalal & Zickar, 2012; Enders, 2013; Hofmann & Gavin, 1998; McCoach, 2010; Peugh, 2010; Raudenbush & Bryk, 2002). They further stated that this practice is appropriate even when the slope of the level-1 predictor is known to be an uninterpretable blend of disparate level-specific effects. With the workload/well-being example, they added a level-2 predictor (organization size or \( SIZE \)) to the model with a conflated slope of workload and stated that:

\[
\gamma_{01} \text{ regression coefficient quantifies the influence of } SIZE, \text{ controlling for individual workload. The } \gamma_{10} \text{ regression still gives a distorted view of the level-1 regression of well-being on workload, but this is not a concern if the substantive focus is on the level-2 covariate and the } \gamma_{01} \text{ coefficient. (Enders & Tofighi, 2007, p. 130)}
\]

The implicit rationale is that, even when the slope is conflated, the level-1 variable can still serve as an appropriate control variable for a level-2 predictor. Researchers commonly follow this advice and elect not to disaggregate the effects of level-1 control variables when examining effects of level-2 predictors (e.g., Dettmers, Trautwein, Lüdtke, Kunter, & Baumert, 2010; Diefendorff, Erickson, Grandey, & Dahling, 2011; Lenkeit, 2013; Lüdtke, Robitzsch, Trautwein, & Kunter, 2009; Merritt, Wanless, Rimm-Kaufman, Cameron, & Peugh, 2012; Salmivalli, Voeten, & Poskiparta, 2011; Trautwein, & Lüdtke, 2009; Wilkowski, Robinson, & Troop-Gordon, 2010).

The purpose of the current paper is to newly demonstrate that, when level-specific effects of level-1 variables differ, failing to disaggregate them can severely bias estimation of level-2 predictor slopes. This contrasts with previous studies on conflation that have
almost exclusively focused on issues in interpreting and making inferences about, specifically, a conflated slope of a level-1 predictor. For instance, methodologists have often discussed how such a slope is uninterpretable, or how such a slope might lead researchers erroneously to infer that there is an effect of the level-1 variable across individuals (when the apparent effect is actually driven by differences across groups), or, conversely, how such a slope might lead researchers erroneously to infer that there is an effect of the level-1 variable across groups (when the apparent effect is actually driven by differences across persons). However, the distortion inherent in specifying such a slope is not isolated to the level-1 variable itself. This distortion can, in turn, perturb the rest of the model as well. In using level-1 variables with conflated slopes as control variables for level-2 predictors, the level-2 predictor slopes are conditioned on the level-1 variables whose effects are misspecified and thus not controlled for correctly. Although methodologists have previously provided otherwise excellent guidance related to disaggregating level-specific effects in MLM (e.g., Enders & Tofighi, 2007), this particular point needs to be clarified.

The remainder of this paper proceeds as follows. We first consider different methods for assessing a level-2 predictor slope while conditioning on a level-1 variable. In particular, we show mathematically why using conflated level-1 variables as controls for level-2 predictors is inappropriate. We focus on the extent to which this practice yields biased estimation of level-2 predictor slopes. We highlight factors that could exacerbate such bias, and we show that this bias can be avoided altogether by fitting models that disaggregate level-specific effects. We then demonstrate these analytics with a simulation. Ultimately, we recommend that, to control for a level-1 variable when assessing the effects of level-2 predictors, one should include only the cluster mean of the level-1 variable as a control. We conclude with an empirical example wherein we assess the influence of a level-2 predictor while controlling for a level-1 variable, showing that the results differ markedly when the model is fit with a conflated level-1 predictor slope, as opposed to a model that appropriately disaggregates level-specific effects.

Before continuing, we note several caveats. First, to maintain a manageable scope, we restrict our focus to fairly simple models. In particular, we discuss models with only a single level-1 variable, we assume perfectly reliable predictors (consistent with standard MLM assumptions; Raudenbush & Bryk, 2002; Snijders & Bosker, 2012), and we present random intercept models as opposed to random slope models (because researchers typically model control variables with fixed slopes). We nonetheless discuss the generalizability of our results to more complex models in Section 5 (showing, for instance, that our recommendations hold for random slope models). Finally, we restrict our focus to the impact of conflated level-1 predictor slopes on the estimation of level-2 predictor slopes, although we stress that having conflated level-1 predictor slopes can similarly distort other portions of the model. For instance, recent work has shown the adverse effects of conflation on estimating interactions involving level-1 predictors (Loeys, Josephy, & Dewitte, 2018).

2. Methods to assess the effect of a level-2 variable while controlling for a level-1 variable

First consider the scenario in which a level-2 predictor is of primary substantive interest, but the researcher wishes to condition on a level-1 covariate. We compare four different
modelling options in terms of their abilities to assess appropriately the slope of the level-2 variable of interest.

