Routledge
Taylor & Francis Group

Multivariate Behavioral Research

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/hmbr20

On the Common but Problematic Specification of Conflated Random Slopes in Multilevel Models

Jason D. Rights & Sonya K. Sterba

To cite this article: Jason D. Rights & Sonya K. Sterba (2023): On the Common but Problematic Specification of Conflated Random Slopes in Multilevel Models, Multivariate Behavioral Research, DOI: 10.1080/00273171.2023.2174490

To link to this article: https://doi.org/10.1080/00273171.2023.2174490

MULTIVARIATE BEHAVIORAL RESEARCH
https://doi.org/10.1080/00273171.2023.2174490

On the Common but Problematic Specification of Conflated Random Slopes in Multilevel Models

Jason D. Rights[a] and Sonya K. Sterba[b]

[a]University of British Columbia, Vancouver, BC, Canada; [b]Vanderbilt University, Nashville, TN, Canada

ABSTRACT

For multilevel models (MLMs) with fixed slopes, it has been widely recognized that a level-1 variable can have distinct between-cluster and within-cluster fixed effects, and that failing to disaggregate these effects yields a conflated, uninterpretable fixed effect. For MLMs with random slopes, however, we clarify that two different types of slope conflation can occur: that of the fixed component (termed fixed conflation) and that of the random component (termed random conflation). The latter is rarely recognized and not well understood. Here we explain that a model commonly used to disaggregate the fixed component—the contextual effect model with random slopes—troublingly still yields a conflated random component. Negative consequences of such random conflation have not been demonstrated. Here we show that they include erroneous interpretation and inferences about the substantively important extent of between-cluster differences in slopes, including either underestimating or overestimating such slope heterogeneity. Furthermore, we show that this random conflation can yield inappropriate standard errors for fixed effects. To aid researchers in practice, we delineate which types of random slope specifications yield an unconflated random component. We demonstrate the advantages of these unconflated models in terms of estimating and testing random slope variance (i.e., improved power, Type I error, and bias) and in terms of standard error estimation for fixed effects (i.e., more accurate standard errors), and make recommendations for which specifications to use for particular research purposes.

KEYWORDS
Multilevel modeling; random slopes; contextual effect models; group mean centering

Multilevel modeling (MLM; also known as hierarchical linear modeling or linear mixed effects modeling) is a popular and useful tool for analyzing nested data structures, such as patients nested within clinician or repeated observations nested within persons (e.g., Raudenbush & Bryk, 2002; Snijders & Bosker, 2012). Using MLM, researchers can simultaneously examine the influence of observation-level/level-1 predictors (e.g., patient characteristics) and cluster-level/level-2 predictors (e.g., clinician characteristics) on an outcome of interest.

In the methodological literature, it has been long recognized that a level-1 variable can have both a *between-cluster* and a *within-cluster* fixed effect in MLM. To explain, consider that the observed value of a level-1 variable for observation i nested within cluster j is implicitly the sum of two distinct parts: (a) the aggregate score for cluster j, and (b) observation i's deviation from the cluster-aggregated score. These two parts can each exert an influence on a particular outcome; the fixed effect of the former can be termed the between-cluster fixed effect, and that of the latter the within-cluster fixed effect. Importantly, these effects need not be the same nor similar. Because between-cluster and within-cluster fixed effects of a level-1 variable may differ, it is widely recommended to explicitly disaggregate them (e.g., Algina & Swaminathan, 2011; Cronbach, 1976; Curran et al., 2012; Curran & Bauer, 2011; Enders, 2013; Enders & Tofighi, 2007; Hedeker & Gibbons, 2006; Hofmann & Gavin, 1998; Raudenbush & Bryk, 2002; Snijders & Bosker, 2012). An estimate that fails to disaggregate level-specific effects can be said to be *conflated* (e.g., Preacher, 2011).

As a concrete example wherein the between-cluster and within-cluster fixed effects of a level-1 variable differ, Baldwin et al. (2007) examined patients nested within clinicians and predicted patient outcomes from therapeutic alliance, defined as the degree to which the patient-clinician dyad engages in "collaborative,

CONTACT Jason Rights jrights@psych.ubc.ca Quantitative Methods Area, Department of Psychology, University of British Columbia, 2136 West Mall, Vancouver, BC, V6T 0C1, Canada

purposive work" (Hatcher & Barends, 2006). Fitting a random-intercept/fixed-slope MLM, Baldwin et al. (2007) found that clinicians (i.e., clusters) with a higher mean therapeutic alliance across their patients had lower mean depression among their patients; in other words, there was evidence of a *between-cluster* or between-clinician fixed effect. However, they also found that, for patients who had the same clinician, therapeutic alliance was not predictive of depression; in other words, there was no evidence for a *within-cluster* or within-clinician fixed effect. Thus, results suggest the influence of therapeutic alliance to be realized at the clinician-level, and the influence of patient's individual propensity to engage with clinicians may not be as important. Indeed, Baldwin et al. (2007) note that their findings imply "it would behoove therapists to attend to their own contributions to the alliance and focus less on characteristics of the patient that impede the development of the alliance (p. 851)." Prior to the Baldwin et al. (2007) study, earlier researchers had estimated only a conflated effect of therapeutic-alliance on patient outcomes (rather than disaggregating effects) and had interpreted this conflated effect as indicating that patient-level differences, rather than clinician-level differences, were responsible for the relationship (e.g., Mallinckrodt, 2000). The disaggregated-effect results from Baldwin et al. (2007), however, suggest that earlier conclusions were subject to an ecological fallacy (Robinson, 1950) wherein inferences were inappropriately made regarding individuals (patients) based on group (clinician) data. Disaggregation helps researchers to avoid such fallacies and make appropriate, level-specific inferences.

To disaggregate level-specific effects, researchers from fields such as industrial-organizational psychology, sociology, developmental psychology, and social epidemiology have for many years used *contextual effect models* (Burstein, 1980; see Enders, 2013 or Brincks et al., 2017 for historical reviews). Contextual effect models are defined shortly in this paper, but are also detailed in standard MLM textbooks (e.g., Raudenbush & Bryk, 2002; Snijders & Bosker, 2012). In the methodological literature, the disaggregation afforded by the contextual effect model has been discussed mainly for *fixed slopes* (e.g., Boateng, 2016; Enders, 2013; Enders & Tofighi, 2007; Henry & Slater, 2007; Raudenbush & Bryk, 2002). However, the very same points and explanations about the utility and interpretation of contextual effect models for disaggregation have often been directly applied to *random slope* models—without mentioning how the presence of random slopes affects the disaggregation (e.g., Antonakis et al., 2019; Algina & Swaminathan, 2011; Brincks et al., 2017; Hamaker & Muthén, 2019; Hoffman & Stawski, 2009; Hox, 2010; Kreft et al., 1995; Paccagnella, 2006; Snijders & Bosker, 2012). Consequently, *random-slope contextual effect models* have become widely used in practice by researchers interested in disaggregating level-specific effects (e.g., Bliese & Britt, 2001; Deemer et al., 2017; Diez-Roux et al., 2000; Espelage et al., 2003; Fischer et al., 2004; Hoffman & Stawski, 2009; Kidwell et al., 1997; Lee, 2009; Lee & Bryk, 1989; Merlo et al., 2005; Poteat et al., 2007; Schempf & Kaufman, 2012; Titus, 2004).

There are several problems underlying this current situation:

1. It is not well understood that, for a level-1 variable, there are two distinct types of slope conflation that can occur in MLM (as will be detailed later). Namely, there is *fixed conflation*—wherein the *fixed* components of the slope's level-specific portions are implicitly set to equality—and there is *random conflation*—wherein the *random* components of the slope's level-specific portions are implicitly set to equality. The slope of a level-1 variable can be characterized as having only one type of conflation (here termed *partial conflation*) or having both types (here termed *full conflation*), or having neither (here termed *unconflation*). Although fixed conflation has been widely recognized, random conflation has been almost entirely ignored, and the exhaustive possibilities (full conflation vs. partial conflation vs. unconflation) have not previously been fully enumerated.
2. There is little appreciation that contextual effect models with random slopes have a *conflated random* component of the slope, despite having an *unconflated fixed* component of the slope.
3. There are negative consequences of fitting contextual effect models with conflated random slopes that researchers need to understand. Such consequences have not been explained or demonstrated and can include erroneous interpretation and inferences.
4. Researchers need a full delineation of which random slope MLM specifications yield an unconflated random component, and need recommendations on which to use for particular purposes.

This paper addresses each of these problems, with the primary purpose of encouraging researchers to be mindful of avoiding random conflation, similarly to how researchers are already (largely) mindful of avoiding fixed conflation. First we demonstrate the *full conflation* of both the fixed and random component in a random slope MLM with an uncentered (or grand-mean-centered) level-1 predictor. We then show that the conventional random-slope contextual effect model is *random conflated* in that it disaggregates the fixed component, but, importantly, fails to disaggregate the random component of this level-1 predictor's slope. Due to this random conflation, the variance component for the slope of the level-1 predictor in this model actually reflects an uninterpretable blend of *both* slope heterogeneity of the level-1 predictor and intercept heteroscedasticity at level-2. Therefore if, in the population, there is truly slope heterogeneity but there is no intercept heteroscedasticity, we show later that the (random-conflated) variance component estimate for the slope of a level-1 variable in a conventional random-slope contextual effect model will, disturbingly, be weighted toward zero. This, in turn, leads to heightened Type II errors when testing slope heterogeneity for that level-l variable, and also compromises inference for the fixed effect of that level-1 variable. We fully detail this and other negative consequences of such random conflation, including different kinds of erroneous interpretation and inferences pertaining to the estimated slope variance along with inaccurate standard errors for fixed effects. Next, noting that the concept of fixed and/or random conflation is important to consider for any MLM with level-1 variables, we provide a general taxonomy for any such model wherein we distinguish random slope MLM specifications that are fully conflated vs. partially conflated vs. unconflated. We then provide evidence via simulation that these unconflated models provide better Type I error, power, and bias for the random slope variance and provide more accurate standard errors for fixed effects than the widely used conventional random-slope contextual effect model. Finally, we illustrate the concepts presented with an empirical example, provide recommendations for practice, and discuss avenues for future research.

Conflation of the fixed and random components of the slope in a random slope model

Uncentered MLM with random slopes: Conflation of the fixed and random components

Consider a random slope MLM with an uncentered (or grand-mean-centered)[1] level-1 predictor; this remains one of—if not the—most commonly used random slope specification in applied practice (as mentioned in literature reviews by, e.g., Curran et al., 2012; Hamaker & Grasman, 2014; Hoffman, 2015).

$$\begin{gathered} y_{ij} = \gamma_{00} + \gamma_{10}x_{ij} + u_{1j}x_{ij} + u_{0j} + e_{ij} \\ e_{ij} \sim N(0, \sigma^2) \\ \begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim MVN\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_{00} & \\ \tau_{10} & \tau_{11} \end{bmatrix} \right) \end{gathered} \tag{1}$$

Here, we are modeling a continuous outcome y_{ij} for observation i nested in cluster j. The level-1 residual, e_{ij}, is normally distributed with variance σ^2. The fixed component of the intercept is γ_{00} and the random component (residual) of the intercept is u_{0j}; likewise, the fixed component of the slope of the level-1 variable x_{ij} is γ_{10} and the random component is u_{1j}.[2] The level-2 random components u_{0j} and u_{1j} are multivariate normally distributed with covariance matrix **T**. Note that this model is also presented in Table 1, which we refer to later when comparing model specifications.

To make the conflation in Eq. (1) apparent, in Eq. (2) we equivalently express this model by substituting $x_{ij} = x_{ij} - x_{\cdot j} + x_{\cdot j}$. That is, in Eq. (2) we replace x_{ij} with the sum of its within-cluster portion ($x_{ij} - x_{\cdot j}$, observation i's deviation from the cluster-aggregated score) plus its between-cluster portion ($x_{\cdot j}$, the aggregate score for cluster j) . We underscore that this within-cluster portion $x_{ij} - x_{\cdot j}$ and between-cluster portion $x_{\cdot j}$ can represent fundamentally different constructs that are each substantively important (see, e.g., Algina & Swaminathan, 2011; Cronbach, 1976; Curran et al., 2012; Curran & Bauer, 2011; Enders & Tofighi, 2007; Hedeker & Gibbons, 2006; Hofmann & Gavin, 1998; Raudenbush & Bryk, 2002; Snijders & Bosker, 2012).

$$\begin{aligned} y_{ij} &= \gamma_{00} + \gamma_{10}(x_{ij} - x_{\cdot j} + x_{\cdot j}) \\ &\quad + u_{1j}(x_{ij} - x_{\cdot j} + x_{\cdot j}) + u_{0j} + e_{ij} \\ &= \gamma_{00} + \gamma_{10}(x_{ij} - x_{\cdot j}) + u_{1j}(x_{ij} - x_{\cdot j}) \\ &\quad + \gamma_{10}x_{\cdot j} + u_{1j}x_{\cdot j} + u_{0j} + e_{ij} \end{aligned} \tag{2}$$

This re-expression in Eq. (2) shows that—as is already well appreciated in the literature—there is conflation

[1]*Grand*-mean centering (or centering by any constant value) has no impact on whether level-specific effects are disaggregated or conflated. Hence we do not distinguish between uncentered vs. grand-mean-centered predictors.

[2]Throughout this paper, to facilitate the distinction of fixed and random conflation, we define the overall *random slope* of x_{ij} (shorthand: "the x_{ij} slope") as the combination of the fixed component—i.e., the γ term—and the random component—i.e., the u term (thus, the random slope, as a whole, can be subject to either fixed or random conflation). This nomenclature is consistent with that found in the MLM literature (see, e.g., Snijders & Bosker, 2012, p. 92), but as a point of clarification, in practice, often the term *random slope* refers specifically to what we term here the random *component* of the slope.

Table 1. Suite of random slope models characterized by full conflation, partial conflation, or no conflation.

Name	Model specification*	**Conflation**?	Commonly used?	Previously recommended for disaggregating level-specific effects?	Currently recommended for disaggregating level-specific effects?	‡ Likelihood equivalent?
Uncentered (or grand-mean-centered) model						
Uncentered random-slope MLM	$y_{ij} = \gamma_{00} + \gamma_{10}x_{ij} + u_{0j} + u_{1j}x_{ij} + e_{ij}$	fixed & random conflation	Yes		No	
Contextual effect models						
Conventional random-slope contextual effect MLM	$y_{ij} = \gamma_{00} + \gamma_{10}x_{ij} + \gamma_{01}x_{\cdot j} + u_{0j} + u_{1j}x_{ij} + e_{ij}$	random conflation	Yes	Hox (2010); Kreft et al. (1995); Antonakis et al. (2019); Longford (1989); Plewis (1989); Snijders and Bosker (2012)	No	
random-slope contextual effect MLM w/ random contextual effect	$y_{ij} = \gamma_{00} + \gamma_{10}x_{ij} + \gamma_{01}x_{\cdot j} + u_{0j} + u_{1j}x_{ij} + u_{2j}x_{\cdot j} + e_{ij}$	unconflated	No		Yes	a
Cluster-mean-centered models						
random-slope cluster-mean-centered MLM	$y_{ij} = \gamma_{00} + \gamma_{10}(x_{ij} - x_{\cdot j}) + u_{0j} + u_{1j}(x_{ij} - x_{\cdot j}) + e_{ij}$	unconflated	Yes		Yes	
random-slope cluster-mean-centered MLM w/ fixed between effect	$y_{ij} = \gamma_{00} + \gamma_{10}(x_{ij} - x_{\cdot j}) + \gamma_{01}x_{\cdot j} + u_{0j} + u_{1j}(x_{ij} - x_{\cdot j}) + e_{ij}$	unconflated	Yes	Raudenbush (1989); Raudenbush and Bryk (2002); Wang and Maxwell (2015); Hofmann and Gavin (1998); Brauer & Curtin (2018)	Yes†	b
random-slope cluster-mean-centered MLM w/ random between effect	$y_{ij} = \gamma_{00} + \gamma_{10}(x_{ij} - x_{\cdot j}) + \gamma_{01}x_{\cdot j} + u_{0j} + u_{1j}(x_{ij} - x_{\cdot j}) + u_{2j}x_{\cdot j} + e_{ij}$	unconflated	No		Yes	a
Hybrid model						
random-slope hybrid contextual effect & cluster-mean-centered MLM	$y_{ij} = \gamma_{00} + \gamma_{10}x_{ij} + \gamma_{01}x_{\cdot j} + u_{0j} + u_{1j}(x_{ij} - x_{\cdot j}) + e_{ij}$	unconflated	No		Yes	b

Notes: MLM = multilevel model.
*For each reduced form model specification listed in Table 1, the corresponding covariance matrices of the random effects are written out in Appendix C.
† = Note that this model assumes intercept homoscedasticity; in this model intercept heteroscedasticity can bias inference about level-2 fixed effects (see manuscript text for details).
‡ Note that likelihood equivalencies between pairs of models are derived in Appendix D.

of the fixed component (here termed *fixed conflation*) in that the fixed component of the slope of $x_{\cdot j}$ (i.e., γ_{10}) is held equal to the fixed component of the slope of $x_{ij} - x_{\cdot j}$ (i.e., γ_{10}).[3] Hence γ_{10} reflects an "uninterpretable blend" (Cronbach, 1976) of within- and between-cluster fixed effects, and will be a weighted average of the two (Raudenbush & Bryk, 2002; Scott & Holt, 1982). Importantly—though *not* well appreciated in the literature—Eq. (2) also shows that there is conflation of the random component of the slope (here termed *random conflation*) in that the random component of the slope of $x_{ij} - x_{\cdot j}$ (i.e., u_{1j}) is held equal to the random component of the slope of $x_{\cdot j}$ (i.e., u_{1j}).

