Psychological Methods

Aberrant Distortion of Variance Components in Multilevel Models Under Conflation of Level-Specific Effects

Jason D. Rights

Online First Publication, October 6, 2022. http://dx.doi.org/10.1037/met0000514

CITATION

Rights, J. D. (2022, October 6). Aberrant Distortion of Variance Components in Multilevel Models Under Conflation of Level-Specific Effects. *Psychological Methods*. Advance online publication. http://dx.doi.org/10.1037/met0000514

AMERICAN PSYCHOLOGICAL ASSOCIATION

Psychological Methods

© 2022 American Psychological Association
ISSN: 1082-989X

https://doi.org/10.1037/met0000514

Aberrant Distortion of Variance Components in Multilevel Models Under Conflation of Level-Specific Effects

Jason D. Rights
Department of Psychology, University of British Columbia

Abstract

Methodologists have often acknowledged that, in multilevel contexts, level-1 variables may have distinct within-cluster and between-cluster effects. However, a prevailing notion in the literature is that separately estimating these effects is primarily important when there is specific interest in doing so. Consequently, in practice, researchers uninterested in disaggregating these effects (or unaware of their difference) routinely fit models that conflate them. Furthermore, even researchers who properly disaggregate the fixed components in a model (avoid *fixed* conflation) may still inadvertently and unknowingly conflate the random effects (fail to avoid *random* conflation). The purpose of this article is to elucidate an unappreciated consequence of such fixed or random conflation, namely, that it can cause systematic distortion in all variance components, yielding uninterpretable variances that adversely affect the entire model. In this article, I provide novel mathematical derivations, simulations, and pedagogical illustrations of such variance distortion, showing how it leads to several aberrant consequences: (1) error variances at level-1 and level-2 can systematically increase (in the population) with the addition of predictors; (2) there can be a large apparent degree of between-cluster random-effect variability in cases in which there is actually no between-cluster outcome variability; (3) R-squared measures of explained variance can be severely biased, uninterpretable, and well below the logical bound of 0; and (4) inference for all fixed components of the model—not just the conflated slopes themselves—can be compromised. I conclude with recommendations for practice, including cautionary notes on interpreting results from prior research that had specified conflated slopes.

Translational Abstract

Many analyses in psychology and other fields involve multilevel structured data, such as students nested within classrooms or repeated observations nested within individuals. In these contexts, it has been established that observation-level predictor variables may have distinct effects at the within-cluster level (e.g., within classrooms) versus the between-cluster level (e.g., between classrooms). However, in practice, researchers uninterested in disaggregating these effects (or unaware of their difference) routinely fit models that conflate them, meaning they obtain a single estimate for the slope of the observation-level variable that is, implicitly, an uninterpretable weighted average of the level-specific effects. The purpose of this article is to elucidate an unappreciated consequence of such conflation, namely, that it can cause systematic distortion in the variance components, yielding variances that are largely uninterpretable and that adversely affect the entire model. In this article, I explain this concept, and illustrate several of its key implications with pedagogical examples. For instance, I show how conflation leads to: (1) error variances at both level-1 and level-2 systematically increasing with the addition of predictors; (2) a large apparent degree of between-cluster variability in cases in which there is actually no between-cluster variability; (3) severely biased and uninterpretable R-squared measures of explained variance, which can also be well below the logical bound of 0; and (4) compromised inference for all of the slopes included in the model. I conclude with recommendations for practice, including a cautionary note on interpreting results from prior research that had specified conflated slopes.

Keywords: multilevel modeling, linear mixed effects modeling, centering, variance components, intraclass correlation

Supplemental materials: https://doi.org/10.1037/met0000514.supp

For the past several decades, multilevel modeling (MLM; also known as hierarchical linear modeling or linear mixed effects modeling) has been the most popular approach to accommodating nested data structures in the psychological sciences, and has been widely used in many other fields, such as education, biology, and organizational research (Goldstein, 2010; Raudenbush & Bryk,

Jason D. Rights https://orcid.org/0000-0003-3784-9772

I thank Sonya Sterba, Kristopher Preacher, Sun-Joo Cho, Andrew Tomarken, Yingchi Guo, and Jeneesha Dhaliwal for helpful comments.

Correspondence concerning this article should be addressed to Jason D. Rights, Department of Psychology, University of British Columbia, 2136 West Mall, Vancouver, BC V6T1Z4, Canada. Email: jrights@psych.ubc.ca

2002; Snijders & Bosker, 2012). As a prototypical example, an analysis might involve collecting data from students across multiple classrooms. The dataset would then have a two-level structure, in which individual students are at level-1 (i.e., the observation level) and classrooms are at level-2 (i.e., the cluster level). MLM provides an intuitive framework by which researchers can accommodate the dependency of observations within the same cluster (e.g., students within the same class) and simultaneously examine potential effects of level-1 variables (e.g., individual student characteristics) and level-2 variables (e.g., classroom-level characteristics).

Intrinsic to such hierarchical data structures is the possibility that level-1 independent variables can exert distinct level-specific effects, that is, distinct *within-cluster* and *between-cluster* effects. For example, in the United States, data often suggest there to be a positive *between-state* association of income and political conservatism, but a negative *within-state* association—that is, states with higher average income tend to be less politically conservative, whereas people with higher income relative to their state's average tend to be more politically conservative (e.g., Gelman et al., 2007). As another example, suppose a study involves repeated measures collected from individuals, and hence persons are the clustering unit. If looking at the relationship between a person's time spent on a cognitive task and their accuracy, one might expect a positive *between-person* association (because people who are better at the task overall are likely both faster and more accurate than those who are worse at the task) but a negative *within-person* association (because people tend to do worse when they rush, i.e., spend less time on the task than their average time spent; e.g., Murayama et al., 2017). In a third, classic example, consider a dataset of students nested within classrooms with math achievement as the predictor and math self-efficacy the outcome. In such cases, it is often found that the *within-class* relationship is more positive than the *between-class* relationship. Specifically, a student's achievement relative to their classroom is highly positively predictive of self-concept, but when looking across classrooms, holding constant a student's absolute level of achievement, being in a high-achieving classroom (and thus having high-achieving peers to which to compare oneself) is negatively associated with self-concept (this phenomenon is often termed the "big-fish-little-pond effect;" e.g., Marsh et al., 2008).[1]

Failing to appropriately disaggregate such level-specific effects can yield a slope estimate that is an "uninterpretable blend" of the two (Cronbach, 1976; Raudenbush & Bryk, 2002). For example, if one estimated a single slope to describe one of the aforementioned relationships (income and political conservatism, reaction time and accuracy, or math achievement and math self-concept), this estimate would be of questionable utility and interpretability, and would implicitly reflect some weighted average of the two distinct and highly disparate effects. Such a slope is often said to be *conflated* (Preacher et al., 2010; Rights et al., 2020).