2.1. Fully-disaggregated-\(x\) model

A level-1 predictor \(x_{ij}\), with observation \(i\) nested within cluster \(j\), equals the sum of the mean of cluster \(j\), denoted \(\bar{x}_j\), and observation \(i\)’s deviation from \(\bar{x}_j\), denoted \(x_{ij} - \bar{x}_j\) (i.e., \(x_{ij} = (x_{ij} - \bar{x}_j) + \bar{x}_j\)). The latter is commonly called a cluster-mean-centered, or group-mean-centered, level-1 variable (Enders & Tofighi, 2007; Snijders & Bosker, 2012). Assuming there is variability in both of these components, the level-1 variable can exert both within- and between-cluster effects on the outcome. Therefore, we can include all three of these slopes (that of the level-2 predictor of interest, \(z_j\), and those of the two level-specific components of \(x_{ij}\)) in a random-intercept model as

\[
y_{ij} = \gamma_{00} + \gamma_{xw}(x_{ij} - \bar{x}_j) + \gamma_{x}x_{ij} + \gamma_{z}z_j + u_{0j} + u_{1j}(x_{ij} - \bar{x}_j) + e_{ij},
\]

\(e_{ij} \sim N(0, \sigma^2),\)

\[
\begin{bmatrix}
  u_{0j} \\
  u_{1j}
\end{bmatrix} \sim MVN\left(\begin{bmatrix}
  0 \\
  0
\end{bmatrix}, \begin{bmatrix}
  \tau_{00} & \tau_{01} \\
  \tau_{01} & \tau_{11}
\end{bmatrix}\right).
\]

Here the continuous outcome \(y_{ij}\) is modelled as a function of the level-2 predictor of interest, \(z_j\), as well as the two components of \(x_{ij}\). The slope of interest is \(\gamma_{z}\), the within effect of \(x_{ij}\) is \(\gamma_{xw}\), and the between effect of \(x_{ij}\) is \(\gamma_{x}\). The fixed portion of the intercept is given as \(\gamma_{00}\), with the random portion (residual) given as \(u_{0j}\), which is normally distributed with across-cluster variance \(\tau_{00}\). The level-1 residual, \(e_{ij}\), is normally distributed with variance \(\sigma^2\).\(^1\)

Because this model contains separate parameters for the two level-specific effects of \(x_{ij}\), as a shorthand, we will call this the fully-disaggregated-\(x\) model. Of the four models we discuss, this is the most general specification. The remaining models place simple constraints on the parameters of this model. For illustrative purposes, borrowing from the structural equation modelling framework, this model is presented as a path diagram in Figure 1a. Note that the three single-headed arrows going from left to right denote the slopes of the three predictors. The double-headed arrow indicates that the two predictors \(x_{ij}\) and \(z_j\) can be correlated (although this is not an estimated MLM parameter and is thus not labelled). The lack of any double-headed arrow connecting \(x_{ij} - \bar{x}_j\) highlights the fact that \(x_{ij} - \bar{x}_j\) is a purely level-1 predictor with no across-cluster variance and therefore zero correlation with any level-2 predictor (Raudenbush & Bryk, 2002; Snijders & Bosker, 2012). With this in mind, we can make use of covariance algebra to express the slope of interest, \(\gamma_{z}\), in the population as a function of other model components. First consider the model-implied covariance of \(y_{ij}\) and \(z_j\):

\(^1\) Note that this model is equivalent to (albeit a slight reparameterization of) a contextual effect model (Enders & Tofighi, 2007; Kreft, de Leeuw, & Aiken, 1995; Snijders & Bosker, 2012) that includes as predictors \(x_{ij}\) (or grand-mean-centered \(\bar{x}_j\)), \(x_j\), and \(z_j\). Without loss of generality, we focus on the widely recommend cluster-mean-centered approach to disaggregation (Algina & Swaminathan, 2011; Enders & Tofighi, 2007; Hofmann & Gavin, 1998; Raudenbush & Bryk, 2002).
\[
\text{cov}(y_{ij}, z_j) = \text{cov}(\gamma_{00} + \gamma_{x_{0j}}(x_{ij} - x_{0j}) + \gamma_{x_{0j}}x_{0j} + \gamma_{z}z_j + u_{0j} + u_{1j}(x_{ij} - x_{0j}) + e_{ij}, z_j)
\]

\[
= \text{cov}(\gamma_{x_{0j}}x_{0j} + \gamma_{z}z_j, z_j)
\]

\[
= \text{cov}(\gamma_{x_{0j}}x_{0j}, z_j) + \text{cov}(\gamma_{z}z_j, z_j)
\]

\[
= \gamma_{x_{0j}}\text{cov}(x_{0j}, z_j) + \text{cov}(\gamma_{z}z_j, z_j)
\]

\[
= \gamma_{x_{0j}}\text{cov}(x_{0j}, z_j) + \gamma_{z}\text{var}(z_j).
\]

(2)

We can thus express \( \gamma_z \) in the population as

\[
\gamma_z = \frac{\text{cov}(y_{ij}, z_j) - \gamma_{x_{0j}}\text{cov}(x_{0j}, z_j)}{\text{var}(z_j)}.
\]

(3)

Importantly, because we have included both level-specific effects for \( x_{ij} \), when fitting this model we need not worry that estimation of \( \gamma_z \) is adversely affected by either failing to
control for the level-1 variable or controlling for it inappropriately by conflating level-specific effects. However, because $x_{ij} - x_{i*}$ is entirely uncorrelated with $z_j$, its inclusion does not affect the slope of $z_j$, as we show formally in the next section.  