Contextual effect model with random slopes: Unconflation of the fixed component but conflation of the random component

Next we consider the conventional contextual effect model with a random slope of x_{ij}, which continues to be used frequently in applied practice (e.g., Deemer et al., 2017; Hoffman & Stawski, 2009; Lee, 2009; Schempf & Kaufman, 2012). From Eq. (2), we simply add a fixed slope of $x_{\cdot j}$, given by γ_{01} :

$$y_{ij} = \gamma_{00} + \gamma_{10}x_{ij} + \gamma_{01}x_{\cdot j} + u_{1j}x_{ij} + u_{0j} + e_{ij}$$
$$e_{ij} \sim N(0, \sigma^2) \qquad (3)$$
$$\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim MVN\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_{00} & \\ \tau_{10} & \tau_{11} \end{bmatrix}\right)$$

This γ_{01} is the *contextual effect*, that is, the fixed effect of $x_{\cdot j}$ controlling for x_{ij} (which can also be interpreted as the between effect of x_{ij} minus the within effect of x_{ij}). Re-expressing x_{ij} as the sum of its level-specific parts then yields:

$$\begin{aligned} y_{ij} &= \gamma_{00} + \gamma_{10}(x_{ij} - x_{\cdot j} + x_{\cdot j}) \\ &\quad + \gamma_{01}x_{\cdot j} + u_{1j}(x_{ij} - x_{\cdot j} + x_{\cdot j}) + u_{0j} + e_{ij} \\ &= \gamma_{00} + (\gamma_{10} + \gamma_{01})x_{\cdot j} + u_{0j} + u_{1j}x_{\cdot j} \\ &\quad + \gamma_{10}(x_{ij} - x_{\cdot j}) + u_{1j}(x_{ij} - x_{\cdot j}) + e_{ij} \end{aligned} \qquad (4)$$

Here, the within-cluster fixed effect is given as γ_{10}, whereas the between-cluster fixed effect is $\gamma_{10} + \gamma_{01}$. Hence, this re-expression shows that this model unconflates the fixed component of the slope of x_{ij} (since $\gamma_{10} \neq \gamma_{10} + \gamma_{01}$). Unfortunately, however, Eq. (4) shows that this model still *conflates the random component*, as u_{1j} simultaneously represents the random component of the slope of $x_{ij} - x_{\cdot j}$ as well as $x_{\cdot j}$. Hence, researchers fitting this conventional contextual effect model with random slopes for the purposes of disaggregation are paradoxically allowing the fixed components of the slopes of $x_{ij} - x_{\cdot j}$ and $x_{\cdot j}$ to be different but forcing their random components to be equivalent.

The fact that the conventional random-slope contextual effect MLM and the uncentered MLM yield a conflated random component has largely gone unappreciated. It was first briefly noted in an exchange in the *Multilevel Modeling Newsletter* (Raudenbush, 1989; see also Longford, 1989; Plewis, 1989). But in the intervening 30 years it has rarely been mentioned in the methodological literature (Enders & Tofighi, 2007; Hoffman, 2015; Wu & Wooldridge, 2005), and has been virtually ignored in the applied literature—in contrast to the considerable attention and concern that has been paid to conflation of the fixed component in both the methodological and applied literatures. Random conflation may have gone unappreciated because there has been neither elaboration nor demonstration of the specific issues and implications associated with it until now.[4]

As background for explaining why random conflation is problematic, we must first review the distinction between a random component for a purely level-1 predictor (which is already well understood) and that for a purely level-2 predictor (which is not widely understood, according to Goldstein, (2011) and Snijders and Berkhof, (2008)). Having a random component for the slope of a purely level-1 predictor $x_{ij} - x_{\cdot j}$ is well known to imply that the within-cluster effect of $x_{ij} - x_{\cdot j}$ depends on cluster membership, and thus certain clusters have stronger effects than others. This is often called *slope heterogeneity*. Such a situation is illustrated in Figure 1a. In contrast, having a random component for the slope of a purely level-2 predictor may, at first glance, sound counterintuitive

[3]Note that we reserve the use of the word "conflation" for equality constraints between parameters dealing with the within-cluster and the between-cluster portion of x_{ij}; the substitution done in Eq. (2) is simply a mathematically convenient way of demonstrating such implicit equality constraints. Practically speaking, any model implicitly places innumerable constraints, but only those specifically involving $x_{ij} - x_{\cdot j}$ and $x_{\cdot j}$ are relevant to the current discussion.

[4]Prior authors have, however, recognized that, despite the likelihood equivalency of the fixed-slope contextual effect model and fixed-slope cluster-mean centered model (defined later), the random-slope contextual effect model (Equation 3) is not likelihood equivalent to the random-slope cluster-mean-centered model (equal to Equation 3 when replacing x_{ij} with $x_{ij} - x_{\cdot j}$). Methodologists have noted this nonequivalence and have suggested differing implications thereof (see, e.g., Kreft et al., 1995; Enders & Tofighi, 2007; Raudenbush & Bryk, 2002; Snijders & Bosker, 2012). We discuss such explanations later, and explain how the results of the current paper add to the understanding of the differences between these two models.

Panel A: A random component for a slope of a purely level-1 predictor reflects across-cluster *slope heterogeneity*	**Panel B**: A random component for a slope of a purely level-2 predictor reflects across-cluster *intercept heteroscedasticity*

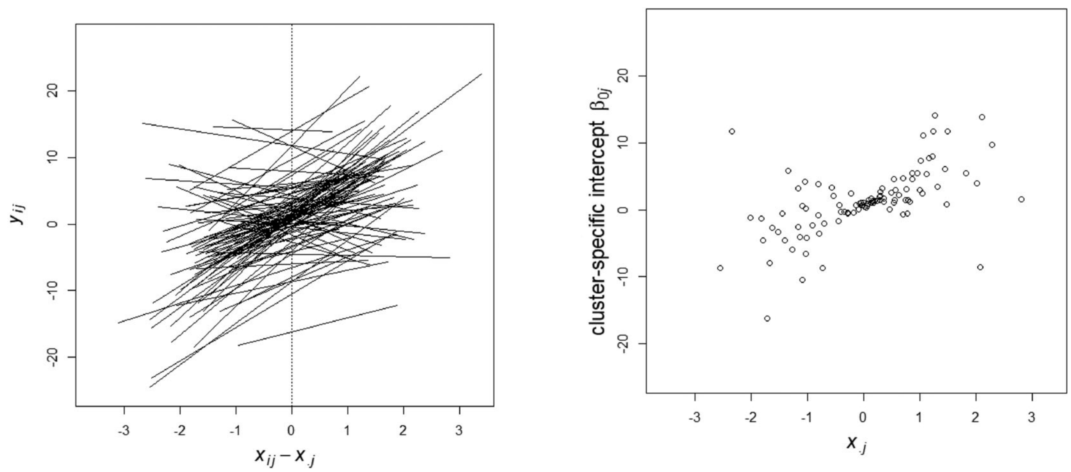

Figure 1. Interpretation of a random component for a slope of a level-1 predictor vs. that of a level-2 predictor. *Notes.* In Panel A, slope heterogeneity is reflected in the fact that the within-cluster effect of $x_{ij} - x_{\cdot j}$ depends on cluster membership. In Panel B, heteroscedastic intercept variance is reflected in the fact that cluster-specific intercepts are more variable at the extremes of $x_{\cdot j}$. Cluster-specific intercepts (i.e., β_{0j} s) were defined in Eq. (A2), and are equivalent to the model-implied cluster means of y_{ij}.

given that every cluster supplies only one score on the level-2 predictor.[5] However, it has been previously shown to be estimable, though with an unfamiliar interpretation—specifically, it has been shown to represent *intercept heteroscedasticity* across clusters (e.g., Goldstein, 2003, 2011 [Chapter 3, p. 85]; Rights & Sterba, 2016; Snijders & Berkhof, 2008; Snijders & Bosker, 1999, 2012 [Chapter 8, pp. 128–129]). To see this, first note that the cluster-specific intercept for the conventional random-slope contextual effect model, β_{0j} (as is defined in Appendix A, where it is equal to the model-implied cluster mean of the outcome; Raudenbush, 1989),[6] implies a *heteroscedastic intercept variance*. This heteroscedastic intercept variance denoted τ_{22j} for cluster j can be expressed as a quadratic function[7] of $x_{\cdot j}$, as derived in detail in Appendix A

$$\begin{aligned}\tau_{22j} &= \text{var}(\beta_{0j}|x_{\cdot j})\\ &= \text{var}(\gamma_{00} + (\gamma_{10} + \gamma_{01})x_{\cdot j} + u_{0j} + u_{1j}x_{\cdot j}|x_{\cdot j})\\ &= \text{var}(u_{0j} + u_{1j}x_{\cdot j}|x_{\cdot j})\\ &= \tau_{00} + 2\tau_{10}x_{\cdot j} + \tau_{11}x_{\cdot j}^2\end{aligned} \tag{5}$$

An example of such heteroscedastic intercept variance is illustrated in Figure 1b, wherein the intercepts are more variable at the extremes of $x_{\cdot j}$.

In the conventional random-slope contextual effect model, the random component of the x_{ij} slope (i.e., $u_{1j)}$ thus simultaneously reflects a blend of *both* slope heterogeneity and intercept heteroscedasticity. That is, though random slope variance in the conventional random-slope contextual effect model is $\text{var}(u_{1j}) = \tau_{11}$, the latter term τ_{11} also appears in Eq. (5). Troublingly, in practice researchers interpret the variance τ_{11} from this conventional random-slope contextual effect model as representing *purely* slope

[5]Indeed, it is commonly taught in introductory MLM courses that one can never include random slopes of level-2 predictors in a two-level model—the intuition being that if each cluster supplies exactly one observation of $x_{\cdot j}$, it does not make sense to estimate a "cluster-specific" effect of $x_{\cdot j}$. However, an inherent aspect to multilevel modeling is that estimation procedures pool across all observations/ clusters when estimating parameters (Gelman & Hill, 2007; Raudenbush & Bryk, 2002), which, in turn, allows estimation of variances and covariances associated with random components of level-2 predictor slopes, as demonstrated later via simulation and via empirical examples.

[6]Though here, and in Appendix A, the cluster-specific intercept $\beta_{0j}$is defined as conditional on the cluster mean of the level-1 predictor (following Raudenbush, 1989), and is thus equivalent to model-implied cluster mean of the outcome for cluster j, the intercept alternatively could be defined unconditionally as $\gamma_{00} + u_{0j}$. The former definition is useful for explicating issues associated with random conflation; however, these issues arise regardless of which intercept definition is used.

[7]In theory the intercept variance could follow some other function (e.g., linear, cubic, etc.) of $x_{\cdot j}$ and/or could vary as a function of some other level-2 variable. Because our goal is to explicate the issues associated with random conflation, here we restrict focus to the intercept variance structure specifically implied by the conventional random-slope contextual effect model.

heterogeneity (i.e., cluster-specific differences in slopes), thus failing to recognize or note that it is conflated with intercept heteroscedasticity.

An alternative (but complimentary) way to conceptualize the random conflation implied by the conventional random slope contextual effect model is that it assumes the variance in y_{ij} across the range of x_{ij} follows the same exact quadratic form as the variance in $y_{\cdot j}$ across the range of $x_{\cdot j}$. See Appendix A for details.

To more precisely and mathematically clarify the highly restrictive assumptions made by the conventional random-slope contextual effect model, in Eq. (6) we introduce a general model expression with separate notation for the two random components associated with the x_{ij} slope, denoted u_{wj} and u_{bj}. Here u_{wj} is the random component associated with the purely within-cluster portion of the x_{ij} slope (i.e., slope of $x_{ij} - x_{\cdot j}$) whereas u_{bj} is the random component associated with the purely between-cluster portion of the x_{ij} slope (i.e., slope of $x_{\cdot j}$). Eq. (6) also introduces separate notation for within-cluster and between-cluster fixed components (γ_w, γ_b).

$$y_{ij} = \gamma_{00} + \gamma_b x_{\cdot j} + \gamma_w(x_{ij} - x_{\cdot j}) + u_{bj}x_{\cdot j} + u_{wj}(x_{ij} - x_{\cdot j}) + u_{0j} + e_{ij}$$
$$e_{ij} \sim N(0, \sigma^2)$$
$$\begin{bmatrix} u_{0j} \\ u_{wj} \\ u_{bj} \end{bmatrix} \sim MVN\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \text{var}(u_{0j}) & & \\ \text{cov}(u_{0j}, u_{wj}) & \text{var}(u_{wj}) & \\ \text{cov}(u_{0j}, u_{bj}) & \text{cov}(u_{wj}, u_{bj}) & \text{var}(u_{bj}) \end{bmatrix}\right) \quad (6)$$

Note that these model terms would be directly estimated in *cluster-mean-centered* (also known as *group-mean-centered*) models discussed later (in which $(x_{ij} - x_{\cdot j})$ is entered as a predictor), but can also be inferred from models with x_{ij} as the predictor (as done, e.g., in Eqs. (2) and (4)). In particular, this representation in Eq. (6) allows us to clarify that the conventional random-slope contextual effect model in Eq. (3) makes the highly restrictive assumption that $u_{bj} = u_{wj}$, yielding random conflation. Appendix B proves that this assumption can be represented in terms of constraints on model parameters as:

a. *equal variances* of u_{wj} and u_{bj} (i.e., $\text{var}(u_{bj}) = \text{var}(u_{wj})$)
b. *perfect correlation* of u_{wj} and u_{bj} ($\text{corr}(u_{bj}, u_{wj}) = 1$).

(with a third constraint implied by the others that $\text{corr}(u_{0j}, u_{wj}) = \text{corr}(u_{0j}, u_{bj})$). Appendix B proves that adding these constraints to the Eq. (6) model yields the conventional random-slope contextual effect model of Eq. (3). When these strict assumptions of equal variance and perfect correlation do not hold, as is likely in practice, the estimated slope variance in the conventional random-slope contextual effect model will be an uninterpretable blend of its 'within-cluster' component (slope heterogeneity) and its 'between-cluster' component (intercept heteroscedasticity), and the conflated random components (u_{1j}'s) will lie between the u_{wj}'s and u_{bj}'s.

Erroneous inferences resulting from fitting conventional random slope contextual effect models

When fitting the conventional random slope contextual effect model, two of the erroneous inferences that are likely to occur as a result of random conflation are: #1) concluding there is no slope heterogeneity when it indeed exists and #2) concluding there is slope heterogeneity when it does not exist. Erroneous inference #1 is likely to arise, for instance, in the commonly theorized situation in which there is indeed slope heterogeneity in the population, but there is no intercept heteroscedasticity in the population. This situation is illustrated in Figure 2, wherein data were generated such that $\text{var}(u_{wj}) > 0$—evident by the heterogeneity in the slope of $x_{ij} - x_{\cdot j}$ depicted in Figure 2a—and $\text{var}(u_{bj}) = 0$—evident by the constant (homoscedastic) intercept variance ($\text{var}(\beta_{0j}|x_{ij})$) across the range of $x_{\cdot j}$ depicted in Figure 2b. In this situation, researchers fitting the conventional random slope contextual effect model will too often erroneously conclude there is a lack of evidence for slope heterogeneity, as demonstrated in Figure 2c where the estimate of $\text{var}(u_{1j})$ is near-zero and non-significant, though slope heterogeneity does truly exist. Erroneous Inference #1 is likely to arise in this situation because the conflated residuals (u_{1j}'s) are weighted *toward* 0 in relation to the u_{wj}'s (since the conflated residuals lie between the u_{wj}'s, which vary about 0, and the u_{bj}'s, which here are all 0). As an applied example of such a situation, consider predicting language test scores from verbal IQ for students nested within school. Across-school slope heterogeneity in the effect of verbal IQ has been theorized (certain schools provide better resources for students to utilize their natural language abilities than other schools) but intercept heteroscedasticity by school-average verbal IQ has not been theorized (e.g., Snijders & Bosker, 2012).

In contrast, Erroneous inference #2 is likely to arise, for instance, when there is *no* slope heterogeneity in the population, but there *is* intercept heteroscedasticity in the population. This situation is illustrated

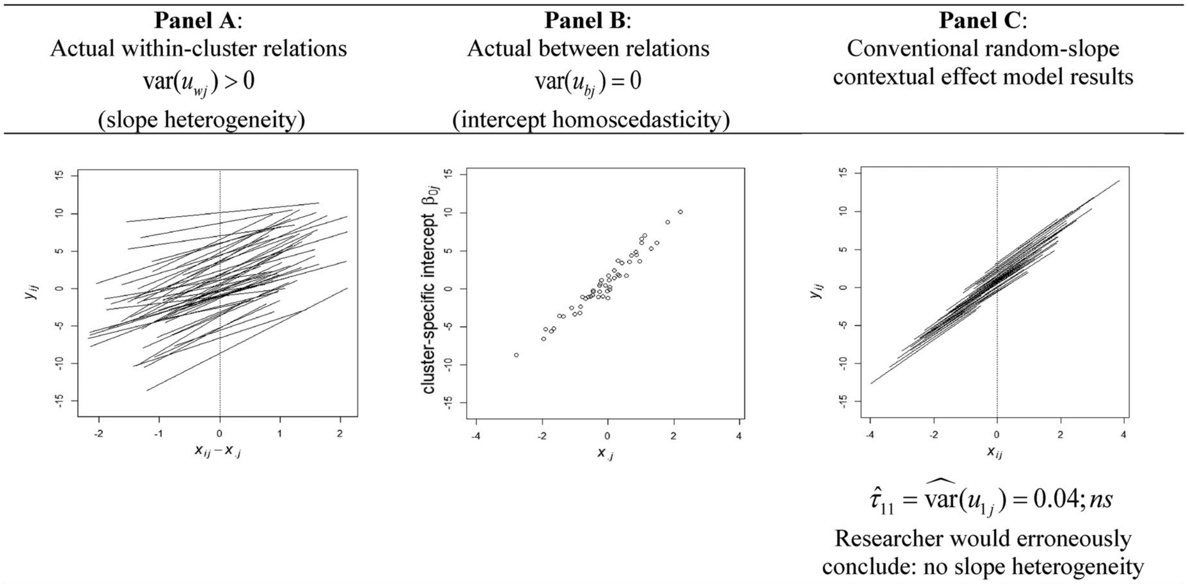

Figure 2. Illustrating Erroneous Inference #1 from the conventional random-slope contextual effect model: Concluding there is no slope heterogeneity when it exists. *Note*. Generating parameters: $\gamma_{00} = 1,\ \gamma_w = 3,\ \gamma_b = 1.5,\ \text{var}(u_{0j}) = 2, \text{var}(u_{wj}) = 2, \text{var}(u_{bj}) = 0$, and $\sigma^2 = 15$ (all other residual (co)variances were 0). Data were generated from Eq. (6) using 50 clusters of size 10. The level-1 x_{ij} was generated as the sum of a within-cluster and between-cluster component, each with variance 1. The significance test for τ_{11} was a mixture chi-square test using a 50:50 mixture of $\chi^2_{df=1}$ and $\chi^2_{df=2}$ as the null reference distribution (Stram & Lee, 1994, 1995). Panel A depicts the generated slope heterogeneity, apparent from the nonparallel cluster-specific regression lines of y_{ij} on $x_{ij} - x_{\cdot j}$ (obtained from the u_{wj}'s). Panel B depicts the generated intercept homoscedasticity, apparent from the constant vertical spread of cluster-specific intercepts on the *y*-axis (obtained from the u_{bj}'s) across the range of $x_{\cdot j}$. Recall that the cluster-specific intercepts (i.e., β_{0j}s) were defined in Eq. (A2), and are equivalent to the model-implied cluster means of y_{ij}. In Panel C we fit the conventional random-slope contextual effect model with *lmer* in R using REML, and find evidence of non-significant slope heterogeneity; this is reflected in the cluster-specific regression lines of y_{ij} on x_{ij} that are nearly parallel.

in Figure 3, wherein data were generated such that $\text{var}(u_{wj}) = 0$—evidenced by the parallel lines in Figure 3a—and $\text{var}(u_{bj}) > 0$—evidenced by the nonconstant/heteroscedastic variance across the range of $x_{\cdot j}$ depicted in Figure 3b. In this situation, a researcher fitting the random slope contextual effect model will too often erroneously conclude there is slope heterogeneity, as demonstrated in Figure 3c wherein the estimate of $\text{var}(u_{1j})$ is significant when there is actually no slope heterogeneity. Erroneous Inference #2 is likely to arise in this situation because the conflated residuals (u_{1j}'s) are weighted *away* from 0 in relation to the u_{wj} (the conflated residuals lie between the u_{wj}'s, which vary about zero, and the u_{bj}'s, which here are all 0). As an applied example of such a situation, consider the prediction of depressive symptoms from therapeutic alliance for patients nested within clinicians. Although no significant slope heterogeneity in the effect of within-clinician deviations in therapeutic alliance on patient depression has been detected (e.g., Baldwin et al., 2007), theory suggests that there could be intercept heteroscedasticity by clinician-mean-alliance. That is, there could be a large amount of across-clinician variability in depression at the extremes of clinician-mean alliance. Extremely high clinician-mean alliance may facilitate depression treatment most for therapists with person-centered orientations (who are following highly individualized treatment plans), and least for behaviorally oriented therapists (who are following pre-established manualized treatment plans). Conversely, low clinician-mean alliance may interfere with treating depression least for behaviorally orientated therapists, but may interfere with treatment most for person-centered therapists. In contrast, moderate levels of clinician-mean alliance may be equally consistent with therapist effectiveness for all orientations and so be associated with less across-clinician variability in patient depression.