Nonetheless, current methodological recommendations are not absolute. On the one hand, numerous sources do recommend disaggregating level-specific effects to ensure appropriate level-specific inferences (e.g., Asparouhov & Muthén, 2019; Brauer & Curtin, 2018; Curran & Bauer, 2011; Enders & Tofighi, 2007; Hamaker & Muthén, 2020; Hoffman, 2019; Raudenbush & Bryk, 2002; Wang & Maxwell, 2015), and this advice is commonly followed in practice. On the other hand, however, even among methodologists who recognize the merit of such disaggregation and largely recommend doing so, it is also commonly suggested that its importance depends on the research context. For instance, it is suggested that disaggregation may not be necessary if one has no theoretical reason to believe within-cluster and between-cluster effects differ, has no substantive interest in looking at these effects separately, or considers the level-1 predictors to be control variables (see, for example, Dalal & Zickar, 2012; Enders, 2013; Enders & Tofighi, 2007; Hamaker & Grasman, 2014; Hofmann & Gavin, 1998; Hox et al., 2018; McCoach, 2010; Paccagnella, 2006; Peugh, 2010; Raudenbush & Bryk, 2002; Snijders & Bosker, 2012). At a more extreme level, a recent methodological article has even stated that disaggregating level-specific effects (in particular, via cluster-mean centering) is a "dangerous practice" that "should be abandoned" (Kelley et al., 2017; this argument, however, was later rebutted by Bell et al., 2018;). Consequently, conflated slopes are routinely specified in published multilevel applications, and have been noted to be the most common way level-1 predictors are incorporated into models in various research contexts (see, e.g., Asparouhov & Muthén, 2019, p. 129; Hoffman, 2015, p. 344; Hox et al., 2018, p. 48; Preacher & Sterba, 2019, p. 253). Adding further concern is the fact that, though nearly all sources discussing conflation focus exclusively on the *fixed* portion of the model (i.e., across-cluster average slopes), there is also the possibility to conflate the *random* portion of the model (i.e., the random effects representing cluster-specific deviations from the across-cluster average slopes), and commonly, even when the fixed portion is disaggregated, the random portion will be conflated (Rights & Sterba, 2020).

The purpose of this article is to elucidate and demonstrate an unappreciated consequence of conflating level-specific effects, namely, that this practice can cause systematic distortion in variance components. Importantly, this distortion can compromise model inferences and interpretation, even when one has no substantive interest in separately considering within-cluster versus between-cluster effects. In this article, I provide descriptions and pedagogical illustrations of how conflation leads to distortion of variances, and provide novel mathematical derivations and formulas that show the degree of distortion to be a direct function of the discrepancy in the level-specific effects. I further explain and illustrate how such distortion can lead to several aberrant and unintuitive issues:

1. Error variances at both level-1 and level-2 can systematically increase (in the population) with the addition of predictors;

2. There can be a large apparent degree of between-cluster random-effect variability in cases in which there is actually no between-cluster outcome variability;

[1] This slope of math self-concept on class-mean math achievement, conditioning on *absolute* (not class-mean-centered) math achievement, is an example of a *contextual effect* (Raudenbush & Bryk, 2002; Snijders & Bosker, 2012). A contextual effect is mathematically equal to the between-cluster effect minus the within-cluster effect of the level-1 variable. Within-cluster effects and contextual effects are directly estimated in so-called *contextual effect models*, whereas within-cluster effects and between-cluster effects are directly estimated in *cluster-mean-centered models* (both types of models will be discussed in detail in the current article).

3. R-squared measures of explained variance, which are often used to quantify effect size, can be severely biased, uninterpretable, and well below the logical lower bound of 0;

4. Testing and inference for all fixed components of the model—not just the conflated slopes themselves—can be compromised, particularly in terms of increased estimation variance and reduced power.

I explain and demonstrate how these issues are avoided in unconflated models.

From an immediately practical standpoint, this article aims to encourage applied researchers to disaggregate level-specific effects of level-1 variables in practice. However, even for researchers who are already accustomed to such disaggregation, it is my hope that this article will have didactic utility in explaining seemingly aberrant behavior that can be encountered when fitting multilevel models, or when interpreting results from published research. That is, this article seeks to clarify that researchers should be cautious in interpreting conflated model results not only in terms of the estimated conflated effects themselves, but also the estimated random effect variances and any metrics that utilize these estimates (e.g., intraclass correlation coefficients and R-squared measures), as well as inferences based on other fixed component estimates in the model.

The remainder of the article proceeds as follows. I first, at a population level, describe and illustrate within-cluster versus between-cluster error terms. I then compare these errors from a model that disaggregates level-specific fixed effects to one that fails to do so, and use this to provide a mathematical derivation showing how the error variances from the conflated model can be distorted. I then give specific analytic explanations of the four aforementioned issues induced by conflating. For each issue, I provide an illustrative example (R scripts and simulated data sets are provided in the online supplemental materials), as well as simulation results that demonstrate the issue across repeated samples (the examples and simulation results can be read independently—the former will be most useful for readers seeking a pedagogical tutorial, whereas the latter will be most useful for readers interested in the extent of distortion to expect under specific generating conditions). I will initially consider random intercept models with fixed slopes; hence, the first part of this article will focus on the impact of *fixed conflation*, defined as imposing the constraint that the *fixed* components associated with the within-cluster and between-cluster portions of level-1 variables are exactly equal. I will then focus on random slope models, discuss how *random conflation*—defined as imposing the constraint that the *random* components of the within-cluster and between-cluster portions of level-1 variables are exactly equal—can cause further distortion, and investigate its impact via simulation. In the Discussion, I provide recommendations, address certain modeling complexities not explicitly included throughout the article, and explain how the results generalize to these situations.

Comparing Level-Specific Error Terms in Conflated Versus Unconflated Models

Inherent to multilevel modeling is the inclusion of not only *observation-level* error terms (which are also included in standard single-level regression), but also *cluster-level* error terms, both defined shortly. The impact that conflation has on variance component distortion can be understood by comparing these two error terms from an unconflated model versus those from a conflated model. I will first consider these models purely from a population standpoint—hence, an important distinction will be made between model *errors* (reflecting deviations between actual scores and expected scores based on population parameters) and model *residuals* (reflecting sample realizations or estimates of the population errors).