2.2. x-mean model

Another way to control for the level-1 variable is simply to include $x_{i*}$ as a predictor without worrying about the within-cluster effect. We call this the x-mean model, and it is given by

$$y_{ij} = \gamma_{00} + \gamma_{x*}x_{i*} + \gamma_z z_j + u_{0j} + e_{ij},$$

$$e_{ij} \sim N(0, \sigma^2), \quad (4)$$

$$u_{0j} \sim N(0, \tau_{00}).$$

This model is also shown in Figure 1b. Note that the slope of $z_j$ is the same as that in the fully-disaggregated-x model. To prove this using covariance algebra, we show that the covariance between $y_{ij}$ and $z_j$ is the same for both of these models. For the x-mean model,

$$\text{cov}(y_{ij}, z_j) = \text{cov}(\gamma_{00} + \gamma_{x*}x_{i*} + \gamma_z z_j + u_{0j} + e_{ij}, z_j)$$

$$= \text{cov}(\gamma_{x*}x_{i*}, z_j) + \text{cov}(\gamma_z z_j, z_j)$$

$$= \gamma_{x*} \text{cov}(x_{i*}, z_j) + \gamma_z \text{var}(z_j). \quad (5)$$

Hence,

$$\gamma_z = \frac{\text{cov}(y_{ij}, z_j) - \gamma_{x*} \text{cov}(x_{i*}, z_j)}{\text{var}(z_j)}, \quad (6)$$

which is the same expression as shown in equation (3). Thus, researchers exclusively interested in the slope of $z_j$ need only include $x_{i*}$ as a predictor.

2.3. No-x model

We next consider the bias that results from failing to include $x_{i*}$ in the model, which we term the no-x model. Bias resulting from omitted variables is, of course, a well-established and recognized phenomenon, not only in MLM but also in single-level regression and other modelling frameworks (Rencher, 2000). We present this here as a basis of comparison for the subsequent conflated model. The no-x model is shown in Figure 1c and is given by

---

2 Some methodologists have correctly noted that including only a cluster-mean-centered/group-mean-centered level-1 variable does not appropriately condition the effects of level-2 variables on the level-1 variable, and, because of this, recommended instead that researchers grand-mean-center level-1 control variables (e.g., Enders & Tofghi, 2007; Hofmann & Gavin, 1998). However, the option of just group- versus grand-mean-centering the level-1 variable is not exhaustive, and we show here that neither of these is appropriate (and researchers should instead include the cluster mean of the level-1 variable as a control).
\[ y_{ij} = \gamma_{00} + \gamma_z z_j + u_{0j} + e_{ij}, \]

\[ e_{ij} \sim N(0, \sigma^2), \]

\[ u_{0j} \sim N(0, \tau_{00}). \]  

We denote the effect of \( z_j \) as \( \gamma_z \) to distinguish it from the true population effect, \( \gamma_z \). As before, we can express this slope as a function of other model parameters. In this no-\( x \) model, the model-implied covariance between \( y_{ij} \) and \( z_j \) is

\[ \text{cov}(y_{ij}, z_j) = \text{cov}(\gamma_{00} + \gamma_z z_j + u_{0j} + e_{ij}, z_j) \]

\[ = \text{cov}(\gamma_z z_j, z_j) \]

\[ = \gamma_z \text{var}(z_j). \]  

Hence, the slope of interest is

\[ \hat{\gamma}_z = \frac{\text{cov}(y_{ij}, z_j)}{\text{var}(z_j)}. \]  

(9)

The bias that results from failing to include \( x_{ij} \) can be computed by subtracting equation (9) from equation (3):

\[ \hat{\gamma}_z - \gamma_z = \frac{\text{cov}(y_{ij}, z_j)}{\text{var}(z_j)} - \frac{\text{cov}(y_{ij}, z_j) - \gamma_x \text{cov}(x_{ij}, z_j)}{\text{var}(z_j)} \]

\[ = \gamma_x \frac{\text{cov}(x_{ij}, z_j)}{\text{var}(z_j)}. \]  

(10)

Thus, assuming there is a non-zero effect of \( x_{ij} \) and a non-zero covariance between \( x_{ij} \) and \( z_j \), there will be bias in the estimation of the slope of \( z_j \) when failing to include \( x_{ij} \) in the model. (Note also that this same bias would result had we included only \( x_{ij} - x_{ij} \) as a control and not \( x_{ij} \)). We next consider how this degree of bias compares to the bias that results from including \( x_{ij} \) as control but failing to disaggregate level-specific effects.

2.4. Conflated-x model

Now suppose that a researcher, wishing to estimate a level-2 variable’s slope while controlling for a level-1 variable, followed current recommendations and simply grand-mean-centered \( x_{ij} \) or left \( x_{ij} \) in its raw form (these are equivalent models differing only in intercept interpretation; Kreft et al., 1995). We call this the conflated-x model. This model is shown in Figure 1d and can be expressed in grand-mean-centered form as

\[ y_{ij} = \gamma_{00} + \gamma_x (x_{ij} - \bar{x}) + \hat{\gamma}_z z_j + u_{0j} + e_{ij}, \]

\[ e_{ij} \sim N(0, \sigma^2), \]

\[ u_{0j} \sim N(0, \tau_{00}). \]  

Here \( \bar{x} \) denotes the mean of \( x_{ij} \) across all observations, with \( \gamma_x \) denoting the conflated slope of \( x_{ij} \), and \( \hat{\gamma}_z \) (distinct from the true \( \gamma_z \)) denotes the slope of the predictor
of substantive interest. This model is more restrictive than the fully-disaggregated-x model as it (implicitly) constrains the within-cluster slope equal to the between-cluster slope.³