In the situations considered thus far, there was either only slope heterogeneity (i.e., only variability in u_{wj}, not u_{bj}, leading to Erroneous Inference #1) or only intercept heteroscedasticity (i.e., only variability in u_{bj}, not u_{wj}, leading to Erroneous Inference #2). Later, in a full-scale simulation, we show how the more general situation wherein both slope heterogeneity and intercept heteroscedasticity occur together (i.e., there is variability in both u_{wj} and u_{bj}) can also lead to erroneous conclusions about the degree of

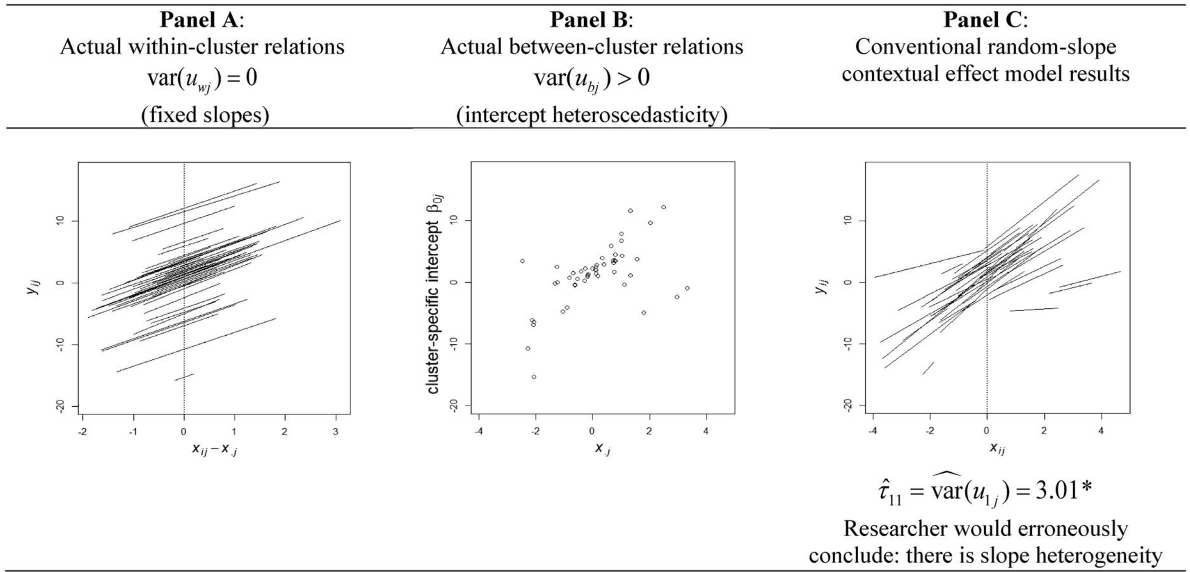

Figure 3. Illustrating Erroneous Inference #2 from the conventional random-slope contextual effect model: Concluding there is slope heterogeneity when it does not exist. *Note*. The significance test for τ_{11} uses the mixture chi-square test described the Figure 2 notes. The generating model, population parameters, and sample size were described in the Figure 2 note with the exception that now var(u_{wj})=0, var(u_{bj}=8), and σ^2=10. Whereas Figure 2 demonstrated the case wherein conflated residuals are weighted toward 0 relative to the u_{wj}'s, the opposite can also occur wherein conflated residuals are weighted away from 0. This is demonstrated in Figure 3, in which there is no generated slope heterogeneity (all u_{wj}'s' are 0; Panel A) but there is generated intercept heteroscedasticity (var(u_{bj})>0; Panel B). Hence in the fitted conventional random slope contextual effect model, the conflated residuals (u_{1j}'s) lie between 0 and the u_{bj}'s, leading to non-parallel cluster-specific regression lines and the conclusion of significant slope heterogeneity (Panel C).

slope heterogeneity (either underestimation or overestimation) when using the conventional random-slope contextual effect model. Such distortion in the random effect structure will then be shown via simulation to result in inaccurate standard errors for fixed components of slopes.

General taxonomy of slope conflation in multilevel models

Though discussed thus far in the context of the conventional random-slope contextual effect model, the terms *fixed conflation* and *random conflation* that we introduced earlier can be used more generally to describe and classify any random-slope MLM containing level-1 variables. Here we do so to provide a general taxonomy of slope conflation for MLMs. Recall that anytime a MLM contains x_{ij}, the MLM can be re-expressed by separating x_{ij} into its pure level-specific parts, $x_{ij} - x_{\cdot j}$ (often called the *cluster-mean-centered*, or *group-mean-centered* predictor) and $x_{\cdot j}$ (the *cluster mean*, or *group mean*). By doing this, one can always express fixed components of a level-1 variable's slope in the form $\gamma_w(x_{ij} - x_{\cdot j}) + \gamma_b x_{\cdot j}$ and random components of a level-1 variable's slope in the form $u_{wj}(x_{ij} - x_{\cdot j}) + u_{bj}x_{\cdot j}$. We provide the taxonomy:

- The MLM is *fully conflated* if it constrains $\gamma_b = \gamma_w$ and $u_{bj} = u_{wj}$.
- The MLM is *partially conflated* if it constrains $\gamma_b = \gamma_w$ (fixed conflation) or $u_{bj} = u_{wj}$ (random conflation), but not both.
- The MLM is *unconflated* if neither of these constraints are made.

For instance, the conventional random-slope contextual effect model constrains $u_{bj} = u_{wj}$ but does not constrain $\gamma_b = \gamma_w$ and is thus partially conflated.

Methods for unconflating random slopes

Here we describe different methods for unconflating random slopes. Corresponding reduced-form and level-specific equations for these unconflated random slope model specifications are given in Appendix C, and reduced-form equations are summarized in Table 1. In Appendix C we also explain why each of these model specifications is fully unconflated. Appendix E provides syntax for fitting each of these unconflated models.

For models that contain x_{ij} as a predictor—i.e., the random-slope contextual effect MLMs or uncentered/grand-mean-centered MLMs—random conflation can be avoided by including a random component for the slope of $x_{\cdot j}$. We demonstrate this in Appendix C. Therein we add the random component u_{2j} to Eq. (3) and show that, in this unconflated model (which we term the *random-slope contextual effect model with random contextual effect* in Table 1, Row 3), $u_{bj} = u_{1j} + u_{2j}$ and $u_{wj} = u_{1j}$ (hence u_{bj} is no longer constrained equal to u_{wj}). Empirical applications almost never fit this model for two reasons. First, including a random component for the slope of $x_{\cdot j}$ in this model (a random contextual effect) had not previously been motivated specifically as a methodological tool for unconflating. Second, including a random component for the slope of $x_{\cdot j}$ simply for the purpose of allowing for intercept heteroscedasticity has also been uncommon because intercept heteroscedasticity is not often considered by substantive theories (Goldstein, 2003, 2011; Rights & Sterba, 2016; Snijders & Berkhof, 2008; Snijders & Bosker, 1999, 2012).

For cluster-mean-centered models that contain a random slope of $x_{ij} - x_{\cdot j}$ (rather than x_{ij}), random conflation is inherently avoided because there is no implicit equality constraint of $u_{wj} = u_{bj}$; hence, unconflation is assured regardless of whether a fixed component for the slope of $x_{\cdot j}$, or random component for the slope of $x_{\cdot j}$, or neither, is included. As such, if one wants to fit a cluster-mean-centered model for unconflation and is interested in only within-cluster effects, one could exclude $x_{\cdot j}$ entirely (noting that $x_{\cdot j}$ is orthogonal to $x_{ij} - x_{\cdot j}$) by fitting the *random-slope cluster-mean-centered model*, given in Appendix C, Eq. (C17) and in Table 1, Row 4. The cluster-mean-centered model remains unconflated if one has substantive interest in also including a fixed between effect for $x_{\cdot j}$ (termed the *random-slope cluster-mean-centered model with fixed between effect* in Appendix C, Eq. (C15) and in Table 1, Row 5). Likewise, the cluster-mean-centered model remains unconflated if one wanted to also include a random between effect for $x_{\cdot j}$ to account for the possibility of intercept heteroscedasticity (termed the *random-slope cluster-mean-centered model with random between effect* in Appendix C, Eq. (C13) and in Table 1, Row 6).

As a final option we consider here, suppose a researcher strongly desired a contextual effect interpretation for the fixed portion of the model, and also wished to avoid random conflation, but did not wish to model a random component for the slope of $x_{\cdot j}$ at level-2. To accomplish this, one could use a novel *hybrid* specification that combines elements of a contextual effect specification (for the fixed portion of the model) with elements of a cluster-mean-centered specification (for the random portion of the model). We call this hybrid specification a *random-slope hybrid contextual effect and cluster-mean-centered MLM* (see Appendix C, Eq. (C19) and Table 1, Row 7). It includes x_{ij} and $x_{\cdot j}$ as predictors for the fixed portion of the model (as in a contextual effect specification), but includes a random component for the slope of $x_{ij} - x_{\cdot j}$ in the random portion of the model to purely reflect slope heterogeneity (as in a cluster-mean-centered specification). Though it is unconventional to utilize a different centering option for the fixed portion of the model vs. the random portion (to our knowledge, no other sources have suggested doing so), this model still affords a sensible—and unconflated—interpretation of parameters. In particular, as shown in Appendix D, in this hybrid specification from the last row in Table 1, γ_{10} represents the within-cluster fixed effect (γ_w), whereas γ_{01} represents the contextual fixed effect ($\gamma_b - \gamma_w$), and the random slope term u_{1j} purely reflects slope heterogeneity (u_{wj}). As derived in Appendix D and illustrated in the upcoming empirical example section, this hybrid specification is likelihood equivalent to the *random-slope cluster-mean-centered MLM with fixed between effect* in Table 1, Row 4, whether x_{ij} (in the fixed portion of the model) is uncentered or centered by any constant value.

In total, Table 1 contains a suite of possible random-slope MLM specifications. For each specification, Column 3 of Table 1 clarifies whether fixed and/or random components are conflated. Table 1 also characterizes MLM specifications by whether they have been commonly used in practice (Column 4) and previously recommended by methodologists (Column 5). Column 4 of Table 1 highlights the noteworthy and concerning fact that the commonly used MLMs include those that are random conflated. Table 1, Column 5, indicates that methodologists have even specifically recommended that the conventional random-slope contextual effect model (Table 1, Row 2) be used to disaggregate level-specific effects, instead of using cluster-mean-centering. Snijders and Bosker (2012), for instance, stated "Generally, one should be reluctant to use cluster-mean-centered random slopes models unless there is a clear theory (or an empirical clue) that not the absolute score x_{ij} but rather the relative score $(x_{ij} - x_{\cdot j})$ is related to y_{ij}."[8] Others have

[8]This statement was made in comparing the *random-slope cluster-mean-centered model with fixed between effect* and the *conventional random-slope contextual effect model*. Breaking this statement down and focusing

claimed that the contextual effect model is less complicated and more interpretable than the cluster-mean-centered model (see our Discussion for more details)—however, importantly they failed to note its inherent random conflation (e.g., Antonakis et al., 2019; Hox, 2010; Kreft et al., 1995). Column 6 of Table 1 highlights that here we recommend only the unconflated MLM specifications as methods for unconflating, and these specifications include those not yet recognized or used for this purpose in practice (i.e., Table 1, Rows 3 and 7). Column 7 of Table 1 indicates which of the specifications in Table 1 are likelihood equivalent (as explained in the next section and as derived in Appendix D). In Appendix E we provide R, SPSS, and SAS syntax to fit each model in Table 1.

Reconciling previous and current understanding of relationships between contextual effect and cluster-mean-centered MLMs

Previous literature has considered relationships between certain of the model specifications listed in Table 1, in particular, the difference between contextual effect models and cluster-mean-centered models. Numerous sources have established, for instance, that the conventional random-slope contextual effect MLM (Table 1, Row 2) is not equivalent to the random-slope cluster-mean-centered MLM with a fixed between effect for $x_{\cdot j}$ (Table 1, Row 5) because of their nonequivalent random effect structures. It had been thought that cluster-mean-centered MLMs and contextual effect MLMs were equivalent *only* when they included fixed slopes, not when they included random slopes (e.g., Kreft et al., 1995; Enders & Tofighi, 2007; Brincks et al., 2017; Curran et al., 2012). However, here we extend this literature by showing that, in fact, a specification of the random-slope cluster-mean-centered MLM and the random-slope contextual effect MLM are indeed analytically equivalent (i.e., analytic reparameterizations of each other, as summarized in Table 1, Column 7). Specifically, we prove mathematically in Appendix D that the following two unconflated models are statistically equivalent—the *random-slope cluster-mean-centered MLM with random between effect* (Table 1, Row 6) and the *random-slope contextual effect MLM with random contextual effect* (Table 1, Row 3). To illustrate, later in an empirical application, we demonstrate how to reparameterize estimates and standard errors from the former model to yield those of the latter model (and vice versa).

Though most prior sources had simply noted the nonequivalence of the conventional random slope contextual effect MLM and random-slope cluster-mean-centered MLMs with a fixed between effect, Raudenbush and Bryk (2002, pp. 145–149) provided a substantive explanation for their discrepancy. They discussed how, when grand-mean-centering (rather than cluster-mean-centering) x_{ij}, estimating the random intercept components entails an extrapolation for clusters that are high or low on x_{ij}, as the zero point of x_{ij} is outside the range of available data. They argued that, because of this extrapolation, the random slope residuals are shrunk toward 0, and hence the random slope variance in a model with grand-mean-centered x_{ij} (including contextual effect models) will be smaller than that of cluster-mean-centered models, which they then demonstrated with an empirical example. However, as we show later via simulation, this explanation does not generally hold when investigating behavior across repeated sampling. If the generating model is, for instance, the random-slope cluster-mean-centered model with fixed between effect (Table 1, Row 5), then the random slope variance of the conventional random slope contextual effect MLM will indeed tend to be smaller than that of the correct fitted model; however, if instead the generating model were the conventional random slope contextual effect MLM (Table 1, Row 2), the random slope variances from both models will, on average, be the *same* value, despite the extrapolation that occurs for the contextual effect model. Furthermore, in certain cases in which there is level-2 heteroscedasticity in the generating model, the average estimated slope variance in the contextual effect model will actually be *much greater* than that of the cluster-mean-centered model. Each of these patterns can be explained by the fact the conflated residuals are, implicitly, a weighted average of two possibly discrepant random components (u_{wj} s and u_{bj} s).

on the random portion of the model (as the fixed portion of both models are analytically equivalent), if x_{ij} is more strongly related to y_{ij} than $(x_{ij} - x_{\cdot j})$ is related to y_{ij}, this implies that $x_{\cdot j}$ is related to y_{ij} via a random component, since the difference between x_{ij} and $(x_{ij} - x_{\cdot j})$ is exactly $x_{\cdot j}$. However, by using the conventional random-slope contextual effect model, the random component of the slopes of both $(x_{ij} - x_{\cdot j})$ and $x_{\cdot j}$ are implicitly assumed to be equivalent. Hence, the analytic result of the current paper shows that this recommendation is logically inconsistent—if, on the random side of the model, x_{ij} is more strongly related to y_{ij} than $(x_{ij} - x_{\cdot j})$, then the underlying random components of $(x_{ij} - x_{\cdot j})$ and $x_{\cdot j}$ *cannot* be the same (if they were the same, as the random conflated model assumes, then x_{ij} and $(x_{ij} - x_{\cdot j})$ would have the *same* strength of relation with y_{ij}).

Table 2. Testing the random slope variance of a level-1 predictor using the conventional random-slope contextual effect model vs. the unconflated random-slope models: Power and Type I error.