Unconflated-x Model

The first multilevel model I consider is what I term the *unconflated-x model*, which can be used to separately estimate the within-cluster and between-cluster slopes associated with a level-1 independent variable:

$$\begin{gathered} y_{ij} = \gamma_{00} + \gamma_b x_{\bullet j} + \gamma_w (x_{ij} - x_{\bullet j}) + u_{0j} + e_{ij} \\ e_{ij} \sim N(0, \sigma^2);\ u_{0j} \sim N(0, \tau_{00}) \end{gathered} \tag{1}$$

Here, *i* denotes level-1 unit and *j* level-2 unit. Hence, y_{ij} denotes the outcome for observation *i* nested within cluster *j*. The level-1 independent variable is given as x_{ij} and predicts *y* via its cluster mean, $x_{\bullet j}$ (entered as a level-2 predictor), as well as its cluster-mean-centered (also commonly called group-mean-centered) version, $x_{ij} - x_{\bullet j}$ (entered as a level-1 predictor). The slope of $x_{\bullet j}$,[2] γ_b, denotes the *between-cluster effect*, whereas the slope of $x_{ij} - x_{\bullet j}$, γ_w, denotes the *within-cluster effect*.[3] The other fixed component, γ_{00}, is the across-cluster average intercept. The level-1 error term, e_{ij}, represents the *within-cluster* deviations of the observed scores of *y* from the *cluster-specific* (or conditional) expected value of *y*, $\gamma_{00} + \gamma_b x_{\bullet j} + u_{0j} + \gamma_w (x_{ij} - x_{\bullet j})$. Hence, the within-cluster error variance is $\text{var}(e_{ij}) = \sigma^2$. The level-2 error term, u_{0j}, represents the *between-cluster* deviations of the cluster-mean scores of *y* from the *across-cluster* (or marginal) expected value of the cluster-mean scores of *y*, given as $\gamma_{00} + \gamma_b x_{\bullet j}$. Hence, the between-cluster error variance is $\text{var}(u_{0j}) = \tau_{00}$.[4]

In this *unconflated-x model*, there is a clear separation of the levels of analysis in that the level-1 predictor, $x_{ij} - x_{\bullet j}$, varies exclusively within-cluster and thus, logically, can only explain within-cluster/level-1 outcome variance (i.e., variance in *y* scores within clusters), whereas the level-2 predictor, $x_{\bullet j}$, varies exclusively between-cluster and thus can only explain between-cluster/level-2 outcome variance (i.e., variance in cluster-mean *y* scores across clusters). Therefore, conceptually, the overall level-1 outcome variance can be cleanly

[2] There must be some between-cluster variance in x_{ij} (i.e., variance in $x_{\bullet j}$) for this model to be identified; otherwise, there would only be a within-cluster effect of x_{ij}, γ_w (and in such a case, there is no risk of conflation).

[3] In an effort to provide concise communication and to maintain consistency with prior literature, I use the word "effect" here, but strictly speaking, these slopes can reflect associations that are not causal.

[4] An equivalent random-intercept model would involve entering both *x* (in its raw form, or centered by any constant) and the cluster-mean of *x* as predictors. This is termed a *contextual effect model* (Kreft et al., 1995; Raudenbush & Bryk, 2002). For simplicity and without loss of generality, I will here focus on the cluster-mean-centered version of this model.

broken down into that which is accounted for (or "explained") by $x_{ij} - x_{\bullet j}$ (i.e., $\text{var}(\gamma_w(x_{ij} - x_{\bullet j})) = \gamma_w^2 \text{var}(x_{ij} - x_{\bullet j})$) and that which is "unexplained" and instead accounted for by the level-1 error term (i.e., $\text{var}(e_{ij}) = \sigma^2$). Similarly, the overall level-2 outcome variance can be broken down into that which is explained by $x_{\bullet j}$ (i.e., $\text{var}(\gamma_b x_{\bullet j}) = \gamma_b^2 \text{var}(x_{\bullet j})$) and that which is instead accounted for by the level-2 error term (i.e., $\text{var}(u_{0j}) = \tau_{00}$).

As an illustration/visualization of level-specific effects and error terms, Figure 1 shows a hypothetical subset of data from an *unconflated-x model* and the corresponding regression lines that depict the within-cluster and between-cluster effects. Figure 1, Panel A (first considering only the solid lines in the plot) demonstrates the *positive within-cluster effect*, whereas Figure 1, Panel B (the solid line), in contrast, demonstrates the *negative between-cluster effect* (this situation is consistent, for instance, with the earlier-described example of reaction time predicting accuracy). Figure 1 also provides an illustration of the two error terms in the *unconflated-x model*. An example level-1 error, e_{ij}, is highlighted in Panel A, showing the difference between a single observation's actual outcome and its cluster-specific expected outcome. An example level-2 error, u_{0j}, is highlighted in Panel B, showing the difference between a cluster's actual mean outcome and their model-predicted mean outcome. Taken together, for each panel, the variance in the expected scores (defined by the regression lines) is the level-specific variance explained by the predictor, and the variance of the error terms is the remaining level-specific variance (i.e., that which is not explained by the predictor).

Conflated-x Model

What I term here the *conflated-x model* is nested within the *unconflated-x model*, imposing the constraint that the level-specific effects of x are equivalent by setting $\gamma_b = \gamma_w$:

$$\begin{gathered} y_{ij} = \gamma_{00}^* + u_{0j}^* + \gamma_c x_{\bullet j} + \gamma_c(x_{ij} - x_{\bullet j}) + e_{ij}^* \\ e_{ij}^* \sim N(0, \sigma^{2*});\ u_{0j}^* \sim N(0, \tau_{00}^*) \end{gathered} \tag{2}$$