Methodologists have long recognized that the conflated slope $\gamma_\text{c}$ is implicitly a weighted average of the level-specific effects $\gamma_\text{x}_\ell$ and $\gamma_\text{xw}$ (Burstein, 1980; Raudenbush & Bryk, 2002; Scott & Holt, 1982; Snijders & Bosker, 2012). The conflated slope can thus be expressed as

$$\gamma_\text{c} = \lambda \gamma_\text{x} + (1 - \lambda) \gamma_\text{xw},$$

with $\lambda$, ranging from 0 to 1, denoting the weight given to the between-effect. For single-level regression via ordinary least squares (OLS), $\lambda$ in equation (12) is simply the intraclass correlation coefficient of $x_{ij}$ (i.e., the proportion of total variance in $x_{ij}$ that is between clusters; Duncan, Cuzzort, & Duncan, 1961; Raudenbush & Bryk, 2002; Scott & Holt, 1982). Thus the OLS estimator for the conflated slope is equal to

$$\hat{\gamma}_\text{c} = ICC_x \hat{\gamma}_\text{x} + (1 - ICC_x) \hat{\gamma}_\text{xw}. \quad (13)$$

This implies that when a great deal of across-cluster outcome variance (or a large ICC) exists, the conflated estimate will be similar to the between-cluster estimate, whereas when little across-cluster outcome variance (or a small ICC) is present, the conflated estimate will be similar to the within-cluster estimate.

For multilevel contexts, the formula for $\lambda$ is a bit more complex. Raudenbush and Bryk (2002) showed that, for a simple random-intercept model with a single level-1 variable and non-varying cluster sizes (fit with maximum likelihood estimation), $\lambda$ is the precision of the between-cluster slope over the sum of the precisions of the between-cluster and within-cluster slopes, thus making the conflated slope a precision-weighted average of the two. More formally,

$$\hat{\gamma}_\text{c} = \left( \frac{\text{var}(\hat{\gamma}_\text{x})^{-1}}{\text{var}(\hat{\gamma}_\text{x})^{-1} + \text{var}(\hat{\gamma}_\text{xw})^{-1}} \right) \hat{\gamma}_\text{x} + \left( 1 - \frac{\text{var}(\hat{\gamma}_\text{x})^{-1}}{\text{var}(\hat{\gamma}_\text{x})^{-1} + \text{var}(\hat{\gamma}_\text{xw})^{-1}} \right) \hat{\gamma}_\text{xw} = \left( \frac{\sum_{j=1}^{J} (x_{ij} - \bar{x}_\bullet)^2 / (\bar{\tau}_00 + \hat{\sigma}^2 / N_j)}{\sum_{j=1}^{J} (x_{ij} - \bar{x}_\bullet)^2 / (\bar{\tau}_00 + \hat{\sigma}^2 / N_j) + \sum_{j=1}^{J} \sum_{i=1}^{N_j} (x_{ij} - \bar{x}_\bullet)^2 / \hat{\sigma}^2} \right) \hat{\gamma}_\text{x} + \left( 1 - \frac{\sum_{j=1}^{J} (x_{ij} - \bar{x}_\bullet)^2 / (\bar{\tau}_00 + \hat{\sigma}^2 / N_j)}{\sum_{j=1}^{J} (x_{ij} - \bar{x}_\bullet)^2 / (\bar{\tau}_00 + \hat{\sigma}^2 / N_j) + \sum_{j=1}^{J} \sum_{i=1}^{N_j} (x_{ij} - \bar{x}_\bullet)^2 / \hat{\sigma}^2} \right) \hat{\gamma}_\text{xw}, \quad (14)$$

with $J$ denoting the total number of clusters and $N_j$ denoting the cluster size (assumed equal across clusters).

Here, our primary focus is not specifically on the degree to which the two effects are weighted via $\lambda$ in estimating the conflated slope. The overarching point of equations (12–14) is to highlight that, when $\gamma_\text{x}$ and $\gamma_\text{xw}$ are different, $\gamma_\text{c}$ is a weighted average of the two and will be somewhere in between. Although other methodologists have noted this

³ Constraining the level-specific slopes equal in the fully-disaggregated-x model yields a single slope for $x_{ij}$, noting that $\gamma_\text{x}(x_{ij} - \bar{x}_\bullet) + \gamma_\text{xw} x_{ij} = \gamma_\text{x} x_{ij}$. Also, grand-mean-centering this $x_{ij}$ as in the conflated-x model changes only the intercept and does not change the slope (Raudenbush & Bryk, 2002; Snijders & Bosker, 2012).
Conflated level-1 control variables

(Algina & Swaminathan, 2011; Curran & Bauer, 2011; Curran et al., 2012; Enders & Tofghi, 2007; Hofmann & Gavin, 1998; Raudenbush & Bryk, 2002; Snijders & Bosker, 2012), none have specifically noted how this conflation impacts estimation of slopes of level-2 predictors.