		Power	Type I error rate
Random-conflated	Conventional random-slope contextual effect model	.69	.13
Unconflated	Random-slope contextual effect model w/ random contextual effect [a]	.87	.03
	Random-slope cluster-mean-centered model	.92	.04
	Random-slope cluster-mean-centered model w/ fixed between effect [b]	.91	.04
	Random-slope cluster-mean-centered model w/ random between effect [a]	.87	.03
	Random-slope hybrid contextual effect & cluster-mean-centered MLM [b]	.91	.04

Notes. The pair of models with an "a" superscript are likelihood equivalent, as derived in Appendix D. The pair of models with a "b" superscript are likelihood equivalent, as also derived in Appendix D.

Comparing the performance of the conventional random-slope contextual effect MLM to the unconflated MLMs

Earlier, we explained why the conventional random-slope contextual effect model should be especially susceptible to the following issues: concluding there is no slope heterogeneity when it exists (Type II error); concluding there is slope heterogeneity when it does not exist (Type I error); under- or overestimating the degree of slope heterogeneity; and under- or overestimating standard errors (SE) of fixed effects. Here we show via simulation that the unconflated models provide markedly better performance in all of these domains, as these unconflated models avoid the bias induced by random conflation.

Unconflated models provide better power and type I error rates for testing across-cluster slope heterogeneity

To investigate power and Type I error rates across repeated samples, we extend our single-sample illustrations in Figures 2 and 3 by generating 5000 samples using population parameters listed in the Figure 2 notes (for investigating power for detecting slope heterogeneity when $\text{var}(u_{wj}) > 0$ and $\text{var}(u_{bj}) = 0$) and listed in Figure 3 notes (for investigating Type I error for detecting slope heterogeneity when $\text{var}(u_{wj}) = 0$ and $\text{var}(u_{bj}) > 0$).[9] Then we fit each of the models in Table 1 rows 2-7 using *lmer* in R.[10] Restricted maximum likelihood (REML) estimation was used to obtain point estimates and SEs because maximum likelihood (ML) estimation provides biased estimation of random effect variances. ML was used separately in computing mixture LRTs of random slope variances, as the derived null distribution for this test (a 50:50 mixture of $\chi^2_{df=q-1}$ and $\chi^2_{df=q-2}$ where q = # of random effects) assumes ML (Stram & Lee, 1994, 1995). Others recommend using REML for this test (West et al., 2014), though results using REML vs. ML are very similar (e.g., Morrell, 1998).

Power for detecting true slope heterogeneity was computed as the proportion of samples wherein the random slope variance of the level-1 predictor (x_{ij} for the contextual effect models and $x_{ij} - x_{\cdot j}$ for the cluster-mean-centered models) was significant. As is clear from Table 2, Column 1, the widely used conventional random-slope contextual effect model yields much lower power (.69) than all the unconflated models (.87–.92). This is because the lack of intercept heteroscedasticity in the population (see Figure 2) causes the conflated random slope residuals to be weighted toward 0 in relation to the u_{wj}'s, unlike in the unconflated models.

Type I error for detecting slope heterogeneity when it does not exist was computed as the proportion of samples wherein the random slope variance of the level-1 predictor was significant. As evident in Table 2, Column 2, the widely used conventional random-slope contextual effect model yields a large Type I error rate (nearly three times the nominal rate of .05). This is because the presence of intercept heteroscedasticity in the population (see Figure 3) means that the conflated residuals are weighted away from 0 in relation to the u_{wj}'s. All the unconflated models avoid this issue and yield a rate close to the nominal level.

Unconflated models provide less bias in estimating the degree of across-cluster slope heterogeneity

In the above simulation, there was either slope heterogeneity or intercept heteroscedasticity, not both. Here we show how erroneous conclusions about the degree of slope heterogeneity in the conventional random-slope contextual effect model also arise under more general conditions wherein both slope heterogeneity and intercept heteroscedasticity occur together (i.e.,

[9]The set of population effect sizes (in the form of variance explained/ R-squared) for this and all subsequent conditions are provided in Online Appendix A (Rights & Sterba, 2019; Rights & Sterba, 2021).

[10]When any model failed to converge, that sample was excluded. Across all simulations and all conditions, at least 96% of samples were retained.

where there is variability in both the u_{wj} and u_{bj} that contribute to the conflated residual u_{1j}). Specifically, we consider three conditions: when var(u_{wj})=var(u_{bj}), when var(u_{wj})>var(u_{bj}), and when var(u_{wj})<var(u_{bj}). For each condition, we vary the correlation of u_{wj} and u_{bj} across its possible range. We include the first condition (var(u_{wj})=var(u_{bj})) simply to provide a proof-of-concept demonstration that there is no risk of bias only when the assumptions of the conventional random-slope contextual effect model are exactly met. We include the second condition because in empirical practice there can be a larger amount of slope heterogeneity relative to the amount of intercept heteroscedasticity (see e.g., Rights & Sterba, 2016). We include the third condition because in empirical practice there also can be a smaller amount of slope heterogeneity relative to the amount of intercept heteroscedasticity (our upcoming Empirical Example illustrates this).

In the first condition $\text{var}(u_{wj}) = \text{var}(u_{bj}) = 2$; in the second condition we changed $\text{var}(u_{bj}) = 1$ and in the third condition we changed $\text{var}(u_{bj}) = 8$. We generated 5000 repeated samples per condition. Other generating parameters and sample size were the same as previously

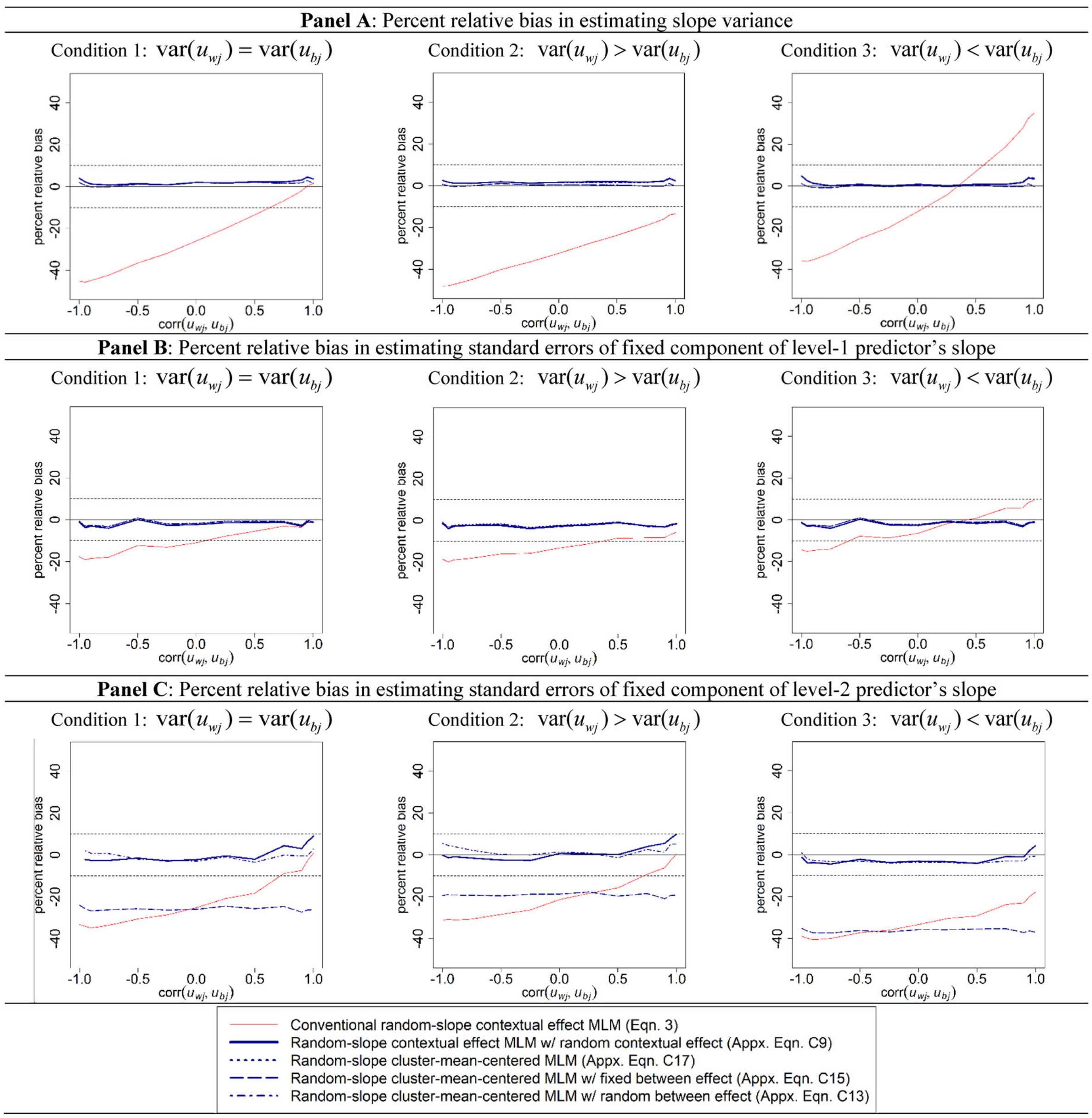

Figure 4. Percent relative bias in random slope variance estimates and fixed effect standard errors for the conventional random-slope contextual effect model versus the unconflated random-slope models. *Notes.* The *random-slope cluster-mean-centered MLM* does not appear in Panel C as it is inapplicable (it does not contain any level-2 predictor). The *random-slope hybrid contextual effect and cluster-mean-centered MLM* does not appear in this figure because its results would be equivalent to those of the *random-slope cluster-mean-centered MLM w/fixed between effect*. The three horizontal (grey) reference lines denote 10, 0, and -10 percent relative bias.

described in the Figure 2 notes. In each condition we fit models in Table 1 rows 2-6 using *lmer* in R. Figure 4 shows the % relative bias in estimating the slope variance $(100 \times (\widehat{\text{var}}(u_{1j}) - \text{var}(u_{wj}))/\text{var}(u_{wj})))$ at each $\text{corr}(u_{wj}, u_{bj})$ of −1, −.95, −.9 −.75, −.5, −.25, 0, .25, .5, .75, .9, .95, or 1. The three horizontal lines in each panel denote where % relative bias is 10%, 0%, and −10%; values outside of this range conventionally indicate meaningfully substantial bias.

Results for Condition 1 indicated that, as a proof-of-concept demonstration supporting our derivations, there is no bias in the estimation of slope heterogeneity when the two strict assumptions of the conventional random-slope contextual effect model are met. This is evident in Figure 4a, Column 1 (i.e., where $\text{var}(u_{wj}) = \text{var}(u_{bj})$) in that the thin (red) line reaches 0 only when $\text{corr}(u_{bj}, u_{wj}) = 1$. The further $\text{corr}(u_{bj}, u_{wj})$ is from 1, the more downward bias. This is because when the correlation is not 1, conflated residuals are, on average, weighted toward 0. For instance, at $\text{corr}(u_{bj}, u_{wj}) = -1$, a u_{wj} of 1 would correspond with a u_{bj} of −1, and hence the conflated residual would be between −1 and 1. In contrast, Figure 4a, Column 1 shows that all unconflated models avoid this issue and yield negligible bias in estimating the degree of slope heterogeneity.

In Condition 2 (where $\text{var}(u_{wj})>\text{var}(u_{bj})$), the conventional random-slope contextual effect model *always underestimates* true slope heterogeneity (see Figure 4a, Column 2). This occurs even when $\text{corr}(u_{bj}, u_{wj}) = 1$ because, for any nonzero value of u_{wj}, the corresponding u_{bj} for cluster j has the same sign but smaller magnitude, hence weighting the conflated residual *closer* to 0 than u_{wj}. In contrast, all unconflated models again avoid this issue yielding negligible bias in Condition 2.

In Condition 3 (where $\text{var}(u_{wj})<\text{var}(u_{bj})$) the conventional random-slope contextual effect model can either *underestimate or overestimate* true slope heterogeneity (see Figure 4a, Column 3). We emphasize this possibility of overestimation, as previous authors have stated that the slope variance of x_{ij} will be smaller than that of $x_{ij} - x_{\cdot j}$ (e.g., Enders & Tofighi, 2007; Raudenbush & Bryk, 2002), though this is not always the case. As for why overestimation is also possible, consider the extreme of $\text{corr}(u_{bj}, u_{wj}) = 1$; for any nonzero value of u_{wj}, the corresponding u_{bj} for cluster j will be of the same sign but larger magnitude, and thus the conflated residual will be weighted *farther* from 0 relative to u_{wj}. All unconflated models again avoid this bias in Condition 3.

Unconflated models provide more accurate standard errors for fixed components of slopes

Thus far we have primarily considered the impact of random conflation on the testing and estimation of the random slope variance itself. Researchers who are interested moreso in fixed effect estimates need to realize, however, that random conflation can adversely affect the fixed portion of the model as well, in that it can yield inappropriate SEs for fixed components of slopes. It is well known in the MLM literature that misspecification of the random effect structure can yield inaccurate SEs for fixed effects (e.g., Raudenbush & Bryk, 2002; Snijders & Bosker, 2012). Typically, such misspecification is discussed in terms of incorrectly omitting a random effect, which, in turn, can lead to estimated fixed effect SEs that do not properly account for across-sample variability in slope heterogeneity, and are thus inaccurate (e.g., leading to elevated Type I error for fixed effects; Barr et al., 2013). To our knowledge, however, no one has demonstrated the adverse impact of random conflation specifically on fixed effect SEs. Due to the lack of general closed-form expressions to obtain SEs in random slope MLMs, here we investigate this adverse impact via simulation. Specifically, under the three conditions described in the previous section, we computed % relative bias for fixed effect SEs (i.e., $100 \times (\widehat{SE_{\hat{\gamma}}} - SD_{\hat{\gamma}})/SD_{\hat{\gamma}})$ where $\widehat{SE_{\hat{\gamma}}}$ is the across-sample average of the analytic SE and $SD_{\hat{\gamma}}$ is the standard deviation of the fixed effect estimate across repeated samples.

Regarding % relative bias for the level-1 fixed effect SE (Figure 4b), the pattern of results mirrors exactly the pattern found for the random slope variance estimates in Figure 4a. That is, for the conventional random-slope contextual effect model, when the random slope variance is underestimated, SEs for $\hat{\gamma}_{10}$ are too small, and when the random slope variance is overestimated, SEs for $\hat{\gamma}_{10}$ are too large. In contrast, all unconflated models provide accurate level-1 fixed effect SEs regardless of the degree of intercept heteroscedasticity or the correlation of u_{wj} and u_{bj}.

Lastly, regarding SEs for fixed effects of level-2 predictors (Figure 4c), the conventional random-slope contextual effect model almost always underestimates these SEs. In contrast, the random-slope contextual effect model with random contextual effect and the cluster-mean-centered model with random between effect provide accurate SEs across conditions (a slight exception being at the extreme boundary of the correlation space). However, it is worth noting that, so long as $\text{var}(u_{bj}) > 0$, the cluster-mean-centered model with only a fixed between effect for $x_{\cdot j}$ (Appendix C, Eq. (C4) or Table 1, Row 5) can underestimate the SE for the fixed effect of $x_{\cdot j}$ (see dashed line in Figure 4c) to an extent proportional to the degree of heteroscedasticity. Hence, unlike the unconflated models in Table 1 Rows 3 or 6, the latter model in Table 1, Row 5 requires intercept homoscedasticity for accurate inference about fixed effects of level-2 predictors.

Impact of sample size and ICC

The simulation depicted graphically in Figure 4 focused primarily on the impact of the variance of, and the correlation between, the random components u_{wj} and u_{bj}. Researchers may also wonder, however, how other factors (e.g., sample size) may impact results, particularly regarding the difference between the unconflated and random conflated models. To assess this, we reran the simulation while changing one of the following: number of clusters (100 vs. 25, as opposed to the original 50), cluster sizes (25 vs. 5, as opposed to the original 10), or average[11] *ICC* ($\tau_{00} = 4$ and $\sigma^2 = 7$ vs. $\tau_{00} = 1$ and $\sigma^2 = 14$, as opposed to the original $\tau_{00} = 2$ and $\sigma^2 = 10$). Results are provided in Online Appendix B, which provides six different versions of the results depicted in Figure 4, each showing results for one of the aforementioned generating conditions. Note that, across all conditions (even the smaller sample sizes), convergence rates remained above 95%.

In terms of level-2 sample size, note that either increasing or decreasing the number of clusters had virtually no impact on the results. This is sensible theoretically, given that changing the number of clusters while holding the cluster size constant retains the relative amount of within-cluster vs. between-cluster information (i.e., the ratio of the number of level-1 units to the number of level-2 units is unchanged). Hence, the extent to which the conflated random slope residuals are weighted toward u_{wj} vs. u_{bj} is not changed from the original simulation.

In terms of the level-1 sample size, in contrast, changing the cluster size did have a meaningful impact on the results. When the cluster size increased to 25 (up from the original 10), the random slope and fixed effect standard error bias induced by random conflation was still present and followed the general pattern of Figure 4, often falling outside the bound of plus-or-minus 10% relative bias; however, the degree of bias was less pronounced. This is consistent with the fact that increasing the number of level-1 units relative to level-2 units in turn increases the relative precision within-cluster vs. between-cluster; hence the conflated residuals are weighted more toward u_{wj}. Conversely, when the cluster size decreased to 5 (down from the original 25), the random slope bias induced by random conflation was much worse than in Figure 4—here, there is decreased precision within-cluster vs. between-cluster, and hence the conflated residuals are weighted more toward u_{bj}. The fixed effect standard error bias for the original conditions and the smaller cluster sizes remained fairly similar.

In terms of average *ICC*, the pattern of results mirror those of cluster size. When the average *ICC* increased, the same general bias pattern held for the random slope variance and the fixed effect SEs for the random-conflated model as was observed in Figure 4, but the bias was less pronounced because the greater relative amount of residual variance between- vs. within-cluster leads to more within-cluster precision (and hence conflated residuals weighted more toward u_{wj}). Conversely, when the average *ICC* decreased, the bias in the random slope for the random conflated model is much worse, as here there is less within-cluster precision and hence the conflated residuals are weighted more toward u_{bj}.

Empirical examples

To concretely illustrate the differences in results obtained from fully, partially, and unconflated MLMs, here we predict math scores using a data set from Kreft and de Leeuw (1998) widely used multilevel modeling textbook. The data set consists of 519 students (level-1) nested with 23 schools (level-2), yielding an average cluster size of approximately 23. We first, for didactic purposes, fit models that match exactly those listed in Table 1 (i.e., same number of predictors and random effects as in Table 1). Then, to provide demonstrations more similar to empirical research practice, we expand and include additional predictors and random effects.