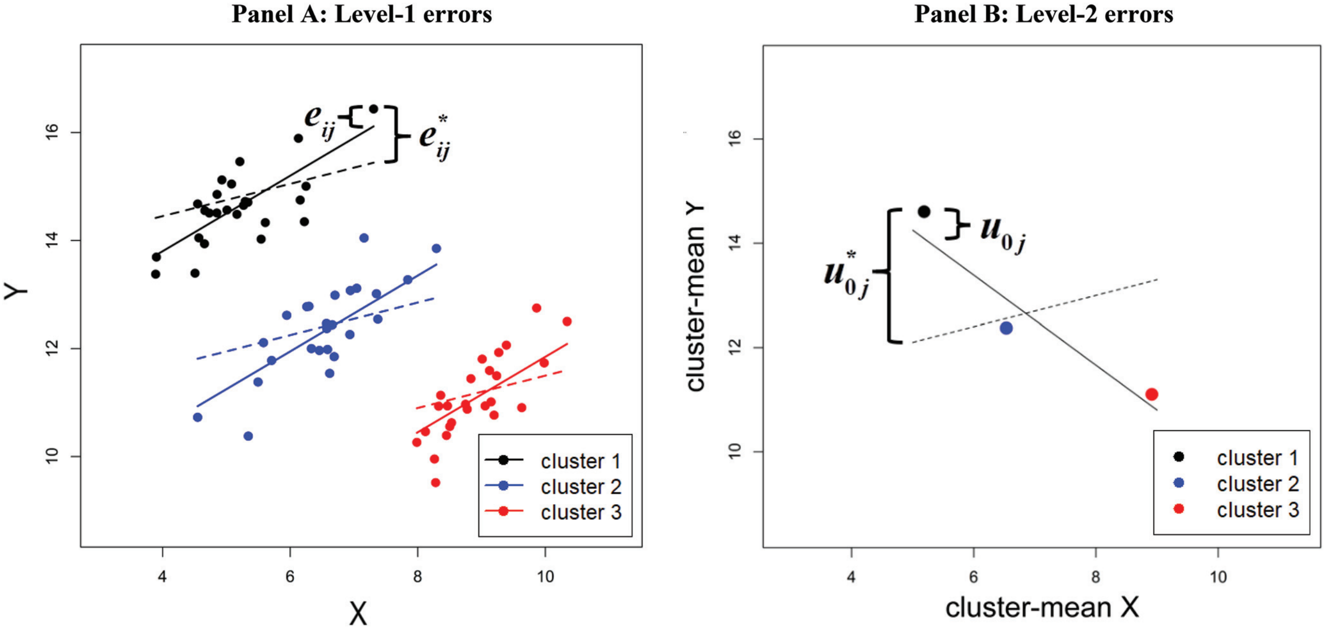
Here γ_c is called the *conflated effect*, in that it simultaneously represents the within-cluster effect (i.e., slope of $x_{ij} - x_{\bullet j}$) and the between-cluster effect (i.e., slope of $x_{\bullet j}$). For the other model terms, I use asterisks to distinguish them from the unconflated model in Equation 1. A mathematically equivalent way of writing this conflated model is to leave x in its raw form, as follows:

$$y_{ij} = \gamma_{00}^* + u_{0j}^* + \gamma_c x_{ij} + e_{ij}^* \tag{3}$$

Hence, entering x in its raw form (or centering it by any constant value, such as the grand mean), as is typical of current practice, creates the implicit assumption that level-specific effects are equal.

Figure 1

Heuristic Illustration of Level-1 Versus Level-2 Errors From Unconflated Versus Conflated Models



Note. The solid lines in Panel A reflect the within-cluster effect from the *unconflated-x model*, whereas the solid line in Panel B reflects the between-cluster effect from the *unconflated-x model*. The dashed line (in both panels) represents the conflated effect (an ambiguous blend of the within-cluster and between-cluster effects) from the *conflated-x model*. Hence, example level-1 and level-2 errors from the *unconflated-x model* are given by e_{ij} and u_{0j}, respectively, and example level-1 and level-2 errors from the *conflated-x model* are given by e_{ij}^* and u_{0j}^*, respectively. Note that, without loss of generality, data are presented on the raw metric (rather than being centered) for illustrative visualization purposes. See the online article for the color version of this figure.

Here, there is no longer a clear separation of the levels of analysis—the uncentered predictor x varies both within- and between-cluster, and its single conflated effect is used to simultaneously explain both the within- and between-cluster outcome variance.

One result that is well-established in the current literature is that the conflated slope, γ_c, is some weighted average of the two level-specific effects, γ_w and γ_b. Hence, generically:

$$\gamma_c = \lambda\gamma_b + (1 - \lambda)\gamma_w \tag{4}$$

where λ is a number from 0 to 1 and represents the degree to which the conflated effect is weighted toward the between-cluster effect. There is not an established closed-form population expression for λ general to any MLM (for specific sample-level case examples, see, e.g., Raudenbush & Bryk, 2002; Scott & Holt, 1982), but it is known that having more precision in estimating γ_w relative to γ_b will lead to a small λ, whereas having more precision in estimating γ_b relative to γ_w will lead to a large λ (Raudenbush & Bryk, 2002). Because, by definition, there are always more level-1 units than level-2 units in hierarchical data sets, λ will typically be closer to 0 than 1, and hence, the conflated effect will tend to be closer to γ_w than γ_b.

An illustration of this conflated slope is also found in Figure 1 and is given by the dashed regression lines. Hence, in Panel A, the dashed regression lines are equal to $\gamma^*_{00} + u^*_{0j} + \gamma_c x_{ij}$, and in Panel B, the dashed regression line is $\gamma^*_{00} + \gamma_c x_{\bullet j}$. As is apparent in this plot, the conflated effect is somewhere in the middle of the within-cluster and between-cluster effects (per Equation 4). This conflated effect thus provides a distorted view of the association between x and y by forcing the (positive) within-cluster effect and the (negative) between-cluster effect to be equivalent.

The novel result of this article, however, pertains to the relationship between the errors of the two models, e_{ij} and u_{0j} versus e^*_{ij} and u^*_{0j}. Figure 1, Panel A provides a visualization of the conflated level-1 error, e^*_{ij}. As can be gleaned from the plot, e^*_{ij} is equal to the unconflated level-1 error, e_{ij}, plus the difference in the cluster-specific expectation from the *unconflated-x model* versus the *conflated-x model*. A proof of this is provided in Appendix A, yielding the following expression for the conflated level-1 error:

$$e^*_{ij} = e_{ij} + (\gamma_w - \gamma_c)(x_{ij} - x_{\bullet j}) \tag{5}$$

Similarly, as illustrated in Figure 1, Panel B, the conflated level-2 error, u^*_{0j}, is equal to the unconflated level-2 error, u_{0j}, plus the difference in the expected mean score from the *unconflated-x model* and the *conflated-x model*. Hence, as shown in Appendix A, the conflated level-2 error is:

$$u^*_{0j} = u_{0j} + \gamma_{00} - \gamma^*_{00} + (\gamma_b - \gamma_c)x_{\bullet j} \tag{6}$$

This thus yields the following level-1 and level-2 error variances for the *conflated-x model*:

$$\sigma^{2*} = \sigma^2 + (\gamma_w - \gamma_c)^2 \text{var}(x_{ij} - x_{\bullet j}) \tag{7}$$

$$\tau^*_{00} = \tau_{00} + (\gamma_b - \gamma_c)^2 \text{var}(x_{\bullet j}) \tag{8}$$