We can solve for the population $\gamma_z$ in the conflated-$x$ model similarly to how we solved for $\gamma_z$ and $\tilde{\gamma}_z$ above by first considering the model-implied covariance between $y_{ij}$ and $z_j$:

$$\text{cov}(y_{ij}, z_j) = \text{cov}(\gamma_{00} + \gamma_{x}x_{ij} - x_{i1}z_j + u_{0j} + e_{ij}, z_j)$$

$$= \text{cov}(\gamma_{x}x_{ij} + \gamma_zz_j, z_j)$$

$$= \gamma_{x} \text{cov}(x_{ij}, z_j) + \gamma_z \text{var}(z_j)$$

$$= \gamma_{x} \text{cov}(x_{ij} - x_{i1}, z_j) + \gamma_z \text{var}(z_j)$$

$$= \gamma_{x} \text{cov}(x_{ij}, z_j) + \gamma_z \text{var}(z_j).$$

Hence,

$$\tilde{\gamma}_z = \frac{\text{cov}(y_{ij}, z_j) - \gamma_{x} \text{cov}(x_{ij}, z_j)}{\text{var}(z_j)},$$

Comparing this quantity to the true slope, $\gamma_z$, yields the bias induced by the conflation in the slope of $x_{ij}$:

$$\tilde{\gamma}_z - \gamma_z = \frac{\text{cov}(y_{ij}, z_j) - \gamma_{x} \text{cov}(x_{ij}, z_j)}{\text{var}(z_j)} - \frac{\text{cov}(y_{ij}, z_j) - \gamma_{x} \text{cov}(x_{ij}, z_j)}{\text{var}(z_j)}$$

$$= (\gamma_{x} - \gamma_{x}) \frac{\text{cov}(x_{ij}, z_j)}{\text{var}(z_j)}.$$

This expression in equation (17) makes clear the primary point of this paper: when the conflated slope of the level-1 variable is different than the between-cluster slope of the level-1 variable (i.e., when level-specific effects of $x_{ij}$ are different), the estimation of the slope of a level-2 variable will be biased, assuming there is a non-zero correlation between the level-1 and level-2 variables. As we will show, this bias can be substantial.

To emphasize further the adverse impact of using a conflated level-1 variable as a control, we compare the bias in estimation of the level-2 predictor slope in the conflated-$x$ model given in equation (17) to the bias from the no-$x$ model given in equation (10). The difference in the absolute values of these expressions is

$$|\tilde{\gamma}_z - \gamma_z| - |\gamma_z - \gamma_x| = \left| (\gamma_{x} - \gamma_{x}) \frac{\text{cov}(x_{ij}, z_j)}{\text{var}(z_j)} - \frac{\text{cov}(x_{ij}, z_j)}{\text{var}(z_j)} \right|$$

$$= \frac{\text{cov}(x_{ij}, z_j)}{\text{var}(z_j)} \left( |\gamma_{x} - \gamma_{x}| - \frac{|\gamma_{x}|}{|\gamma_{x} - \gamma_{x}|} \right).$$

Note that this expression is greater than or equal to 0 when
\[
\frac{\text{cov}(x_{ij}, z_j)}{\text{var}(z_j)} \left( |\gamma_{x0} - \gamma_{x} | - |\gamma_{x0} | \right) \geq 0
\]

(19)

Thus, whenever the absolute difference in the between effect and conflated effect is greater than the absolute value of the between effect, the bias will be worse for the conflated-x model than for the no-x model. In other words, in some cases, researchers would be better off not controlling for the level-1 variable at all than controlling for it incorrectly by letting it have a conflated slope. Examples of such situations are shown in the simulation (Section 3).

2.5. **Summary**

Three conclusions emerge for situations in which substantive interest is in the slope of a level-2 variable, \( z_j \). First, controlling for a level-1 variable by including its between-cluster slope (i.e., including \( x_{ij} \) as a predictor) appropriately conditions the slope of \( z_j \) on the level-1 variable; however, including the within-cluster slope (i.e., \( x_{ij} - x_{s,j} \)) as a predictor does not and is unnecessary. Second, failing to control for \( x_{s,j} \) by excluding it from the model yields bias in the slope of \( z_j \), provided \( x_{s,j} \) and \( z_j \) are correlated and the slope of \( x_{s,j} \) is non-zero. Third, controlling for \( x_{ij} \) without disaggregating level-specific effects (e.g., by grand-mean-centering \( x_{ij} \)) yields bias in the slope of \( z_j \), provided \( x_{s,j} \) and \( z_j \) are correlated and the within-cluster and between-cluster slopes of \( x_{ij} \) are not equal. Importantly, such bias can be greater than the bias that results from failing to include the level-1 variable in the model at all. We thus recommend that researchers who are interested solely in slopes of level-2 predictors condition on level-1 variables by including the cluster mean of \( x_{ij} \) as a control. Furthermore, we stress that including a conflated slope of \( x_{ij} \) as a control is not appropriate.

3. **Simulation**

We now present simulation results to explicitly demonstrate the aforementioned points. We show that the bias from the conflated-x model is particularly large whenever (a) the absolute value of the correlation between the level-2 predictor \( z_j \) and the between-cluster portion of \( x_{ij} \) is large and/or (b) the level-specific slopes of \( x_{ij} \) are highly dissimilar.