We first fit the popular (but fully conflated) uncentered random slope MLM (i.e., Eq. (1), Table 1, Row 1, with x_{ij} being parents' level of education) using *lmer* in R with REML estimation. The estimated fixed component of the slope was positive and significant (2.90; $p < .05$), suggesting that students with more educated parents tend to have higher math scores. However, this fixed effect estimate is conflated (*fixed conflation*) in that it implicitly reflects both a within-

[11]Note that the *ICC* (i.e., proportion of outcome variance that is between-cluster) is not constant across all conditions in the original simulation. This is because the u_{bj} component contributes to between-cluster outcome variance, and thus the *ICC* is larger when var(u_{bj}) is larger.

cluster fixed effect—the fixed component of a student's parental education relative to their schoolmates—and a between-cluster fixed effect—the fixed component of the school's overall level of parental education. As for the random component of the slope, the slope variance was small and non-significant (1.29; LRT using a 50:50 mixture of $\chi^2_{df=1}$ and $\chi^2_{df=2}$ null distribution yielded $p > .05$). This might lead one to think that there is little across-cluster variability in the association of parent education and math scores. However, this estimate is also conflated (*random conflation*) in that it implicitly reflects both across-cluster slope heterogeneity in the association of parent education with math scores as well as intercept heteroscedasticity by school-mean parent education.

We next fit the conventional random-slope contextual effect model by adding a fixed slope of school-mean parent education (i.e., Eq. (3); Table 1, Row 2). Though this popular model unconflates the fixed component, it still yields a conflated random component. In this case, the conflated slope variance is roughly the same as in the previous model, still small and non-significant (1.51; LRT using a 50:50 mixture of $\chi^2_{df=1}$ and $\chi^2_{df=2}$ null distribution yielded $p > .05$). Researchers fitting this model in practice might erroneously believe that level-specific effects were effectively disaggregated, based on current literature that only emphasizes the fixed component.

Lastly, to fully unconflate, we fit the random slope contextual effect model with random contextual effect (Appendix C, Eq. (C9) or Table 1, Row 3) and the random-slope cluster-mean-centered model with random between effect (Appendix C, Eq. (C13) or Table 1, Row 6). We present results of both in Table 3 to highlight their equivalencies (the table note explains how parameter estimates and SEs from one model can be expressed in terms of those of the other model; corresponding analytic derivations for these equivalencies are in Appendix D). By unconflating the fixed component, we see that both the within fixed effect (2.85) and between fixed effect (5.90) are positive and significant, but that there is a contextual effect ($5.90 - 2.85 = 3.05$) in that the between effect is larger. In other words, overall school-average parent education is more predictive of math scores (in terms of slope magnitude) than is an individual student's parent education. By unconflating the random component, results now indicate significant slope heterogeneity (2.31; an LRT using a 50:50 mix of $\chi^2_{df=2}$, $\chi^2_{df=3}$ null distribution yielding $p < .05$ for either model in Table 3) and significant intercept heteroscedasticity (with the school-specific intercept variance for school j given as $15.38 + 23.90x_{\cdot j} + 10.07x^2_{\cdot j}$; an LRT using a 50:50 mix of $\chi^2_{df=2}$, $\chi^2_{df=3}$ null distribution yielding $p < .05$ for either model in Table 3). In other words, (a) certain schools had a stronger association between average parent education and math score than other schools, and (b) schools with either low or high average parent education had more variability in math scores.

Importantly, results obtained by unconflating the random component are in sharp contrast to what was found in the (widely used) conventional random-slope contextual effect model, wherein the slope variance was small and nonsignificant. When fitting either the widely used *un*centered random slope MLM or the conventional random-slope contextual effect model, the conflated residuals are weighted heavily toward 0 in comparison to the u_{wj}'s. Thus, as in our simulation demonstration, conflating the random component can combine together and inextricably mix up two substantively distinct phenomena (slope heterogeneity and intercept heteroscedasticity), which here gave researchers the problematic impression that there was no slope heterogeneity in the association of parent education with math scores.

It is important to realize that random conflation can cause similar distortion in more complex models with more predictors, and can compromise model results and interpretation in additional ways. Suppose, for instance, one was interested in assessing *why* the slope of parent's level of education may differ across schools, and thus sought to examine cross-level interactions (i.e., interactions between level-1 and level-2 predictors) to account for the slope heterogeneity. When adding cross-level interactions, researchers will often assess the change in the random slope variance of the level-1 predictor going from a model without the interaction to a model with the interaction (Hoffman, 2015; Rights & Sterba, 2020). As an illustration, we will consider school-level SES (i.e., average socioeconomic status among students from the same school) as a possible moderator of the slope of parent education (full results are provided in Table 4). It could be the case, for instance, that parent's level of education is less impactful in schools that are higher in SES and thus likely have more resources available to students. To examine this, we first added a fixed slope of school-level SES to both the aforementioned conventional random-slope contextual effect model and the unconflated random-slope contextual effect model; it was non-significant in both models (point estimates of 11.15 and 9.01, respectively; $p > .05$ for both).

More pertinent to this research question, however, the slope variance of parent education is relatively small

Table 3. Empirical example results for illustrative unconflated random-slope models: Assessing the relationship between parent education and math scores.

	Random-slope contextual effect MLM w/ random contextual effect (Appx. Eq. (C9))	Random-slope cluster-mean-centered MLM w/ random between effect (Appx. Eq. (C13))
Fixed effects		
intercept ($\hat{\gamma}_{00}$)	52.38 (1.08)	52.38 (1.08)
parent education (within effect, $\hat{\gamma}_{10}$)	2.85 (0.49)	—
school-mean-centered parent education (within effect, $\hat{\gamma}_{10}$)	—	2.85 (0.49)
contextual effect of parent education ($\hat{\gamma}_{01}$)	3.05 (1.27)	—
between effect of parent education ($\hat{\gamma}_{01}$)	—	5.90 (1.24)
Random effects		
var. of u_{0j} ($\hat{\tau}_{00}$)	15.38	15.38
var. of parent education slope ($\hat{\tau}_{11}$)	2.31	—
var. of school-mean-centered parent education slope ($\hat{\tau}_{11}$)	—	2.31
var. of contextual effect of parent education ($\hat{\tau}_{22}$)	8.46	—
var. of between effect of parent education ($\hat{\tau}_{22}$)	—	10.07
cov. of u_{0j} & parent education slope residual ($\hat{\tau}_{10}$)	0.71	—
cov. of u_{0j} & school-mean-centered parent education slope residual ($\hat{\tau}_{10}$)	—	0.71
cov. of u_{0j} & contextual effect of parent education slope residual ($\hat{\tau}_{20}$)	11.18	—
cov. of u_{0j} & between effect of parent education slope residual ($\hat{\tau}_{20}$)	—	11.88
cov. of parent education slope residual with contextual effect of parent education slope residual ($\hat{\tau}_{21}$)	−0.35	—
cov. of school-mean-centered parent education slope residual with between effect of parent education slope residual ($\hat{\tau}_{21}$)	—	1.96
level-1 residual variance ($\hat{\sigma}^2$)	70.95	70.95

Note. As derived in Appendix D, $\hat{\gamma}_{01}$, $\hat{\tau}_{20}$, $\hat{\tau}_{21}$, $\hat{\tau}_{22}$, and the standard error of $\hat{\gamma}_{01}$ ($se_{\hat{\gamma}_{01}}$) from the random-slope cluster-mean-centered model w/random between effect are equal to $(\hat{\gamma}_{10}+\hat{\gamma}_{01})$, $(\hat{\tau}_{10}+\hat{\tau}_{20})$, $(\hat{\tau}_{11}+\hat{\tau}_{21})$, $(\hat{\tau}_{11}+\hat{\tau}_{22}+2\hat{\tau}_{21})$, and $\sqrt{se^2_{\hat{\gamma}_{10}}+se^2_{\hat{\gamma}_{01}}+2\text{cov}(\hat{\gamma}_{10},\hat{\gamma}_{01})}$ (with $\text{cov}(\hat{\gamma}_{10},\hat{\gamma}_{01})=-0.16$), respectively, from the random-slope contextual effect model with random contextual effect. All fixed effects were significant (*t*-test with alpha = .05). All random effect variances were significant using mixture likelihood ratio tests described in the Empirical Example section.

and non-significant in the random conflated model (1.66, $p > .05$), whereas it is larger and significant in the unconflated model (2.35, $p < .05$). Similar to the earlier demonstration, the estimate of the heteroscedastic component (for school-mean parent education) is comparatively large and significant (9.57, $p < .05$), and thus the random conflated model's slope variance may be distorted and misleadingly low. As such, in considering school-level SES as a possible explanation of across-school heterogeneity in the effect of parent's education, the conflated model could lead researchers to erroneously think there is no significant such slope variance to begin with. For the unconflated model, adding a fixed slope for the cross-level interaction of school-level SES × parent education reveals a significant fixed effect for the cross-level interaction (−2.18; $p < .05$).[12] This suggests that the effect of parent education is indeed smaller for schools with higher SES—and inclusion of this cross-level interaction in the unconflated model leads to an interpretable decrease in the random slope variance (2.35 to 0.52).[13]

Discussion

This paper began by highlighting the common practice in which researchers use and recommend the conventional random-slope contextual effect model expecting it will disaggregate level-specific effects (e.g., Antonakis et al., 2019; Hox, 2010; Kreft et al.,1995; Snijders & Bosker, 2012). We then showed how this model (and the also-common uncentered random slope model) actually fail to disaggregate the random component of the slope of x_{ij} (i.e., these models conflate slope heterogeneity and intercept

[12]In order to disaggregate the *fixed* component of this cross-level interaction, we additionally included a fixed slope of school-mean SES × school-mean parent education in both the unconflated and random-conflated models. This level-2 interaction itself is not central to the research question (i.e., why the slope of parent's education differs across school), and was non-significant in both models (see Table 4).

[13]If, for the conflated model, one were to ignore the non-significance of the slope variability in the reduced model and nonetheless assess the change in the slope variance after adding the cross-level interaction, one would see the slope variance go to nearly 0 (from 1.66 to <0.01). Importantly, however, this decrease is driven not only by the impact of the product term, but also the extent to which there is random conflation, compromising the utility and interpretability of the result.

Table 4. Empirical example results: Assessing school-mean SES as a possible moderator of the relationship between parent education and math scores.

	Conventional random-slope contextual effect MLM + school-mean SES	Random-slope contextual effect MLM w/ random contextual effect + school-mean SES	Conventional random-slope contextual effect MLM + cross-level interaction	Random-slope contextual effect MLM w/ random contextual effect + cross-level interaction
fixed intercept	51.96 (0.88)*	52.42 (1.04)*	51.52 (1.18)*	52.75 (1.16)*
fixed within effect of parent education	2.77 (0.46)*	2.85 (0.49)*	2.54 (0.36)*	2.59 (0.39)*
fixed contextual effect of school-mean parent education	−4.88 (4.61)	−2.75 (3.90)	−3.58 (4.47)	−1.64 (4.25)
fixed effect of school-mean SES	11.15 (7.38)	9.01 (5.73)	8.78 (6.99)	6.65 (6.82)
fixed effect of parent education × school-mean SES	—	—	−2.40 (0.62)*	−2.18 (0.70)*
fixed effect of school-mean parent education × school-mean SES	—	—	2.35 (1.69)	0.06 (2.15)
variance of random intercept ($\widehat{\text{var}}(u_{0j})$)	12.67*	13.00*	11.65*	13.06*
variance of random slope of parent education	1.66	2.35*	<0.01	0.52
variance of random slope of school-mean parent education	—	9.57*	—	13.74*
cov. of u_{0j} & parent education slope residual	−0.69	0.33	0.22	−0.08
cov. of u_{0j} & school-mean parent education slope residual	—	11.15	—	13.13
cov. of random parent educ & school-mean parent educ slope residual	—	0.19	—	0.51
level-1 residual variance	71.41	70.99	72.13	71.52

Note. $^{*}p < .05$.

heteroscedasticity and are hence termed *random conflated*). After providing a taxonomy differentiating fixed conflation vs random conflation, we demonstrated negative interpretational and inferential consequences of random conflation. Next, we identified a suite of unconflated models to avoid these problems, explaining why they are unconflated (in Appendix C), providing software code for implementing them (in Appendix E), and showing via simulation that, compared to the commonly used random-conflated models, these unconflated models show improved Type I error, power, accuracy of estimates of slope heterogeneity, and accuracy of SEs for fixed effects. Below we provide recommendations for practice, including comments on which unconflated models to use for particular goals as well as remedies if non-convergence is encountered. We then supply future directions.

Recommendations for practice: choosing among unconflated MLMs

In light of our analytic and simulation results, we do not recommend that the (popular) conflated models (Table 1, Rows 1–2) be used in practice. Hence, the key decision-making process centers on which of the unconflated specifications (Table 1 Rows 3–7, with associated software syntax to implement them in Appendix E) to use—key factors to consider are model parsimony, model estimability, and model interpretability.

The first key factor *is model parsimony*, as some unconflated specifications (Table 1 Rows 3 and 6 involving a random component for the slope of $x_{\cdot j}$) have the additional bonus of being able to account for the presence of intercept heteroscedasticity, but are consequently more complex specifications in terms of the number of estimated parameters. In contrast, other unconflated specifications (Table 1, Rows 4, 5, and 7) are more restrictive in assuming intercept homoscedasticity, but are thus more parsimonious specifications in terms of the number of estimated parameters.

A second key factor to consider is *model estimability*. Unconflated specifications that involve a random component for the slope of $x_{\cdot j}$ (in Table 1, Rows 3 and 6) pose a greater chance of nonconverged or improper solutions compared to the simpler unconflated specifications (in Table 1 Rows 4, 5, and 7), especially with larger numbers of random effects, fewer clusters, and less between-cluster variability in the cluster means of predictors and/or less between-cluster variability in the outcome (e.g., Korendijk et al., 2008). If researchers encounter convergence problems with unconflated specifications in Table 1 involving a random component for the slope of $x_{\cdot j}$ (Table 1, Rows 3 and 6), researchers can instead opt for unconflated specifications that do not require a random component for the slope of $x_{\cdot j}$ (Table 1, Rows 4, 5, and 7). In such cases, researchers could of course still also consider accounting for any substantively theorized intercept heteroscedasticity via other approaches, such as those reviewed in Hedeker et al. (2012).

The third key factor to consider is *model interpretability.* If one wants the ability to not only unconflate the random slope, but also account for the potential presence of intercept heteroscedasticity, deciding between the unconflated specifications in Table 1, Rows 3 and 6 is primarily a matter of interpretational preference, as the two are reparameterizations of each other (see Appendix D). One may prefer to interpret the fixed component of the slope of $x_{\cdot j}$ as the contextual effect (in the former model), or interpret it as the fixed between effect of $x_{\cdot j}$ (in the latter model). Likewise one may prefer to interpret the random component of the slope of $x_{\cdot j}$ (i.e., the cluster specific residual u_{2j}) as the cluster-specific difference in u_{wj} and u_{bj}, akin to a cluster-specific contextual effect residual (in the former model), or interpret it as the heteroscedastic intercept component u_{bj} (in the latter model). Some have argued that contextual effect models in general allow more straightforward interpretation and easier testing for the existence of a contextual effect than cluster-mean-centered models (e.g., Algina & Swaminathan, 2011; Hox, 2010; Kelley et al., 2017; Kreft et al., 1995; Snijders & Bosker, 2012). Nonetheless, the presence of a contextual effect can also easily be detected by testing the equality of the fixed within- and between-cluster effects (e.g., Wang & Maxwell, 2015). Another widely cited suggestion (from Enders & Tofighi, 2007) is that grand-mean-centered level-1 predictors be used (without also adding the cluster mean) when the goal is to assess the effect of a level-2 predictor while controlling for a level-1 variable; however, Rights et al. (2020) recently critiqued this practice, showing it can yield severe bias in estimating fixed slopes of level-2 predictors.

On the other hand, if the researcher is willing to assume intercept homoscedasticity and is interested in effects *only* at level-1, the random-slope cluster-mean-centered model (Table 1, Row 4) can be fit to unconflate random slopes. However, if the researcher is willing to assume intercept homoscedasticity and is interested in fixed effects of *both* level-1 and level-2 predictors, deciding between the unconflated specifications in Table 1 rows 5 and 7 is again primarily a matter of interpretational preference—i.e., whether or not one wants a contextual effect interpretation of the fixed effects (Table 1, Row 7) or not (Table 1, Row 5)—as the two specifications are reparameterizations of each other (see Appendix D). Note that if this assumption of intercept homoscedasticity were violated, SEs for the fixed slope of $x_{\cdot j}$ would be inappropriate (as shown earlier in Figure 4c).

Future directions

First, although the assumptions of the conventional random-slope contextual effect model (i.e., $\text{var}(u_{wj}) = \text{var}(u_{bj})$ and $\text{corr}(u_{wj}, u_{bj}) = 1$) are highly restrictive, in practice, it is possible that sample estimates of the random effect covariances could happen to be consistent with these constraints. If desired, one could test if the deviations from these constraints (i.e., unequal variance and imperfect correlation) are statistically significant. For instance, one could compare conventional vs. unconflated random-slope contextual effect models using an adjusted LRT or information criteria. In future work, the power to detect unequal variances and imperfect correlations of u_{wj} and u_{bj} could be investigated. However, in our opinion, the most straightforward approach to avoid random conflation is not to undertake potentially fallible testing of these restrictive assumptions, but rather to simply avoid making them by instead fitting one of the unconflated MLMs.

Second, though we provided results for a variety of possible patterns and correlations of u_{wj}, u_{bj} in Figure 4, future research can capitalize on growing data sharing and open data policies in order to undertake reanalyses of empirical applications from diverse substantive fields that previously fit random-conflated contextual effect models. This will aid in determining which patterns of results (e.g., those resulting from Conditions 1, 2, or 3 from Figure 4) are most common empirically when unconflating the random slope. Such information about $\text{var}(u_{wj})$, $\text{var}(u_{bj})$ and $\text{corr}(u_{wj}, u_{bj})$ cannot be gleaned from output of existing empirical applications that report only the conflated $\text{var}(u_{1j})$.