As shown in Equations 7 and 8, the error variances of the *conflated-x model* will be larger than those of the *unconflated-x model*—more troubling though is the fact that, unlike the unconflated variances, when the underlying level-specific effects differ, the conflated variances *do not actually represent the portion of level-specific outcome variance that is not explained by the predictor*. The reason this standard interpretation no longer holds is that, in the population, the actual level-specific outcome variance is no longer equal to the simple sum of variance explained by the level-specific portion of x and the variance accounted for by the error term. In fact, as will be shown later, often times the conflated error variance alone will be *much greater* than the *overall* level-specific variance (this can be seen, for instance, in Figure 1, Panel B, in which the across-cluster variance of the u^*_{0j} s is greater than the across-cluster variance of the cluster means of y). Instead, these conflated variances in Equations 7 and 8 reflect not only the "true" unexplained variance (i.e., variances of e_{ij} and u_{0j}) but also the degree to which level-specific effects of x differ. Hence, as is often described of the conflated slope, the variances from a conflated model are a sort of "uninterpretable blend" themselves. Secondarily, taking together Equations 7 and 8 as well as Equation 4 clarifies that there is a trade-off in that, holding all else constant, *less* distortion in the level-1 error variance (i.e., more similarity in γ_w and γ_c) is associated with *more* distortion in the level-2 variance (i.e., more discrepancy in γ_b and γ_c), and vice versa (as demonstrated later).

A potential critique of this analytic result is that the error variances from the *conflated-x model* and the *unconflated-x model* are, in some sense, representing two different things, and as such, perhaps it is not too troubling that they differ both in value and in interpretation. I would argue, however, that when level-specific effects truly differ, the *conflated-x model* is definitively misspecified, and the variances from this model have no useful interpretation, and can negatively impact the model as a whole. To underscore this position, I now present several unappreciated issues that result from these conflated variances, and illustrate how fitting conflated models can lead to aberrant and seemingly paradoxical results.

Issue 1: Conflating Level-Specific Effects Can Cause Both Level-1 and Level-2 Error Variances to Increase in the Population When Adding Predictors

In both single-level regression and MLM, researchers are accustomed to the aforementioned concept that the error variance is, in theory, that which is "unexplained" by the predictors included in the model. Hence, conceptually, when adding a predictor to a model whose slope is nonzero, the error variance should decrease, as one is adding—not removing—information that can be used in explaining variability in the outcome. In fact, in single-level regression via ordinary least squares (OLS), the population error variance mathematically cannot increase when adding a predictor to the model, nor can the sample sum of squared residuals.[5] As explained and illustrated in this section, however, both the level-1

[5] Using OLS in a sample, however, it is possible for a slight decrease in the estimated *residual* variance after adding a predictor if the decrease in sum of squared residuals is small and not outweighed by the decrease in degrees of freedom (Hoffman, 2015; Rencher & Schaalje, 2007).

and level-2 error variance in MLM *can* increase when adding a predictor, in a way that can be mathematically explained by the impact of conflating level-specific effects.

In terms of what has been established in prior literature, the first is that level-2 *residual* variances (i.e., in a given sample, rather than the population) in MLM can increase when adding a level-1 predictor, in a way that is simply due to how maximum likelihood estimation obtains an estimate for this variance (Hoffman, 2015; Lahuis et al., 2014; Snijders & Bosker, 2012). Specifically, it is known that, mathematically, the observed between-cluster outcome variation will reflect not only true (population) differences across clusters, but also sampling variability resulting from within-cluster variation. To formalize this, consider the *random-intercept-only model* in which there are no predictors and a total of three parameters—the fixed component of the intercept, the level-1 variance (σ^2_{null}), and the level-2 variance ($\tau_{00,null}$). With equal cluster sizes (n_j), the between-cluster outcome variance is then given as:

$$\text{var}(y_{\bullet j}) = \tau_{00,null} + \frac{\sigma^2_{null}}{n_j} \tag{9}$$

Which thus implies

$$\tau_{00,null} = \text{var}(y_{\bullet j}) - \frac{\sigma^2_{null}}{n_j} \tag{10}$$

(see also Snijders & Bosker, 2012, Chapter 3 for expressions not assuming equal cluster sizes). Hence, obtaining estimates for the level-2 variance via maximum likelihood involves subtracting from the observed between-cluster outcome variance the estimated within-cluster outcome variance divided by cluster size, which in certain cases can lead to expected increases in the level-2 residual variance after adding predictors. For instance, when adding a level-1 predictor that exclusively varies within-cluster, this can explain within-cluster outcome variance (leading to a decrease in the level-1 residual variance), but will not explain any between-cluster outcome variance, and hence, the estimate of the level-2 variance will increase. Though this type of increase in variance is useful to understand, the increase resulting from this relationship goes away as cluster size increases, and is distinct from that which is the focus of the current article, that is, the impact on the error variance in the population.

Second, some authors have additionally acknowledged that the level-2 error variances can increase (in the population) when adding a level-1 predictor with a conflated slope, providing examples with corresponding explanations for this phenomenon (e.g., Gelman & Hill, 2007, p. 280; Hoffman, 2015, p. 407). The novelty of the current article is that I (a) clarify specifically the degree to which this increase in level-2 variance occurs as a direct mathematical function of the underlying level-specific effects, and (b) clarify that not only the level-2 error variance, but the level-1 error variance can systematically increase when adding predictors.

Analytic Explanation of Level-2 Variance Increase

I will first consider the behavior of the level-2 error variance increasing when adding a level-1 predictor with a conflated slope, specifically in comparing the level-2 variance from the random-intercept-only model (i.e., the null model with no predictors) to that containing a level-1 predictor (i.e., the *conflated-x model* in Equation 3). The null model level-2 variance, $\tau_{00,null}$, represents the overall between-cluster outcome variance, and assuming data are generated from the *unconflated-x model*, $\tau_{00,null}$ is (Rights & Sterba, 2019):

$$\tau_{00,null} = \tau_{00} + \gamma_b^2 \text{var}(x_{\bullet j}) \tag{11}$$

When adding the level-1 predictor x to form the *conflated-x model*, the change in the level-2 variance is then given as the difference between Equation 8 and 11:

$$\tau^*_{00} - \tau_{00,null} = ((\gamma_b - \gamma_c)^2 - \gamma_b^2)\text{var}(x_{\bullet j}) \tag{12}$$

As shown in Equation 12, the level-2 variance will increase going from the null model to the *conflated-x model* whenever the squared difference in the between-cluster and the conflated effect (i.e., $(\gamma_b - \gamma_c)^2$) is greater than the squared between-cluster effect (i.e., γ_b^2), as this implies Equation 12 is positive. In contrast, when going from the null model to the *unconflated-x model*, the change in level-2 error variance would never be positive (as τ_{00} minus Equation 11 is at most 0).