3.1. **Generating conditions**

Data were generated from the most general model presented here, that is, the fully-disaggregated-x model in equation (1). Note that \( x_{ij} \) was generated from a latent within-cluster component, \( x_{wj,ij} \), with only within-cluster variance, and a latent between-cluster component, \( x_{bj,j} \), with only between-cluster variance, such that \( x_{ij} = x_{wj,ij} + x_{bj,j} \). The outcome \( y_{ij} \) was then generated as a function of the observed \( x_{ij} \) cluster means, \( x_{s,j} \), and observation-specific deviations, \( x_{ij} - x_{s,j} \), consistent with standard MLM assumptions.

In designing the simulation, we focused on conditions that would affect the bias in estimating the slope of \( z_j \) with the conflated-x model, based on the derivation in the previous section. The two such conditions apparent from equation (17) are the magnitude of the difference in within and between effects of \( x_{ij} \) and the correlation between \( z_j \) and \( x_{s,j} \). Secondarily, we also manipulated cluster size and number of clusters, as these would
impact the precision weighting that occurs in estimating the conflated slope of \( x_{ij} \) (Raudenbush & Bryk, 2002). Finally, to highlight the point that the conflated-\( x \) model can be problematic both when there is a true, non-zero effect of \( z_j \) and even when there is no effect of \( z_j \) at all, we manipulated the slope of \( z_j \) to be either zero or non-zero.

Each of \( x_{w,j} \), \( x_{b,j} \), and \( z_j \) was generated with a variance of 1, with \( x_{w,j} \) normally distributed and \( x_{b,j} \) and \( z_j \) bivariate normally distributed with correlation \( \rho_{x_wz_j} \) (note that although \( x_{ij} \) has across-cluster variance, \( x_{w,j} \) by definition varies only within cluster and thus necessarily is uncorrelated with the level-2 variables). We varied \( \rho_{x_wz_j} \) across the parameter space \((-0.9, -0.6, -0.3, 0, .3, .6, \text{ and } .9)\). The fixed effects (with varying conditions in parentheses) were

\[
\begin{bmatrix}
\gamma_{00} \\
\gamma_{xw} \\
\gamma_{xb} \\
\gamma_z
\end{bmatrix}
= \begin{bmatrix}
1 \\
2 \\
(-2, 0, 2) \\
(0, 2)
\end{bmatrix}.
\] (20)

The varying values of \( \gamma_{xb} \) created conditions wherein \( \gamma_{xw} \) and \( \gamma_{xb} \) were either highly dissimilar \((2, -2)\), dissimilar \((2, 0)\), or identical \((2, 2)\). The level-1 and level-2 residual variances were \( \sigma^2 = 10 \) and \( \tau_{00} = 4 \), respectively. The average cluster sizes were either 10 (discrete uniformly distributed from 5 to 15) or 25 (discrete uniformly distributed from 15 to 35), with total number of clusters either 50 or 100.

In total, this yielded \( 7 \times 3 \times 2 \times 2 \times 2 = 168 \) conditions. For each of these, we generated 500 samples and fit the four models presented in Figure 1 using lmer in the lme4 package in R (Bates, Maechler, Bolker, & Walker, 2004) with restricted maximum likelihood estimation. Based on our regression results, the slope of \( z_j \) was of particular interest, we focused on the extent to which each model yielded biased estimation of this slope. Secondarily, we also considered the variance of the estimates of the slope of \( z_j \) across samples.

3.2. Results

For a parsimonious representation of the results, we focus on a subset of the conditions. We first note that results pertaining to bias were virtually identical across sample sizes; the effects of average cluster size and number of clusters were negligible. We thus present results from the largest sample size to highlight that the bias induced by the conflated-\( x \) model is not a small-sample issue. We also note that (as expected) the bias and variance results for the slope of \( z_j \) were identical across all conditions for the fully-disaggregated-\( x \) model and the \( x \)-mean model, thus we will simply present these as the \( x \)-mean model results.

We first present the results for the conditions in which the true slope of \( z_j \) is non-zero in Figure 2. Each panel corresponds to one of the three different magnitudes of difference in the within and between effects of \( X_{ij} \): in (a), the level-specific effects are highly dissimilar, in (b) they are moderately dissimilar, and in (c) they are identical. Each of three plots shows the bias (\( y \)-axis) as a function of \( \rho_{x_wz_j} \) (\( x \)-axis). The two horizontal dashed lines denote the point at which the relative bias is 10% (a commonly used criterion for bias to be deemed substantial).

As expected based on equation (17), the bias for the conflated-\( x \) model was greater at larger absolute values of \( \rho_{x_wz_j} \) (and zero when \( \rho_{x_wz_j} = 0 \)) and when the level-specific effects of \( X_{ij} \) were more dissimilar (and zero when they were identical). Noting again that the horizontal dashed lines denote relative bias of 10%, the bias shown is well beyond what
can be considered acceptable (and goes up to nearly 200%). As expected, the linear relation between bias and $\rho_{x_{ij}z_j}$ was moderated by the difference between $\gamma_{x_{ij}}$ and $\gamma_{z_j}$, as shown in equation (17). Comparing the bias between the conflated-$x$ model and the no-$x$ model, we see that, with the exception of Figure 2c, the bias is worse for the conflated-$x$ model than for the no-$x$ model for each condition (except when $\rho_{x_{ij}z_j} = 0$ and neither yield bias). For instance, in Figure 2b, $x_{ij}$ had no actual between effect; thus excluding $x_{ij}$ yielded no bias. The conflated-$x$ model, however, distorted the relation of $x_{ij}$ and $y_{ij}$. Therefore, the slope of $z_j$ was conditioned on an apparent (but illusory) between effect of $x_{ij}$, yielding extreme bias across the correlation space of $\rho_{x_{ij}z_j} \neq 0$. Lastly, as expected, the $x$-mean model yielded virtually no bias for any condition, despite technically being an underspecified model (i.e., excluding $x_{ij} - x_{ij}$). Thus the $x$-mean model is appropriate for all conditions whether or not level-specific effects of $x_{ij}$ differ; the conflated-$x$ model is appropriate only in the highly restrictive condition in which these are exactly the same.