The topic of the present paper has been on presenting various methods for unconflating, not on presenting various methods for modeling intercept heteroscedasticity. If the latter topic was of primary interest, there are, of course, alternative methods for modeling intercept heteroscedasticity other than inclusion of a random component for the slope of $x_{\cdot j}$ (e.g., Hedeker et al., 2012). Importantly, however, such alternatives are not directly relevant to the notion of random conflation discussed in the current paper, as unconflating the random component for contextual effect models requires explicitly including a random component for the slope of $x_{\cdot j}$. Nonetheless, future research can instead focus on comparing alternative methods of accounting for heteroscedasticity, not only at level-2, but also level-1 (the latter of which was not relevant for the current paper; e.g., Snijders & Berkhof, 2008). Relatedly, such research could investigate possible causes for such heteroscedasticity. For instance, the quadratic pattern of intercept heteroscedasticity discussed here may directly reflect changes in random intercept variance along the cluster mean of x_{ij}, however, it is also possible that this pattern arises due to an omitted interaction between $x_{\cdot j}$ and some additional level-2 predictor.

Last, although here we focused on the impact of random conflation on point estimation, significance testing, and SEs, future work can further explicate its impact on measures of effect size for MLMs, such as R^2 (see Rights & Sterba, [2019] for an integrative framework of MLM R^2). Since conflation can distort either the fixed or random components of level-1 predictor slopes, the estimated proportion of variance attributable to level-1 predictors can similarly be distorted (Rights, 2022). We caution researchers to be wary of interpreting R^2s from conflated models (Rights, 2022).

Conclusion

Although decades of methodological literature has focused on the importance of unconflating level-specific *fixed* effects, it has continued ignoring the possibility of conflating *random* effects. We hope this work helps researchers understand the distinction between random effects associated with level-1 vs. level-2 variables, makes clear the adverse impact of conflating the two, and encourages the use of fully unconflated MLMs in practice.

Acknowledgments

The authors would like to thank Kristopher Preacher, Sun-Joo Cho, and Andrew Tomarken for their helpful comments. Portions of this manuscript are based on the lead author's dissertation. The ideas and opinions expressed herein are those of the authors alone, and endorsement by the authors' institutions or the Natural Science and Engineering Research Council of Canada is not intended and should not be inferred.

Conflict of Interest Disclosures

Each author signed a form for disclosure of potential conflicts of interest. No authors reported any financial or other conflicts of interest in relation to the work described.

Ethical principles

The authors affirm having followed professional ethical guidelines in preparing this work. These guidelines include obtaining informed consent from human participants, maintaining ethical treatment and respect for the rights of human or animal participants, and ensuring the privacy of participants and their data, such as ensuring that individual participants cannot be identified in reported results or from publicly available original or archival data.

Role of the funders/sponsors

None of the funders or sponsors of this research had any role in the design and conduct of the study; collection, management, analysis, and interpretation of data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication.

Funding

This work was supported by Grant RGPIN-2020-06132 from the Natural Science and Engineering Research Council of Canada.

References

Algina, J., & Swaminathan, H. (2011). Centering in two-level nested designs. In J. Hox & J. K. Roberts (Eds.), *Handbook of advanced multilevel analysis* (pp. 285–312). Taylor and Francis. https://doi.org/10.4324/9780203848852.ch15

Antonakis, J., Bastardoz, N., & Rönkkö, M. (2019). On ignoring the random effects assumption in multilevel models: Review, critique, and recommendations. *Organizational Research Methods*, *24*(2), 443–483. https://doi.org/10.1177/1094428119877457

Baldwin, S. A., Wampold, B. E., & Imel, Z. E. (2007). Untangling the alliance-outcome correlation: Exploring the relative importance of therapist and patient variability in the alliance. *Journal of Consulting and Clinical Psychology*, *75*(6), 842–852. https://doi.org/10.1037/0022006X.75.6.842

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. [Database] https://doi.org/10.1016/j.jml.2012.11.001

Bliese, P. D., & Britt, T. W. (2001). Social support, group consensus and stressor–strain relationships: Social context matters. *Journal of Organizational Behavior*, *22*(4), 425–436. https://doi.org/10.1002/job.95

Brauer, M., & Curtin, J. J. (2018). Linear mixed-effects models and the analysis of nonindependent data: A unified framework to analyze categorical and continuous independent variables that vary within-subjects and/or within-items. *Psychological Methods*, *23*(3), 389–411.

Boateng, F. D. (2016). Neighborhood-level effects on trust in the police: A multilevel analysis. *International Criminal Justice Review*, *26*(3), 217–236. https://doi.org/10.1177/1057567716630792

Brincks, A. M., Enders, C. K., Llabre, M. M., Bulotsky-Shearer, R. J., Prado, G., & Feaster, D. J. (2017). Centering predictor variables in three-level contextual models. *Multivariate Behavioral Research*, *52*(2), 149–163. https://doi.org/10.1080/00273171.2016.1256753

Burstein, L. (1980). The analysis of multilevel data in educational research and evaluation. *Review of Research in Education*, *8*, 158–233. https://doi.org/10.3102/0091732X008001158

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

Curran, P. J., & Bauer, D. J. (2011). The disaggregation of within-person and between-person effects in longitudinal models of change. *Annual Review of Psychology*, *62*, 583–619. https://doi.org/10.1146/annurev.psych.093008.100356

Curran, P. J., Lee, T., Howard, A. L., Lane, S., & MacCallum, R. A. (2012). Disaggregating within-person and between-person effects in multilevel and structural equation growth models. In J. R. Harring & G. R. Hancock (Eds.), *Advances in longitudinal methods in the social and behavioral sciences* (pp. 217–253). Information Age.

Deemer, E. D., Marks, L. R., & Miller, K. A. (2017). Peer science self-efficacy: A proximal contextual support for college students' science career intentions. *Journal of Career Assessment*, *25*(3), 537–551. https://doi.org/10.1177/1069072716651620

Diez-Roux, A. V., Link, B. G., & Northridge, M. E. (2000). A multilevel analysis of income inequality and cardiovascular disease risk factors. *Social Science & Medicine*, *50*(5), 673–687. https://doi.org/10.1016/S0277-9536(99)00320-2

Enders, C. K. (2013). Centering predictors and contextual effects. In M. A. Scott, J. S. Simonof, & B. P. Marx (Eds.), *The SAGE handbook of multilevel modeling* (pp. 89–107). Sage. https://doi.org/10.4135/9781446247600.n6

Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods*, *12*(2), 121–138. https://doi.org/10.1037/1082-989X.12.2.121

Espelage, D. L., Holt, M. K., & Henkel, R. R. (2003). Examination of peer–group contextual effects on aggression during early adolescence. *Child Development*, *74*(1), 205–220. https://doi.org/10.1111/1467-8624.00531

Fischer, A. H., Rodriguez Mosquera, P. M., van Vianen, A. E. M., & Manstead, A. S. R. (2004). Gender and culture differences in emotion. *Emotion (Washington, D.C.)*, *4*(1), 87–94. https://doi.org/10.1037/1528-3542.4.1.87

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press.

Goldstein, H. (2003). *Multilevel statistical models* (3rd ed.). Arnold.

Goldstein, H. (2011). *Multilevel statistical models* (4th ed.). Wiley.

Hamaker, E. L., & Grasman, R. P. (2014). To center or not to center? Investigating inertia with a multilevel autoregressive model. *Frontiers in Psychology*, *5*, 1492. https://doi.org/10.3389/fpsyg.2014.01492

Hamaker, E. L., & Muthén, B. (2019). The fixed versus random effects debate and how it relates to centering in multilevel modeling. *Psychological Methods*, *25*(3), 365–379. https://doi.org/10.1037/met0000239

Hatcher, R. L., & Barends, A. W. (2006). How a return to theory could help alliance research. *Psychotherapy: Theory, Research, Practice, Training*, *43*(3), 292–299. https://doi.org/10.1037/00333204.43.3.292

Hedeker, D., & Gibbons, R. (2006). *Longitudinal data analysis*. Wiley. https://doi.org/10.1002/0470036486

Hedeker, D., Mermelstein, R. J., & Demirtas, H. (2012). Modeling between-subject and within-subject variances in ecological momentary assessment data using mixed-effects location scale models. *Statistics in Medicine*, *31*(27), 3328–3336. https://doi.org/10.1002/sim.5338

Henry, K. L., & Slater, M. D. (2007). The contextual effect of school attachment on young adolescents' alcohol use. *Journal of School Health*, *77*(2), 67–74. https://doi.org/10.1111/j.17461561.2007.00169.x

Hoffman, L. (2015). *Longitudinal analysis: Modeling within-person fluctuation and change*. Routledge. https://doi.org/10.4324/9781315744094

Hoffman, L., & Stawski, R. S. (2009). Persons as contexts: Evaluating between-person and within-person effects in longitudinal analysis. *Research in Human Development*, *6*(2-3), 97–120. https://doi.org/10.1080/15427600902911189

Hofmann, D. A., & Gavin, M. B. (1998). Centering decisions in hierarchical linear models: Implications for research in organizations. *Journal of Management*, *24*(5), 623–641. https://doi.org/10.1177/014920639802400504

Hox, J. J. (2010). *Multilevel analysis. Techniques and applications* (2nd ed.) Routledge.

Kelley, J., Evans, M. D. R., Lowman, J., & Lykes, V. (2017). Group-mean-centering independent variables in multilevel models is dangerous. *Quality & Quantity*, *51*(1), 261–283. https://doi.org/10.1007/s11135-015-0304-z

Kidwell, R. E., Jr, Mossholder, K. W., & Bennett, N. (1997). Cohesiveness and organizational citizenship behavior: A multilevel analysis using work groups and individuals. *Journal of Management*, *23*(6), 775–793. https://doi.org/10.1177/014920639702300605

Korendijk, E., Maas, C., Moerbeek, M., & van der, H. (2008). The influence of misspecification of the heteroscedasticity on multilevel regression parameter and standard error estimates. *Methodology*, *4*(2), 67–72. https://doi.org/10.1027/1614-2241.4.2.67

Kreft, I. G., & de Leeuw, J. (1998). *Introducing multilevel modeling*. Sage. https://doi.org/10.4135/9781849209366

Kreft, I. G. G., de Leeuw, J., & Aiken, L. S. (1995). The effects of different forms of centering in hierarchical

linear models. *Multivariate Behavioral Research*, *30*(1), 1–21. https://doi.org/10.1207/s15327906mbr3001_1
Lee, C. C, (2009). Hierarchical linear modeling to explore the influence of satisfaction with public facilities on housing prices. *International Real Estate Review*, *12*(3), 252–272. https://doi.org/10.2307/2112866
Lee, V. E., & Bryk, A. S. (1989). A multilevel model of the social distribution of high school achievement. *Sociology of Education*, *62*(3), 172–192. https://doi.org/10.2307/2112866
Longford, N. (1989). To center or not to center. *Multilevel Modelling Newsletter*, *1*, 7.
Mallinckrodt, B. (2000). Attachment, social competencies, social support, and interpersonal process in psychotherapy. *Psychotherapy Research*, *10*(3), 239–266. https://doi.org/10.1093/ptr/10.3.239
Merlo, J., Chaix, B., Yang, M., Lynch, J., & Råstam, L. (2005). A brief conceptual tutorial on multilevel analysis in social epidemiology: interpreting neighbourhood differences and the effect of neighbourhood characteristics on individual health. *Journal of Epidemiology and Community Health*, *59*(12), 1022–1028. https://doi.org/10.1136/jech.2004.028035
Morrell, C. H. (1998). Likelihood ratio testing of variance components in the linear mixed-effects model using restricted maximum likelihood. *Biometrics*, *54*(4), 1560–1568. https://doi.org/10.2307/2533680
Paccagnella, O. (2006). Centering or not centering in multilevel models? The role of the group mean and the assessment of group effects. *Evaluation Review*, *30*(1), 66–85. https://doi.org/10.1177/0193841X05275649
Plewis, I. (1989). Comment on "centering" predictors. *Multilevel Modelling Newsletter*, *1*, 6–& 11.
Poteat, V. P., Espelage, D. L., & Green, H. D. Jr, (2007). The socialization of dominance: Peer group contextual effects on homophobic and dominance attitudes. *Journal of Personality and Social Psychology*, *92*(6), 1040–1050. https://doi.org/10.1037/0022-3514.92.6.1040
Preacher, K. J. (2011). Multilevel SEM strategies for evaluating mediation in three-level data. *Multivariate behavioral Research*, *46*(4), 691–731. https://doi.org/10.1080/00273171.2011.589280
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.) Sage.
Rights, J. D. (2022). Aberrant distortion of variance components in multilevel models under conflation of level-specific effects. *Psychological Methods*. https://doi.org/10.1037/met0000514
Rights, J. D., Preacher, K. J., & Cole, D. A. (2020). The danger of conflating level-specific effects of control variables when primary interest lies in level-2 effects. *British Journal of Mathematical and Statistical Psychology*, *73*(S1), 194–211. https://doi.org/10.1111/bmsp.12194
Rights, J. D., & Sterba, S. K. (2016). The relationship between multilevel models and nonparametric multilevel mixture models: Discrete approximation of intraclass correlation, random coefficient distributions, and residual heteroscedasticity. *The British Journal of Mathematical and Statistical Psychology*, *69*(3), 316–343. https://doi.org/10.1111/bmsp.12073
Rights, J. D., & Sterba, S. K. (2019). Quantifying explained variance in multilevel models: An integrative framework for defining R-squared measures. *Psychological Methods*, *24*(3), 309–338. https://doi.org/10.1037/met0000184
Rights, J. D., & Sterba, S. K. (2020). New recommendations on the use of R-squared differences in multilevel model comparisons. *Multivariate Behavioral Research*, *55*(4), 568–599. https://doi.org/10.1080/00273171.2019.1660605
Rights, J. D., & Sterba, S. K. (2021). Effect size measures for longitudinal growth analyses: Extending a framework of multilevel model R-squareds to accommodate heteroscedasticity, autocorrelation, nonlinearity, and alternative centering strategies. *New Directions for Child and Adolescent Development*, *2021*(175), 65–110. https://doi.org/10.1002/cad.20387
Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review*, *15*(3), 351–357. https://doi.org/10.2307/2087176
Schempf, A. H., & Kaufman, J. S. (2012). Accounting for context in studies of health inequalities: a review and comparison of analytic approaches. *Annals of Epidemiology*, *22*(10), 683–690. https://doi.org/10.1016/j.annepidem.2012.06.105
Scott, A. J., & Holt, D. (1982). The effect of two-stage sampling on ordinary least squares methods. *Journal of the American Statistical Association*, *77*(380), 848–854. https://doi.org/10.1080/01621459.1982.10477897
Snijders, T. A. B., & Berkhof, J. (2008). Diagnostic checks for multilevel models. In J. de Leeuw & E. Meijer (Eds.), *Handbook of multilevel analysis* (pp. 141–176). Springer. https://doi.org/10.1007/978-0-387-73186-5_3
Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. Sage.
Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling* (2nd ed.). Sage.
Stram, D. O., & Lee, J. W. (1994). Variance components testing in the longitudinal mixed effects model. *Biometrics*, *50*(4), 1171–1177. https://doi.org/10.2307/2533455
Stram, D. O., & Lee, J. W. (1995). Corrections: Variance components testing in the longitudinal mixed effects model. *Biometrics*, *51*, 1196. https://doi.org/10.2307/2533038
Titus, M. A. (2004). An examination of the influence of institutional context on student persistence at 4-year colleges and universities: A multilevel approach. *Research in Higher Education*, *45*(7), 673–699. https://doi.org/10.1023/B:RIHE.0000044227.17161.fa
Wang, L. P., & Maxwell, S. E. (2015). On disaggregating between-person and within-person effects with longitudinal data using multilevel models. *Psychological Methods*, *20*(1), 63–83. https://doi.org/10.1037/met0000030
West, B. T., Welch, K. B., & Galecki, A. T. (2014). *Linear mixed models: A practical guide using statistical software*. CRC Press. https://doi.org/10.1201/b17198
Wu, Y. W. B., & Wooldridge, P. J. (2005). The impact of centering first-level predictors on individual and contextual effects in multilevel data analysis. *Nursing research*, *54*(3), 212–216. https://doi.org/10.1097/00006199-200505000-00009

Appendix A: Derivation of the heteroscedastic intercept variance in Equation 5 that is implied by the conventional random-slope contextual effect model

Here we show that the heteroscedastic intercept variance for the conventional random-slope contextual effect model is given by the quadratic function of $x_{\cdot j}$ in manuscript Equation 5. First we show how the conventional random-slope contextual effect model equation given in Equation 3 can be expressed in terms of cluster-specific intercepts and slopes:

$$y_{ij} = \beta_{0j} + \beta_{1j}(x_{ij} - x_{\cdot j}) + e_{ij} \tag{A1}$$

Here β_{0j} is the cluster-specific intercept, and β_{1j} the cluster-specific slope of $(x_{ij} - x_{\cdot j})$. Defining these as

$$\begin{aligned} \beta_{0j} &= \gamma_{00} + (\gamma_{10} + \gamma_{01})x_{\cdot j} + u_{0j} + u_{1j}x_{\cdot j} \\ \beta_{1j} &= \gamma_{10} + u_{1j} \end{aligned} \tag{A2}$$

yields manuscript Equation 3. Note also that this cluster-specific intercept, β_{0j}, is equal to the model-implied outcome mean for cluster j, as shown in Equation A3:

$$\begin{aligned} E_{i|j}[y_{ij}] &= E_{i|j}\left[\beta_{0j} + \beta_{1j}(x_{ij} - x_{\cdot j}) + e_{ij}\right] \\ &= E_{i|j}\left[\beta_{0j}\right] + E_{i|j}\left[\beta_{1j}\right]E_{i|j}\left[(x_{ij} - x_{\cdot j})\right] + E_{i|j}[e_{ij}] \\ &= E_{i|j}\left[\beta_{0j}\right] \\ &= \gamma_{00} + (\gamma_{10} + \gamma_{01})x_{\cdot j} + u_{0j} + u_{1j}x_{\cdot j} \end{aligned} \tag{A3}$$

Next, noting that, in general, a model's random intercept variance can be expressed as the variance of β_{0j} conditional on predictors (Goldstein, 2011; Rights & Sterba, 2016; Snijders & Bosker, 2012), we compute the random intercept variance of the conventional random-slope contextual effect model as:

$$\begin{aligned} \text{var}(\beta_{0j}|x_{\cdot j}) &= \text{var}(\gamma_{00} + (\gamma_{10} + \gamma_{01})x_{\cdot j} + u_{0j} + u_{1j}x_{\cdot j}|x_{\cdot j}) \\ &= \text{var}(u_{0j} + u_{1j}x_{\cdot j}|x_{\cdot j}) \\ &= \text{var}(u_{0j}) + 2\text{cov}(u_{1j}x_{\cdot j}, u_{0j}|x_{\cdot j}) + \text{var}(u_{1j}x_{\cdot j}|x_{\cdot j}) \\ &= \text{var}(u_{0j}) + 2\text{cov}(u_{1j}, u_{0j})x_{\cdot j} + \text{var}(u_{1j})x_{\cdot j}^2 \\ &= \tau_{00} + 2\tau_{10}x_{\cdot j} + \tau_{11}x_{\cdot j}^2 \end{aligned} \tag{A4}$$

This is equal to τ_{22j} given in Equation 5.