Analytic Explanation of Level-1 Variance Increase

The same concept holds for the level-1 variance—if Equation 1 is the population model, the level-1 variance from the random-intercept-only model (reflecting the overall within-cluster outcome variance) is given as (Rights & Sterba, 2019):

$$\sigma^2_{null} = \sigma^2 + \gamma_w^2 \text{var}(x_{ij} - x_{\bullet j}) \tag{13}$$

When adding the level-1 predictor x to form the *conflated-x model*, the change in the level-1 variance is then given as the difference between Equation 7 and 13:

$$\sigma^{2*} - \sigma^2_{null} = ((\gamma_w - \gamma_c)^2 - \gamma_w^2)\text{var}(x_{ij} - x_{\bullet j})) \tag{14}$$

Equation 14 shows that the level-1 error variance will increase whenever $(\gamma_w - \gamma_c)^2$ is greater than γ_w^2. In contrast, for the *unconflated-x model*, the change in level-1 error variance would never be positive (as σ^2 minus Equation 13 is at most 0).

Illustration #1: Level-2 Variance Increasing

As a concrete illustration of the level-2 variance increasing, consider an analysis of students nested within schools (colleges) in which the outcome of interest is *salary* (i.e., starting salary for first job after college, in thousands of dollars) and the predictor is grade point average (GPA). Table 1 shows results from the random-intercept-only model as well as the *unconflated-x model* and the *conflated-x model* (the dataset, which was simulated for pedagogical purposes, and associated R scripts are available in the online supplemental materials). As with all forthcoming illustrations,

Table 1

Illustration #1: Level-2 Variance Increases When Adding Predictor for Conflated Model (Predicting Starting Salary From College GPA)

Parameter	Null model estimates	Unconflated model estimates	Conflated model estimates
Intercept	49.985 (*SE* = 0.194; $p < .001$)	49.988 (*SE* = 0.186; $p < .001$)	49.980 (*SE* = 0.231; $p < .001$)
Within-school slope of *GPA*	—	3.813 (*SE* = 0.056; $p < .001$)	—
Between-school slope of *GPA*	—	−2.605 (*SE* = 0.710; $p < .001$)	—
Conflated slope of *GPA*	—	—	3.787 (*SE* = 0.056; $p < .001$)
Level-1 error variance	19.890	12.119	12.120
Level-2 error variance	5.247	4.966	7.769

Note. See *salary_exdat.txt* for dataset and *illustrativeexamples.R* for R script. Raw *GPA* centered to have a mean of 0. *p*-values based on *t*-statistics with Satterthwaite degrees of freedom approximation.

models were estimated via restricted maximum likelihood with the *lmer* function in R (Bates et al., 2004). The *unconflated-x model* reveals an estimated *positive within-school* effect (3.813), reflecting the fact that, within a given school, students who have higher GPAs tend to have higher starting salaries. However, there is an estimated *negative between-school* effect (−2.605), which could be explained by the fact that schools with low average GPAs tend to have a greater number of students in programs that give low grades but are associated with high starting salaries (e.g., engineering), and vice versa (see, e.g., Gottard et al., 2007). Accordingly, both the level-1 and level-2 residual variances decrease after adding the two level-specific components of GPA (5.247 to 4.966 and 19.890 to 12.119, respectively), as the school-mean-centered component explains part of the overall within-school variation in salary and its school-mean component explains part of the overall between-school variation in salary. However, the conflated model slope estimate is, as expected, in the middle of the estimated level-specific effects ($-2.605 < 3.787 < 3.813$), and this leads to a large increase in the level-2 residual variance (5.247 to 7.769), thus making it impossible to interpret this conflated variance as the portion of the overall between-school variance that is unexplained. This variance instead reflects some combination of the underlying unexplained between-school variance that would be observed from an unconflated model, as well as the degree of discrepancy in the underlying level-specific effects (see Equation 8).

Illustration #2: Level-1 Variance Increasing

As a related illustration for the *level-1* variance, here I consider a longitudinal dataset consisting of repeated observations nested within persons (i.e., persons serving as clusters), with the outcome of *symptoms* (a measure of physical symptoms including pain, cardiovascular, and others), predicted from both time (quantified by person-mean-centered *session* number) and *mood* (a measure of the degree of one's negative mood at a given session; see Hoffman, 2015 for further description of these data).[6] Table 2 provides results from a baseline unconditional model that includes only *session*, an unconflated model with person-mean-centered and person-mean *mood*, and a conflated model with only grand-mean-centered *mood*. Results from the unconflated model show that there is little apparent within-person effect (as the slope of person-mean-centered *mood* is small and nonsignificant; 0.154) but there is a significant and positive estimate for the between-person effect (2.013), suggesting that people who tend to have higher levels of negative mood on average tend to also have worse symptoms on average. Accordingly, the level-1 variance decreases only a little bit from the baseline model (0.617 to 0.616), whereas the level-2 variance decreases more substantially (1.202 to 0.928). The conflated model, however, yields a conflated slope estimate for *mood* (0.321), which, in turn, yields an increase in the level-1 variance (0.617 to 0.618). Interestingly, the estimated conflated slope of *mood* is over twice the magnitude of the estimated within-cluster slope, but the conflated level-1 variance is larger than the unconflated. This is because the conflated level-1 variance is not simply the portion of within-person variation that is unexplained, but rather is some combination of the unexplained within-person variance and the degree of discrepancy in the underlying level-specific effects (see Equation 7).

Issue 2: Conflating Level-Specific Effects Can Yield Apparent Between-Cluster Random-Effect Variability in Cases in Which There Is No Between-Cluster Outcome Variability

In practice, researchers often seek to assess the degree of outcome variability at the within-cluster level versus the between-cluster level. For instance, if one finds there to be substantial between-cluster variability, this can be used as an indication that there is a meaningful amount of variance that could be explained by adding level-2 predictors (Hoffman, 2015; Raudenbush & Bryk, 2002; Rights & Sterba, 2019). Additionally, finding a large degree of between-cluster variability is often used, in part, as a justification for utilizing multilevel models in the first place, or to highlight the importance of considering individual differences (Lai & Kwok, 2015; Peugh, 2010; Snijders & Bosker, 2012).