The bias results for the condition in which the true slope of $z_j$ is zero are given in Figure 3a–c. These are essentially identical to the results when the true slope of $z_j$ is non-zero (with the slight exception that it no longer makes sense to describe ‘relative bias’ when the parameter is zero). We thus do not discuss these further, but present results to highlight that the conflated-$x$ model can mislead researchers into thinking there is either a positive effect or a negative effect of $z_j$ when there is truly no effect at all.

Although not a focus of the simulation, we also note the across-sample variance of the estimates of the slope of $z_j$ for the conflated-$x$ model compared to the $x$-mean model. We will consider this only for the conditions in which the conflated-$x$ model yielded no bias, that is, when level-specific effects of $x_{ij}$ were identical (the bias of the conflated-$x$ model was otherwise too extreme to make a comparison of variances worthwhile). As noted by other authors, when the assumption underlying the conflated-$x$ model holds (i.e., $\gamma_{x_{ij}} = \gamma_{z_j}$), this model yields the most efficient estimator of the slope of $x_{ij}$ (Raudenbush & Bryk, 2002). Our results indicate that the conflated-$x$ model also yielded the most efficient estimation of the slope of $z_j$ for the conditions in which the level-specific effects of $x_{ij}$ were exactly equal. Averaging across all such conditions, the average standard deviation of the estimated slope of $z_j$ was 0.21 for the conflated-$x$ model and 0.31 for the $x$-mean (and fully-disaggregated-$x$) model; however, the difference between the two is less pronounced (0.21 vs. 0.28) when focusing on conditions in which $\rho_{x_{ij}z_j}$ was less extreme (not ±.9) and

---

### Figure 2. Bias in estimating the slope (equal to 2 in the population) of a level-2 variable. Notes. In Panel B the no-$x$ model and $x$-mean model overlap entirely. In Panel C the conflated-$x$ model and $x$-mean model overlap entirely. Outside of the two horizontal dashed lines, relative bias is >10%.
likely more realistic. We caution readers, however, that this improvement is fairly small when compared to the bias resulting from the conflated-$x$ model with unequal level-specific effects, in which the relative bias was almost 200% for some conditions. Thus, we maintain our recommendation to use the $x$-mean model, noting that it may be unrealistic to assume that level-specific effects are exactly equal in practice (Preacher et al., 2010).

4. Empirical example

To illustrate the potential distortion of results that can occur with the conflated-$x$ model, we present an empirical example in which substantive interest is in the effect of a level-2 variable after controlling for a level-1 variable. We present an abbreviated version here. More extensive details can be seen in the Appendix S1 to this article and in Cole et al. (2019).

Data were obtained via an adaptation of Cole et al.’s (2014) Behind Your Back procedure. In this procedure, 272 adolescents listened to recordings of 21 conversations of a boy and a girl talking about an absent third person, with the gender of the third person matching that of the participant. The content of the conversations ranged from mild to mean. Participants were instructed to imagine they were overhearing conversations about themselves and to answer conversation-specific questions about their negative appraisal, their cognitive reaction, and their emotional reaction of sadness to each event. The 21 conversation-specific observations were nested within adolescent. Of substantive interest is the effect of an adolescent’s propensity for negative appraisal (NA) of events (level-2 predictor) on negative cognitive reaction (COG) to events, above and beyond that attributable to one’s event-specific emotional reaction of sadness (SAD; level-1 control variable).

We fit each of the four models in Figure 1 to these data, with the associated slope estimates provided in Figure 4. In both the fully-disaggregated-$x$ and $x$-mean models, the level-specific effects of SAD are positive and significant; however, no effect of NA was detectable. The no-$x$ model failed to control for SAD altogether, and the effect of NA was significant. In the conflated-$x$ model (the most common approach in practice), the within- and between-cluster effects of SAD are conflated into a single slope and the effect of NA is significant. The effect of NA was appropriately conditioned on person-mean SAD in the

---

**Figure 3.** Bias in estimating the slope (equal to 0 in the population) of a level-2 variable. Notes: In Panel B the no-$x$ model and $x$-mean model overlap entirely. In Panel C the conflated-$x$ model and $x$-mean model overlap entirely. The horizontal dashed line denotes zero bias.
former two models (with identical estimates, as expected), and was distorted in the latter two.  
This example shows that using conflated slopes of level-1 variables as a control for level-2 variables can severely distort results and profoundly alter the substantive implications. Controlling inappropriately with the conflated-$x$ model would suggest that one’s propensity for $NA$ has a direct impact on $COG$ that is comparable to the direct impact of $SAD$. Controlling appropriately with either the fully-disaggregated-$x$ model or the $x$-mean model suggests that the association of $NA$ and $COG$ can be better understood through the relation with $SAD$.