An alternative (but complimentary) way to conceptualize the random conflation implied by the conventional random slope contextual effect model is that it assumes the variance in y_{ij} across the range of x_{ij} follows the same exact same (heteroscedastic) quadratic form as the variance in $y_{\cdot j}$ across the range of $x_{\cdot j}$. In other words, it is well known that, in general, including a random slope in an MLM implies a type of heteroscedasticity in y_{ij} itself, as shown here for the conventional random slope contextual effect model:

$$\begin{aligned} \text{var}(y_{ij}|x_{ij}) &= \text{var}(\gamma_{00} + \gamma_{01}x_{\cdot j} + u_{0j} + \gamma_{10}x_{ij} + u_{1j}x_{ij} + e_{ij}) \\ &= \text{var}(u_{0j} + u_{1j}x_{ij} + e_{ij}) \\ &= \tau_{00} + 2\tau_{01}x_{ij} + \tau_{11}x_{ij}^2 + \sigma^2 \end{aligned} \tag{A5}$$

In fact, this expression is nearly identical to the variance of $y_{\cdot j}$ across the range of $x_{\cdot j}$—i.e., $\text{var}(y_{\cdot j}|x_{\cdot j})$—which is shown in Equation A4, given that $y_{\cdot j}$ is equal to β_{0j}, as delineated in Equation A3. Hence the random slope variance of this random conflated model is simultaneously defining the degree of heteroscedasticity in y_{ij} and in $y_{\cdot j}$.

Appendix B: The conventional random-slope contextual effect model (Equation 3) is nested within the (unconflated) random-slope contextual effect model with random contextual effect (Appendix C Equation C1)

Here we show that the (random-conflated) conventional random-slope contextual effect model is nested within the (unconflated) random-slope contextual effect model with random contextual effect via two constraints on Equation 6: $\text{var}(u_{wj}) = \text{var}(u_{bj})$ and $\text{corr}(u_{wj}, u_{bj}) = 1$ (which technically implies a third constraint that $\text{corr}(u_{0j}, u_{wj}) = \text{corr}(u_{0j}, u_{bj})$). We do this by showing that these two models are equivalent when imposing these constraints.

We first show that $\text{var}(u_{wj}) = \text{var}(u_{bj})$ and $\text{corr}(u_{wj}, u_{bj}) = 1$ together implies $u_{wj} = u_{bj}$.
Note first that, by definition:

$$\begin{aligned} E[u_{wj} - u_{bj}] &= E[u_{wj}] - E[u_{bj}] \\ &= 0 - 0 \\ &= 0 \end{aligned} \tag{B1}$$

Next, if the two constraints hold:

$$\begin{aligned} \text{var}(u_{bj} - u_{wj}) &= \text{var}(u_{bj}) + \text{var}(u_{wj}) - 2\text{cov}(u_{bj}, u_{wj}) \\ &= \text{var}(u_{bj}) + \text{var}(u_{wj}) - 2\sqrt{\text{var}(u_{bj})\text{var}(u_{wj})}\,\text{corr}(u_{bj}, u_{wj}) \\ &= \text{var}(u_{bj}) + \text{var}(u_{wj}) - 2\sqrt{\text{var}(u_{bj})\text{var}(u_{wj})} \\ &= \left(\sqrt{\text{var}(u_{bj})} - \sqrt{\text{var}(u_{wj})}\right)^2 \\ &= \left(\sqrt{\text{var}(u_{bj})} - \sqrt{\text{var}(u_{bj})}\right)^2 \\ &= 0 \end{aligned} \tag{B2}$$

Together, Equations B1 and B2 imply that $u_{wj} - u_{bj} = 0$ for all j, which implies that $u_{wj} = u_{bj}$ for all j. We can thus set the u_{wj} and u_{bj} terms in the unconflated model in Equation 6 generically to u_{1j}, which yields the following model:

$$\begin{aligned} y_{ij} &= \gamma_{00} + \gamma_w(x_{ij} - x_{\cdot j}) + \gamma_b x_{\cdot j} + u_{1j}(x_{ij} - x_{\cdot j}) + u_{1j}x_{\cdot j} + u_{0j} + e_{ij} \\ & e_{ij} \sim N(0, \sigma^2) \\ & \begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim MVN\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_{00} & \tau_{10} \\ \tau_{10} & \tau_{11} \end{bmatrix}\right) \end{aligned} \tag{B3}$$

This is equivalent to the (random-conflated) conventional random-slope contextual effect model, as shown in the parameterization given in manuscript Equation 4. With manuscript Equation 4, we can simply set $\gamma_{10} = \gamma_w$ and $\gamma_{01} = \gamma_b - \gamma_w$, and the expression is the same as Equation B3.

Appendix C: Reduced form and level-specific expressions for all models in Table 1

Here we provide full expressions for each of the models outlined in Table 1, showing both the reduced form and level-specific expressions. For both, we write the model in standard format (in which the level-1 predictor is left in its raw form) as well an equivalent but re-expressed format in which the level-1 predictor is decomposed into a purely level-1 and purely level-2 portion. The latter format is useful in clarifying why each model is either fully conflated, partially conflated, or unconflated (for further details on the latter categories, see the manuscript section titled *General taxonomy of slope conflation in multilevel models*).

The *uncentered random-slope MLM* (Table 1, row 1) was defined in manuscript Equation 1 as:

$$y_{ij} = \gamma_{00} + \gamma_{10}x_{ij} + u_{1j}x_{ij} + u_{0j} + e_{ij}$$
$$e_{ij} \sim N(0, \sigma^2) \tag{C1}$$
$$\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim MVN\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_{00} & \\ \tau_{10} & \tau_{11} \end{bmatrix}\right)$$

This model can, equivalently, be written in level-specific format as follows:

$$\text{Level-1}: y_{ij} = \beta_{0j} + \beta_{1j}x_{ij} + e_{ij}$$
$$\begin{aligned} \text{Level-2}: \beta_{0j} &= \gamma_{00} + u_{0j} \\ \beta_{1j} &= \gamma_{10} + u_{1j} \end{aligned} \tag{C2}$$

where β_{0j} and β_{1j} represent the cluster-specific intercept and slope, respectively. The conflation in this model is more apparent, however, when substituting x_{ij} with the equivalent $(x_{ij}-x_{\cdot j}) + x_{\cdot j}$ and re-writing, yielding a reduced-form expression of:

$$\begin{aligned} y_{ij} &= \gamma_{00} + \gamma_{10}((x_{ij}-x_{\cdot j}) + x_{\cdot j}) + u_{1j}((x_{ij}-x_{\cdot j}) + x_{\cdot j}) + u_{0j} + e_{ij} \\ &= \gamma_{00} + \gamma_{10}(x_{ij}-x_{\cdot j}) + \gamma_{10}x_{\cdot j} + u_{1j}(x_{ij}-x_{\cdot j}) + u_{1j}x_{\cdot j} + u_{0j} + e_{ij} \end{aligned} \tag{C3}$$

and a level-specific expression—with the purely level-1 portion of x_{ij} in the level-1 model and the purely level-2 portion of x_{ij} in the level-2 model—as follows:

$$\text{Level-1}: y_{ij} = \beta_{0j} + \beta_{1j}(x_{ij}-x_{\cdot j}) + e_{ij}$$
$$\begin{aligned} \text{Level-2}: \beta_{0j} &= \gamma_{00} + \gamma_{10}x_{\cdot j} + u_{1j}x_{\cdot j} + u_{0j} \\ \beta_{1j} &= \gamma_{10} + u_{1j} \end{aligned} \tag{C4}$$

This model has fixed conflation because, using the subscripting convention introduced in the more general model expression in manuscript Equation 6, both γ_w and γ_b are equal to γ_{10}. The model similarly has random conflation because both u_{wj} and u_{bj} are equal to u_{1j}.

The *conventional random-slope contextual effect MLM* (Table 1, row 2) was defined in manuscript Equation 3 as:

$$y_{ij} = \gamma_{00} + \gamma_{10}x_{ij} + \gamma_{01}x_{\cdot j} + u_{1j}x_{ij} + u_{0j} + e_{ij}$$
$$e_{ij} \sim N(0, \sigma^2) \tag{C5}$$
$$\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim MVN\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_{00} & \\ \tau_{10} & \tau_{11} \end{bmatrix}\right)$$

And can be written in level-specific format as:

$$\text{Level-1}: y_{ij} = \beta_{0j} + \beta_{1j}x_{ij} + e_{ij}$$
$$\begin{aligned} \text{Level-2}: \beta_{0j} &= \gamma_{00} + \gamma_{01}x_{\cdot j} + u_{0j} \\ \beta_{1j} &= \gamma_{10} + u_{1j} \end{aligned} \tag{C6}$$

Substituting x_{ij} with $(x_{ij}-x_{\cdot j}) + x_{\cdot j}$ and re-writing yields a reduced-form expression of:

$$\begin{aligned} y_{ij} &= \gamma_{00} + \gamma_{10}((x_{ij}-x_{\cdot j}) + x_{\cdot j}) + \gamma_{01}x_{\cdot j} + u_{1j}((x_{ij}-x_{\cdot j}) + x_{\cdot j}) \\ &\quad + u_{0j} + e_{ij} \\ &= \gamma_{00} + \gamma_{10}(x_{ij}-x_{\cdot j}) + (\gamma_{10} + \gamma_{01})x_{\cdot j} + u_{0j} + u_{1j}(x_{ij}-x_{\cdot j}) \\ &\quad + u_{1j}x_{\cdot j} + e_{ij} \end{aligned} \tag{C7}$$

and level-specific expressions of:

$$\text{Level-1}: y_{ij} = \beta_{0j} + \beta_{1j}(x_{ij}-x_{\cdot j}) + e_{ij}$$
$$\begin{aligned} \text{Level-2}: \beta_{0j} &= \gamma_{00} + (\gamma_{10} + \gamma_{01})x_{\cdot j} + u_{1j}x_{\cdot j} + u_{0j} \\ \beta_{1j} &= \gamma_{10} + u_{1j} \end{aligned} \tag{C8}$$

The fixed component is disaggregated because $\gamma_w = \gamma_{10}$ and $\gamma_b = (\gamma_{10} + \gamma_{01})$ (and hence γ_w is not constrained equal to γ_b), but the random component is conflated because $u_{wj} = u_{bj} = u_{1j}$.

The *random-slope contextual effect MLM with random contextual effect* (Table 1 Row 3) adds to the *conventional random-slope contextual effect MLM* a random component for the slope of $x_{\cdot j}$, u_{2j}, and is given as:

$$y_{ij} = \gamma_{00} + \gamma_{10}x_{ij} + u_{1j}x_{ij} + \gamma_{01}x_{\cdot j} + u_{2j}x_{\cdot j} + u_{0j} + e_{ij}$$
$$e_{ij} \sim N(0, \sigma^2)$$
$$\begin{bmatrix} u_{0j} \\ u_{1j} \\ u_{2j} \end{bmatrix} \sim MVN\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_{00} & & \\ \tau_{10} & \tau_{11} & \\ \tau_{20} & \tau_{21} & \tau_{22} \end{bmatrix}\right) \tag{C9}$$

This model can be written in level-specific format as:

$$\text{Level-1}: y_{ij} = \beta_{0j} + \beta_{1j}x_{ij} + e_{ij}$$
$$\begin{aligned} \text{Level-2}: \beta_{0j} &= \gamma_{00} + \gamma_{01}x_{\cdot j} + u_{0j} + u_{2j}x_{\cdot j} \\ \beta_{1j} &= \gamma_{10} + u_{1j} \end{aligned} \tag{C10}$$

The disaggregation of both the fixed and random components is apparent when substituting x_{ij} with $(x_{ij}-x_{\cdot j}) + x_{\cdot j}$, yielding the following reduced-form expression:

$$\begin{aligned} y_{ij} &= \gamma_{00} + \gamma_{10}(x_{ij}-x_{\cdot j} + x_{\cdot j}) + u_{1j}(x_{ij}-x_{\cdot j} + x_{\cdot j}) + \gamma_{01}x_{\cdot j} \\ &\quad + u_{2j}x_{\cdot j} + u_{0j} + e_{ij} \\ &= \gamma_{00} + \gamma_{10}(x_{ij}-x_{\cdot j}) + u_{1j}(x_{ij}-x_{\cdot j}) + (\gamma_{10} + \gamma_{01})x_{\cdot j} \\ &\quad + (u_{1j} + u_{2j})x_{\cdot j} + u_{0j} + e_{ij} \end{aligned} \tag{C11}$$

and the following level-specific expressions:

$$\text{Level-1}: y_{ij} = \beta_{0j} + \beta_{1j}(x_{ij}-x_{\cdot j}) + e_{ij}$$
$$\begin{aligned} \text{Level-2}: \beta_{0j} &= \gamma_{00} + (\gamma_{10} + \gamma_{01})x_{\cdot j} + u_{0j} + (u_{1j} + u_{2j})x_{\cdot j} \\ \beta_{1j} &= \gamma_{10} + u_{1j} \end{aligned} \tag{C12}$$

The fixed component is disaggregated because $\gamma_w = \gamma_{10}$ and $\gamma_b = (\gamma_{10} + \gamma_{01})$, and the random component is disaggregated because $u_{wj} = u_{1j}$ and $u_{bj} = (u_{1j} + u_{2j})$ (and hence $u_{wj} \neq u_{bj}$).

The *random-slope cluster-mean-centered MLM with random between effect* (Table 1 Row 6) is analytically equivalent to the *random-slope contextual effect MLM with random between effect* (as shown in Appendix D), and the former is given as:

$$y_{ij} = \gamma_{00} + \gamma_{10}(x_{ij} - x_{\cdot j}) + u_{1j}(x_{ij} - x_{\cdot j}) + \gamma_{01}x_{\cdot j} + u_{2j}x_{\cdot j} + u_{0j} + e_{ij}$$

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

$$\begin{bmatrix} u_{0j} \\ u_{1j} \\ u_{2j} \end{bmatrix} \sim MVN\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_{00} & & \\ \tau_{10} & \tau_{11} & \\ \tau_{20} & \tau_{21} & \tau_{22} \end{bmatrix} \right) \tag{C13}$$

which can be written in level-specific format (noting that x_{ij} is already decomposed into a purely level-1 and purely level-2 portion) as:

$$\text{Level-1}: y_{ij} = \beta_{0j} + \beta_{1j}(x_{ij} - x_{\cdot j}) + e_{ij}$$

$$\begin{aligned} \text{Level-2}: \beta_{0j} &= \gamma_{00} + \gamma_{01}x_{\cdot j} + u_{0j} + u_{2j}x_{\cdot j} \\ \beta_{1j} &= \gamma_{10} + u_{1j} \end{aligned} \tag{C14}$$

Here there is no conflation, as $\gamma_w = \gamma_{10}$, $\gamma_b = \gamma_{01}$, $u_{wj} = u_{1j}$, and $u_{bj} = u_{2j}$.

The *random-slope cluster-mean-centered MLM with fixed between effect* (Table 1 Row 5) is more parsimonious than the *random-slope cluster-mean-centered MLM with random between effect* in that it excludes the random component for the slope of $x_{\cdot j}$, and is thus given as:

$$y_{ij} = \gamma_{00} + \gamma_{10}(x_{ij} - x_{\cdot j}) + \gamma_{01}x_{\cdot j} + u_{1j}(x_{ij} - x_{\cdot j}) + u_{0j} + e_{ij}$$

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

$$\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim MVN\left(\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix}, \begin{bmatrix} \tau_{00} & \\ \tau_{10} & \tau_{11} \end{bmatrix} \right) \tag{C15}$$

and can be written in level-specific format as:

$$\text{Level-1}: y_{ij} = \beta_{0j} + \beta_{1j}(x_{ij} - x_{\cdot j}) + e_{ij}$$

$$\begin{aligned} \text{Level-2}: \beta_{0j} &= \gamma_{00} + \gamma_{01}x_{\cdot j} + u_{0j} \\ \beta_{1j} &= \gamma_{10} + u_{1j} \end{aligned} \tag{C16}$$

Here there is no conflation, as $\gamma_w = \gamma_{10}$, $\gamma_b = \gamma_{01}$, $u_{wj} = u_{1j}$, and $u_{bj} = 0$.

The *random-slope cluster-mean-centered MLM* (Table 1 Row 4) is more parsimonious than the *random-slope cluster-mean-centered MLM with fixed between effect* in that it excludes a fixed component for the slope of $x_{\cdot j}$, and is thus given as:

$$y_{ij} = \gamma_{00} + \gamma_{10}(x_{ij} - x_{\cdot j}) + u_{1j}(x_{ij} - x_{\cdot j}) + u_{0j} + e_{ij}$$

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

$$\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim MVN\left(\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix}, \begin{bmatrix} \tau_{00} & \\ \tau_{10} & \tau_{11} \end{bmatrix} \right) \tag{C17}$$

which can be written in level-specific format as:

$$\text{Level-1}: y_{ij} = \beta_{0j} + \beta_{1j}(x_{ij} - x_{\cdot j}) + e_{ij}$$

$$\begin{aligned} \text{Level-2}: \beta_{0j} &= \gamma_{00} + u_{0j} \\ \beta_{1j} &= \gamma_{10} + u_{1j} \end{aligned} \tag{C18}$$

Here there is no conflation, as $\gamma_w = \gamma_{10}$, $\gamma_b = 0$, $u_{wj} = u_{1j}$, and $u_{bj} = 0$.