Such quantification of between-cluster variance, however, is compromised when one is conflating level-specific effects. Indeed, one important implication of the analytic results here is that it is possible to observe a large degree of across-cluster differences even in cases where there are truly *no* underlying between-cluster differences. To illustrate, consider the *residual intraclass correlation coefficient* (residual *ICC*, sometimes termed the conditional *ICC*), which is defined as the ratio of the level-2 (random intercept) variance to the sum of the level-1 and level-2 variances, that is:

[6] Data for this specific example are available at https://www.pilesofvariance.com/index.html.

Table 2

Illustration #2: Level-1 Variance Increases When Adding Predictor for Conflated Model (Predicting Level of Symptoms From Mood)

Parameter	Unconditional model estimates	Unconflated model estimates	Conflated model estimates
Intercept	1.294 (*SE* = 0.113; $p < .001$)	1.288 (*SE* = 0.100; $p < .001$)	1.293 (*SE* = 0.109; $p < .001$)
Slope of *session*	−0.021 (*SE* = 0.025; *p* = .398)	−0.021 (*SE* = 0.025; *p* = .408)	−0.020 (*SE* = 0.025; *p* = .419)
Within-person slope of *mood*	—	0.154 (*SE* = 0.128; *p* = .231)	—
Between-person slope of *mood*	—	2.013 (*SE* = 0.379; $p < .001$)	—
Conflated slope of *mood*	—	—	0.321 (*SE* = 0.122; *p* = .009)
Level-1 error variance	0.617	0.616	0.618
Level-2 error variance	1.202	0.928	1.123

Note. See *illustrativeexamples.R* for R script. Raw *mood* centered to have a mean of 0. *p*-values based on *t*-statistics with Satterthwaite degrees of freedom approximation.

$$ICC_r = \frac{\tau_{00}}{\tau_{00} + \sigma^2} \tag{15}$$

This is a commonly used index in practice when communicating the relative amount of between-cluster variance versus within-cluster variance, controlling for the included covariates (e.g., Kuo et al., 2013; Krull & MacKinnon, 1999; Nakagawa et al., 2017).[7]

Analytic Explanation of Distortion in Residual ICC

When conflating level-specific effects in the *conflated-x model*, this residual *ICC* becomes distorted. Using Equations 7 and 8, the *conflated* residual *ICC* can be written as

$$\begin{aligned} ICC_r^* &= \frac{\tau_{00}^*}{\tau_{00}^* + \sigma^{2*}} \\ &= \frac{\tau_{00} + (\gamma_b - \gamma_c)^2 \text{var}(x_{\bullet j})}{\tau_{00} + (\gamma_b - \gamma_c)^2 \text{var}(x_{\bullet j}) + \sigma^2 + (\gamma_w - \gamma_c)^2 \text{var}(x_{ij} - x_{\bullet j})} \end{aligned} \tag{16}$$

The degree to which this conflated ICC_r^* differs from the unconflated ICC_r is hence a function of the squared differences between γ_b and γ_c and between γ_w and γ_c, as well as the amount of variance in x at the within-cluster level ($\text{var}(x_{ij} - x_{\bullet j})$) and at the between-cluster level ($\text{var}(x_{\bullet j})$).

Whenever there are no between-cluster differences on the outcome (i.e., the population cluster means are equivalent), the unconflated ICC_r is, by definition, equal to 0, as $\tau_{00} = 0$. However, using the formula in Equation 16, the conflated ICC_r^* would be given as

$$ICC_r^* \mid_{\tau_{00}, \gamma_b = 0} = \frac{\gamma_c^2 \text{var}(x_{\bullet j})}{\sigma^2 + \gamma_c^2 \text{var}(x_{\bullet j}) + (\gamma_w - \gamma_c)^2 \text{var}(x_{ij} - x_{\bullet j})} \tag{17}$$

Hence, whenever the conflated effect is nonzero, in the population, the conflated ICC_r^* will erroneously indicate that there are between-cluster differences.

This scenario is illustrated graphically in Figure 2, which shows plots similar to those in Figure 1, but here such that there are no mean differences between clusters on the outcome. However, as shown in Panel A, here there is a strong negative within-cluster effect, which in turn leads the conflated effect (depicted with the dashed lines) to be negative. As shown in Panel B, despite the fact that the true level-2 errors are all 0, the conflated errors are markedly nonzero, which will naturally lead to an expected nonzero ICC_r^*.

Illustration #3: Large ICC Observed When There Is No Between-Cluster Variation

As an illustration of such residual *ICC* distortion, consider an analysis of employees nested within companies and years spent working at company (*yearsworked*) as the independent variable and perceived authority within the company (*authority*) as the dependent variable (dataset and R script are available in the online supplemental materials). Here, the data (which were, again, simulated for illustrative purposes) have only chance outcome variability across companies (i.e., in the population, companies had the same average perceived level of authority across workers). As shown in Table 3, the unconflated model suggests a positive within-company effect, but a negligible (and nonsignificant) between-company effect (as expected, because there is almost no between-company outcome variance to explain), and correctly suggests that there is no apparent residual between-company variability (the residual *ICC* is exactly equal to 0). The *conflated-x model*, on the other hand, reveals a distorted view of both the effect of time spent working as well as the degree of between-company variability—here, the residual *ICC* suggests that over one fourth of the total unexplained variance is at the between-company level (residual *ICC* is 0.272), driven exclusively by the discrepancy in the within-company and between-company effects of *yearsworked.*