5. Discussion
We have argued that, despite methodological recommendations suggesting otherwise, it is inappropriate to use level-1 predictors as control variables for level-2 predictors if level-specific effects of the level-1 predictors are conflated. We showed mathematically why doing so will yield biased estimation of level-2 predictor slopes and provided corroborative simulation results. We also demonstrated with real data the distortion of results for level-2 predictor slopes that occurs with conflated level-1 control variables. In sum, we

![Diagram of empirical example results](image)

**Figure 4.** Empirical example results: Predicting cognitive response using each of the four modelling approaches. $^*p < .005$. 
recommend that, if researchers are solely interested in the effects of level-2 predictors, they can use the cluster means of level-1 predictors as control variables.\footnote{This implies that one could use aggregated data (i.e., analyse only cluster means of variables) in a single-level analysis to avoid conflation when interest lies primarily in level-2 effects. However, MLM has advantages in that one can also include level-1 predictors of secondary interest, and can quantify the amount of within-cluster and between-cluster outcome variation, as well as the proportion of within-cluster and between-cluster variance accounted for by predictors (Rights & Sterba, 2019). Additionally, within a multilevel framework, one can model cluster means as latent variables, which can avoid bias when cluster means are measured with error (Lüdtke \textit{et al.}, 2008; Preacher \textit{et al.}, 2010).}

A caveat is that, when level-specific effects of level-1 variables are identical in the population, it is technically no longer necessary to disaggregate the two, and when disaggregating, a superfluous parameter is freed up which yields more variance in estimation (as discussed in the simulation in Section 3). We feel, however, that the assumption of equal level-specific effects is highly restrictive. Indeed, others have argued that this assumption rarely holds in practice (Preacher \textit{et al.}, 2010). Thus, we maintain our recommendation not to conflate effects of level-1 control variables. At the very least, we strongly encourage researchers not to blindly assume that the two effects are equal.

Regarding the generalizability of our simulation results, we first note that, in practice, researchers may have multiple level-1 control variables. In these cases, the expected bias in estimating level-2 slopes induced by conflation would be more complex than that shown in equation (17), as multiple conflated slopes would distort the model, and the amount of bias could potentially be exacerbated. One could avoid this by including the cluster mean of each level-1 control variable as a separate predictor. Second, despite our aforementioned focus on fixed slope models, none of the analytic results discussed in this paper would change if we had considered random slope models; that is, the addition of a random component of the slope of the level-1 predictors would not affect the last line of equations (6), (10) and (17). We thus expected that the inclusion of random slopes would not meaningfully affect the simulation results. To confirm, we reran the entire simulation with random slopes of level-1 predictors (with the slope variance set to be a quarter of the random intercept variance) and found virtually identical results (i.e., the plots in Figures 2 and 3 were unchanged). Finally, another consideration is our choice of parameter values, in particular the degree of discrepancy between the within-cluster and between-cluster effects of the level-1 variable. Making general statements about the expected discrepancy in practice is difficult: in certain contexts, highly disparate level-specific effects might be expected (e.g., Baldwin, Wampold, & Imel, 2007; Marini \textit{et al.}, 2013; Wang & Maxwell, 2015), whereas in other contexts differences between the two might be negligible or non-existent. Future work can systematically investigate the research contexts in which disparate level-specific effects are most likely, in order to determine the contexts in which conflation is likely to have biased the estimation of level-2 effects.

Future work can also extend the points made here to latent variable models. Particularly with multilevel structural equation modelling, researchers can decompose level-1 variables into latent within- and between-cluster components, rather than using observed cluster means (Asparouhov & Muthén, 2019; Lüdtke \textit{et al.}, 2008; Preacher \textit{et al.}, 2010). The basic idea would still apply: when modelling $x_{ij}$ as a latent variable on which to condition the effect of a level-2 predictor, the latent $x_{ij}$ level-specific effects should still be disaggregated; that is, an explicit latent between-cluster component of $x_{ij}$ ($x_{b,ij}$) should be included in the model.

Our hope is that this work will discourage researchers from specifying models with conflated level-1 variables as controls for level-2 variables. Researchers historically have
not worried about decomposing level-1 control variables due to a lack of interest in their potential disparate level-specific effects. However, failing to appropriately disaggregate these effects can yield serious distortion of results pertaining to other predictors in a model, regardless of one’s interest in the effect(s) of the level-1 variable itself. We encourage researchers to be mindful of the fact that controlling for a variable merely by including it as a predictor is not sufficient; it is necessary to model its effects appropriately.

Acknowledgements
The authors would like to thank Sonya Sterba for helpful comments.

References
Dettmers, S., Trautwein, U., Lüdtke, O., Kunter, M., & Baumert, J. (2010). Homework works if homework quality is high: Using multilevel modeling to predict the development of
achievement in mathematics. *Journal of Educational Psychology*, *102*, 467–482. https://doi.org/10.1037/a0018453


Received 3 August 2018; revised version received 21 October 2019

---

**Supporting Information**

The following supporting information may be found in the online edition of the article:

**Appendix S1.** Empirical example predicting cognitive response using each of the four modeling approaches.