Lastly, the *hybrid random-slope contextual effect cluster-mean-centered MLM* (Table 1 Row 7) is analytically equivalent to the *random-slope cluster-mean-centered MLM with fixed between effect* (Table 1 Row 5). The former specifies the fixed effects as in a conventional contextual effect model, but rather than a random slope of x_{ij}, there is a random slope of $x_{ij} - x_{\cdot j}$. It is thus given as:

$$y_{ij} = \gamma_{00} + \gamma_{10}x_{ij} + \gamma_{01}x_{\cdot j} + u_{0j} + u_{1j}(x_{ij} - x_{\cdot j}) + e_{ij} \tag{C19}$$

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

$$\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim MVN\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_{00} & \\ \tau_{10} & \tau_{11} \end{bmatrix} \right)$$

Writing this model is traditional multilevel format yields an atypical expression, given that both x_{ij} and $(x_{ij} - x_{\cdot j})$ are level-1 predictors, but their slopes don't each have fixed and random components:

$$\text{Level-1}: y_{ij} = \beta_{0j} + \beta_{1j}x_{ij} + \beta_{2j}(x_{ij} - x_{\cdot j}) + e_{ij}$$

$$\begin{aligned} \text{Level-2}: \beta_{0j} &= \gamma_{00} + \gamma_{01}x_{\cdot j} + u_{0j} \\ \beta_{1j} &= \gamma_{10} \\ \beta_{2j} &= u_{1j} \end{aligned} \tag{C20}$$

The disaggregation of both the fixed and random components is apparent when substituting x_{ij} with $(x_{ij} - x_{\cdot j}) + x_{\cdot j}$, yielding the following reduced-form expression:

$$\begin{aligned} y_{ij} &= \gamma_{00} + \gamma_{10}(x_{ij} - x_{\cdot j} + x_{\cdot j}) + \gamma_{01}x_{\cdot j} + u_{0j} + u_{1j}(x_{ij} - x_{\cdot j}) + e_{ij} \\ &= \gamma_{00} + \gamma_{10}(x_{ij} - x_{\cdot j}) + (\gamma_{10} + \gamma_{01})x_{\cdot j} + u_{0j} + u_{1j}(x_{ij} - x_{\cdot j}) + e_{ij} \end{aligned} \tag{C21}$$

and the following (less atypical) level-specific expressions:

$$\text{Level-1}: y_{ij} = \beta_{0j} + \beta_{1j}(x_{ij} - x_{\cdot j}) + e_{ij}$$

$$\begin{aligned} \text{Level-2}: \beta_{0j} &= \gamma_{00} + (\gamma_{10} + \gamma_{01})x_{\cdot j} + u_{0j} \\ \beta_{1j} &= \gamma_{10} + u_{1j} \end{aligned} \tag{C22}$$

Here there is no conflation, as $\gamma_w = \gamma_{10}$, $\gamma_b = \gamma_{10} + \gamma_{01}$, $u_{wj} = u_{1j}$, and $u_{bj} = 0$.

Appendix D: Deriving equivalencies between unconflated MLMs: Analytic equivalency between Table 1 Row 3 & 6 MLM and analytic equivalency between Table 1 Row 5 & 7 MLMs

In this section we first show that the (unconflated) *random-slope contextual effect MLM with random between effect* (Table 1 Row 3) is equivalent to the (unconflated) *random-slope cluster-mean-centered MLM with random between effect* (Table 1 Row 6), such that the two are re-parameterizations of each other. The *random-slope contextual effect MLM with random contextual effect* can be re-expressed as

$$\begin{aligned}y_{ij} &= \gamma_{00} + \gamma_{10}x_{ij} + u_{1j}x_{ij} + \gamma_{01}x_{\cdot j} + u_{2j}x_{\cdot j} + u_{0j} + e_{ij} \\ &= \gamma_{00} + \gamma_{10}(x_{ij} - x_{\cdot j} + x_{\cdot j}) + u_{1j}(x_{ij} - x_{\cdot j} + x_{\cdot j}) \\ &\quad + \gamma_{01}x_{\cdot j} + u_{2j}x_{\cdot j} + u_{0j} + e_{ij} \\ &= \gamma_{00} + \gamma_{10}(x_{ij} - x_{\cdot j}) + \gamma_{10}x_{\cdot j} + u_{1j}(x_{ij} - x_{\cdot j}) + u_{1j}x_{\cdot j} \\ &\quad + \gamma_{01}x_{\cdot j} + u_{2j}x_{\cdot j} + u_{0j} + e_{ij} \\ &= \gamma_{00} + \gamma_{10}(x_{ij} - x_{\cdot j}) + (\gamma_{10} + \gamma_{01})x_{\cdot j} + u_{1j}(x_{ij} - x_{\cdot j}) \\ &\quad + (u_{1j} + u_{2j})x_{\cdot j} + u_{0j} + e_{ij}\end{aligned} \tag{D1}$$

Letting γ_{01}^* and u_{2j}^* denote, respectively, the fixed and random component of the slope of $x_{\cdot j}$ from the *random-slope cluster-mean-centered MLM with random between effect*, we define these parameters in terms of *random-slope contextual effect MLM with random contextual effect* parameters as $\gamma_{01}^* = \gamma_{10} + \gamma_{01}$ and $u_{2j}^* = u_{1j} + u_{2j}$ to show the equivalence of the reduced-form expressions in the first lines of Appendix Equations C9 and C13 (all other terms between the models on these lines are the same). This implies that the standard error of $\hat{\gamma}_{01}^*$ can be expressed as the square root of a function of the asymptotic variances and covariances of the estimates from the *random-slope contextual effect MLM with random contextual effect*, as such:

$$\begin{aligned}se_{\hat{\gamma}_{01}^*} &= \sqrt{\text{var}(\hat{\gamma}_{01}^*)} \\ &= \sqrt{\text{var}(\hat{\gamma}_{10} + \hat{\gamma}_{01})} \\ &= \sqrt{\text{var}(\hat{\gamma}_{10}) + \text{var}(\hat{\gamma}_{01}) + 2\text{cov}(\hat{\gamma}_{10}, \hat{\gamma}_{01})} \\ &= \sqrt{se_{\hat{\gamma}_{10}}^2 + se_{\hat{\gamma}_{01}}^2 + 2\text{cov}(\hat{\gamma}_{10}, \hat{\gamma}_{01})}\end{aligned} \tag{D2}$$

Further, we show the equivalence of the random effect covariances by first letting τ_{20}^*, τ_{21}^*, and τ_{22}^* denote variance and covariance components from the *random-slope cluster-mean-centered MLM with random between effect*, and then expressing the random effect covariance structure of this model as

$$\begin{bmatrix} u_{0j} \\ u_{1j} \\ u_{2j}^* \end{bmatrix} = \begin{bmatrix} u_{0j} \\ u_{1j} \\ u_{1j} + u_{2j} \end{bmatrix} \sim MVN\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_{00} & & \\ \tau_{10} & \tau_{11} & \\ \tau_{20}^* & \tau_{21}^* & \tau_{22}^* \end{bmatrix}\right) \tag{D3}$$

with u_{0j}, u_{1j}, τ_{00}, τ_{10}, and τ_{11} being the same for either model. We next express this random effect covariance structure purely in terms of the *random-slope contextual effect MLM with random contextual effect* parameters by considering that

$$\begin{aligned}\tau_{20}^* &= \text{cov}(u_{1j} + u_{2j}, u_{0j}) \\ &= \text{cov}(u_{1j}, u_{0j}) + \text{cov}(u_{2j}, u_{0j}) \\ &= \tau_{10} + \tau_{20}\end{aligned} \tag{D4}$$

and

$$\begin{aligned}\tau_{21}^* &= \text{cov}(u_{1j} + u_{2j}, u_{1j}) \\ &= \text{cov}(u_{1j}, u_{1j}) + \text{cov}(u_{2j}, u_{1j}) \\ &= \text{var}(u_{1j}) + \text{cov}(u_{2j}, u_{1j}) \\ &= \tau_{11} + \tau_{21}\end{aligned} \tag{D5}$$

and

$$\begin{aligned}\tau_{22}^* &= \text{var}(u_{1j} + u_{2j}) \\ &= \text{var}(u_{1j}) + \text{var}(u_{2j}) + 2\text{cov}(u_{1j}, u_{2j}) \\ &= \tau_{11} + \tau_{22} + 2\tau_{21}\end{aligned} \tag{D6}$$

Thus,

$$\begin{bmatrix} u_{0j} \\ u_{1j} \\ u_{1j} + u_{2j} \end{bmatrix} \sim MVN\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_{00} & & \\ \tau_{10} & \tau_{11} & \\ \tau_{10} + \tau_{20} & \tau_{11} + \tau_{21} & \tau_{11} + \tau_{22} + 2\tau_{21} \end{bmatrix}\right) \tag{D7}$$

Hence, the parameters of the *random-slope cluster-mean-centered MLM with random between effect* can be equivalently expressed in terms of *random-slope contextual effect MLM with random contextual effect* (Table 1 Row 5), such that the two are re-parameterizations of each other.

We next show that the (unconflated) *random-slope cluster-mean-centered MLM with fixed between effect* (Table 1 Row 5) is equivalent to the (unconflated) *random-slope hybrid contextual effect and cluster-mean-centered MLM* (Table 1 Row 7). The latter can be re-expressed as

$$\begin{aligned}y_{ij} &= \gamma_{00} + \gamma_{10}x_{ij} + \gamma_{01}x_{\cdot j} + u_{0j} + u_{1j}(x_{ij} - x_{\cdot j}) + e_{ij} \\ &= \gamma_{00} + \gamma_{10}(x_{ij} - x_{\cdot j} + x_{\cdot j}) + \gamma_{01}x_{\cdot j} + u_{0j} + u_{1j}(x_{ij} - x_{\cdot j}) + e_{ij} \\ &= \gamma_{00} + \gamma_{10}(x_{ij} - x_{\cdot j}) + (\gamma_{01} + \gamma_{10})x_{\cdot j} + u_{0j} + u_{1j}(x_{ij} - x_{\cdot j}) + e_{ij}\end{aligned} \tag{D8}$$

Letting γ_{01}^* denote the fixed component of the slope of $x_{\cdot j}$ from the *random-slope cluster-mean-centered MLM with fixed between effect*, we define this parameter in terms of the *random-slope hybrid contextual effect and cluster-mean-centered MLM* parameters as $\gamma_{01}^* = \gamma_{10} + \gamma_{01}$ to show the equivalence of the reduced-form expressions in the first lines of Appendix Equations C15 and C19 (all other terms between the models on these lines are the same). This implies that the standard error of $\hat{\gamma}_{01}^*$ can be expressed as the square root of a function of the asymptotic variances and covariances of the estimates from the *random-slope contextual effect MLM with fixed contextual effect*, as such:

$$\begin{aligned}se_{\hat{\gamma}_{01}^*} &= \sqrt{\text{var}(\hat{\gamma}_{01}^*)} \\ &= \sqrt{\text{var}(\hat{\gamma}_{10} + \hat{\gamma}_{01})} \\ &= \sqrt{\text{var}(\hat{\gamma}_{10}) + \text{var}(\hat{\gamma}_{01}) + 2\text{cov}(\hat{\gamma}_{10}, \hat{\gamma}_{01})} \\ &= \sqrt{se_{\hat{\gamma}_{10}}^2 + se_{\hat{\gamma}_{01}}^2 + 2\text{cov}(\hat{\gamma}_{10}, \hat{\gamma}_{01})}\end{aligned} \tag{D2}$$

Hence, the parameters of the *random-slope cluster-mean-centered MLM with fixed between effect* (Table 1 Row 5), can be equivalently expressed in terms of the *random-slope hybrid contextual effect and cluster-mean-centered MLM* (Table 1 Row 7), such that the two are re-parameterizations of each other.

Note that the aforementioned equivalencies hold whenever any independent variable (for either model) is centered by any constant value, e.g., the grand mean. In general, in both MLM and single-level regression, centering by a constant will change the interpretation of the intercept and, in models with higher-order terms (e.g., interactive models or polynomial models of at least order 2), will change the interpretation of the slopes of lower-order terms, but always results in a likelihood equivalent model to the version that is not centered by said constant.

Appendix E: R, SPSS, and SAS syntax for each model listed in Table 1

Here we provide syntax with which researchers can fit each model listed in Table 1 in either R (using the function *lmer* from the package *lme4*), SPSS (using the MIXED command), or SAS (using the procedure PROC MIXED). The below syntax assumes that the dataset (called "data" below) contains the following variables:

yij = the outcome of interest
xij = the uncentered (or grand-mean-centered) level-1 predictor
xij_clusmean = the cluster mean (or group mean) of the level-1 predictor
xij_gmc = the cluster-mean-centered (or group-mean-centered) level-1 predictor
clusterID = cluster-specific identification

Uncentered random-slope MLM (Table 1 Row 1):

***lmer* syntax:**
```
fit_mod1 <- lmer(yij ~ 1 + xij + (1 + xij | clusterID), data = data)
summary(fit_mod1)
```

SPSS MIXED syntax:
```
MIXED yij WITH xij
/FIXED = INTERCEPT xij
/METHOD = REML
/PRINT = SOLUTION TESTCOV
/RANDOM = INTERCEPT xij | SUBJECT(clusterID) COVTYPE(UN).
```

SAS PROC MIXED syntax:
```
proc mixed data = data covtest noclprint noitprint method = reml dfbw;
class clusterID;
model yij = xij / solution;
random intercept xij / subject = clusterID type = un;
run;
```

Conventional random-slope contextual effect MLM (Table 1 Row 2):

***lmer* syntax:**
```
fit_mod2 <- lmer(yij ~ 1 + xij + xij_clusmean + (1 + xij | clusterID),
data = data)
summary(fit_mod2)
```

SPSS MIXED syntax:
```
MIXED yij WITH xij xij_clusmean
/FIXED = INTERCEPT xij xij_clusmean
/METHOD = REML
/PRINT = SOLUTION TESTCOV
/RANDOM = INTERCEPT xij | SUBJECT(clusterID) COVTYPE(UN).
```

SAS PROC MIXED syntax:
```
proc mixed data = data covtest noclprint noitprint method = reml dfbw;
class clusterID;
model yij = xij xij_clusmean / solution;
random intercept xij / subject = clusterID type = un;
run;
```

Random-slope contextual effect MLM with random contextual effect (Table 1 Row 3):

***lmer* syntax:**
```
fit_mod3 <- lmer(yij ~ 1 + xij + xij_clusmean + (1 + xij + xij_clusmean
| clusterID), data = data)
summary(fit_mod3)
```

SPSS MIXED syntax:
```
MIXED yij WITH xij xij_clusmean
/FIXED = INTERCEPT xij xij_clusmean
/METHOD = REML
/PRINT = SOLUTION TESTCOV
/RANDOM = INTERCEPT xij xij_clusmean | SUBJECT(clusterID)
COVTYPE(UN).
```

SAS PROC MIXED syntax:
```
proc mixed data = data covtest noclprint noitprint method = reml dfbw;
class clusterID;
model yij = xij xij_clusmean / solution;
random intercept xij xij_clusmean / subject = clusterID type = un;
run;
```

Random-slope cluster-mean-centered MLM (Table 1 Row 4):

***lmer* syntax:**
```
fit_mod4 <- lmer(yij ~ 1 + xij_gmc + (1 + xij_gmc | clusterID),
data = data)
summary(fit_mod4)
```

SPSS MIXED syntax:
```
MIXED yij WITH xij_gmc
/FIXED = INTERCEPT xij_gmc
/METHOD = REML
/PRINT = SOLUTION TESTCOV
/RANDOM = INTERCEPT xij_gmc | SUBJECT(clusterID) COVTYPE(UN).
```

SAS PROC MIXED syntax:
```
proc mixed data = data covtest noclprint noitprint method = reml dfbw;
class clusterID;
model yij = xij_gmc / solution;
random intercept xij_gmc / subject = clusterID type = un;
run;
```

Random-slope cluster-mean-centered MLM with fixed between effect (Table 1 Row 5):

***lmer* syntax:**
```
fit_mod5 <- lmer(yij ~ 1 + xij_gmc + xij_clusmean + (1 + xij_gmc |
clusterID), data = data)
summary(fit_mod5)
```

SPSS MIXED syntax:
```
MIXED yij WITH xij_gmc xij_clusmean
/FIXED = INTERCEPT xij_gmc xij_clusmean
/METHOD = REML
/PRINT = SOLUTION TESTCOV
/RANDOM = INTERCEPT xij_gmc | SUBJECT(clusterID) COVTYPE(UN).
```

SAS PROC MIXED syntax:
```
proc mixed data = data covtest noclprint noitprint method = reml dfbw;
class clusterID;
model yij = xij_gmc xij_clusmean / solution;
random intercept xij_gmc / subject = clusterID type = un;
run;
```

Random-slope cluster-mean-centered MLM with random between effect (Table 1 Row 6):

***lmer* syntax:**
```
fit_mod6 <- lmer(yij ~ 1 + xij_gmc + xij_clusmean + (1 + xij_
gmc + xij_clusmean | clusterID), data = data)
summary(fit_mod6)
```

SPSS MIXED syntax:
```
MIXED yij WITH xij_gmc xij_clusmean
/FIXED = INTERCEPT xij_gmc xij_clusmean
/METHOD = REML
```

(*Continued*)

```
/PRINT = SOLUTION TESTCOV
/RANDOM = INTERCEPT xij_gmc xij_clusmean | SUBJECT(clusterID)
COVTYPE(UN).
```

SAS PROC MIXED syntax:

```
proc mixed data = data covtest noclprint noitprint method = reml dfbw;
class clusterID;
model yij = xij_gmc xij_clusmean / solution;
random intercept xij_gmc xij_clusmean / subject = clusterID type = un;
run;
```

Random-slope hybrid contextual effect & cluster-mean-centered MLM (Table 1 Row 7):

***lmer* syntax:**

```
fit_mod7 <- lmer(yij ~ 1 + xij + xij_clusmean + (1 + xij_gmc |
clusterID), data = data)
summary(fit_mod7)
```

(*Continued*)

SPSS MIXED syntax:

```
MIXED yij WITH xij xij_clusmean
/FIXED = INTERCEPT xij xij_clusmean
/METHOD = REML
/PRINT = SOLUTION TESTCOV
/RANDOM = INTERCEPT xij_gmc | SUBJECT(clusterID) COVTYPE(UN).
```

SAS PROC MIXED syntax:

```
proc mixed data = data covtest noclprint noitprint method = reml dfbw;
class clusterID;
model yij = xij xij_clusmean / solution;
random intercept xij_gmc / subject = clusterID type = un;
run;
```