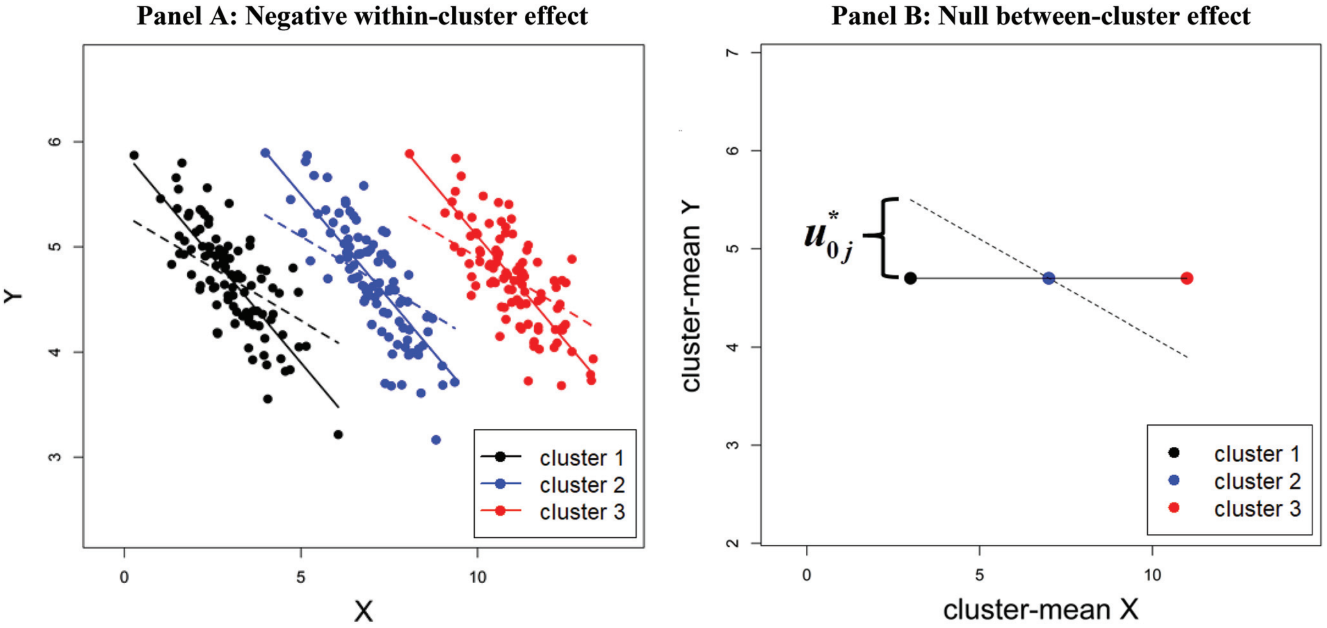
Issue 3: Conflating Level-Specific Effects Can Cause Severe Distortion in R-Squared Measures

Thus far, I have focused on how variance distortion can compromise estimation and inferences related directly to the variances themselves. It is important to realize, however, that these variances are often involved in the computation of distinct, but substantively important metrics—in particular, R-squared measures of explained variance. Even for researchers who aren't directly interested in quantifying the degree of random effect variability, distortion in

[7] As a preliminary step in MLM, researchers often utilize the Equation 15 formula for a random-intercept-only model, commonly calling this the *ICC*. The forthcoming issues I address are relevant specifically for quantifying *ICC* (or just random effect variance more generally) in models with covariates, not for the null model *ICC*.

Figure 2

Heuristic Illustration in Which a Large Degree of Between-Cluster Random-Effect Variance Would be Observed for a Conflated Model Even When There Is No Between-Cluster Outcome Variance



Note. See Figure 1 note for description of each line. See the online article for the color version of this figure.

R-squared can nonetheless provide misleading and inaccurate estimates of the importance (in terms of variance explained) for the fixed components of the model. I will hence here restrict focus to measures that are commonly used in psychological applications to quantify the variance explained specifically by the fixed effects (for a discussion on the utility of using a broader framework utilizing a full decomposition of outcome variance, see Rights & Sterba, 2019).

Some authors have previously acknowledged the seemingly peculiar behavior that certain R-squared *estimates* can sometimes decrease with the addition of predictors, or can even be negative when computed for a single model (e.g., Lahuis et al., 2014; Snijders & Bosker, 2012), attributing this to the increase in level-2 *residual* variances that can occur when adding predictors (explained earlier with reference to Equation 10). Here, I extend this work by showing that the degree to which R-squared values decrease *in the population* when adding predictors (or are negative to begin with) can be understood mathematically through the impact of conflation, and that this can occur for both total and level-specific R-squared measures. I also underscore that such negative estimates are just one realization of the more general distortion caused by conflation.

Analytic Explanation of Distortion in R-Squared

In Table 4, I list a set of commonly used R-squared measures for MLMs, define each, and make note of the specific way conflation will distort the measure (with relevant derivations provided in Appendix B). This table includes total measures—which quantify the proportion of total variance explained—as well as between-cluster and within-cluster measures—which quantify between-cluster and within-cluster variance explained, respectively. As noted in Column 4, certain measures will be underestimated when there is conflation; however, despite the fact that the conflated variances are in general *over*estimated (conceptually consistent with an *under*estimated R-squared), certain R-squared measure can also be *over*estimated.

Table 3

Illustration #3: Large Residual ICC Observed in Conflated Model When No True Between-Cluster Variance Actually Exists (Predicting Perceived Authority From Years Worked)

Parameter	Unconflated model estimate	Conflated model estimate
Intercept	6.027 (*SE* = 0.062; $p < .001$)	5.990 (*SE* = 0.053; $p < .001$)
Within-company slope of *yearsworked*	0.644 (*SE* = 0.013; $p < .001$)	—
Between-company slope of *yearsworked*	−0.010 (*SE* = 0.017; *p* = .549)	—
Conflated slope of *yearsworked*	—	0.593 (*SE* = 0.013; $p < .001$)
Level-1 error variance	1.300	1.308
Level-2 error variance	0.000	0.488
Residual *ICC*	.000	.272

Note. See *authority_exdat.txt* for dataset and *illustrativeexamples.R* for R script. Raw *yearsworked* centered to have a mean of 0. *p*-values based on *t*-statistics with Satterthwaite degrees of freedom approximation.

Table 4

Distortion in R-Squared Measures Induced by Conflation

Measure	Formula	What this measure quantifies	Impact of conflation
Total R-squared measures			
R^2_{SB} (Snijders & Bosker, 2012)	$1-\frac{\tau_{00}+\sigma^2}{\tau_{00,null}+\sigma^2_{null}}$	Proportion of total variance explained by all predictors via fixed components of slopes	Values will be systematically too small; negative conflated values when $(\gamma_b^2-(\gamma_b-\gamma_c)^2)ICC_x$is less than $(\gamma_w^2-(\gamma_w-\gamma_c)^2)(ICC_x-1)$
$R_t^{2(f_1)}$ (Rights & Sterba, 2019)	$\frac{\gamma_w^2\text{var}(x_{ij}-x_{\bullet j})}{\gamma_w^2\text{var}(x_{ij}-x_{\bullet j})+\gamma_b^2\text{var}(x_{\bullet j})+\tau_{00}+\sigma^2}$*	Proportion of total variance explained by the level-1-varying portion of predictors via fixed components of slopes	Values can be either systematically too small or too large; magnitude relative to $R_t^{2(f_2)}$ driven exclusively by ICC_x
$R_t^{2(f_2)}$ (Rights & Sterba, 2019)	$\frac{\gamma_b^2\text{var}(x_{\bullet j})}{\gamma_w^2\text{var}(x_{ij}-x_{\bullet j})+\gamma_b^2\text{var}(x_{\bullet j})+\tau_{00}+\sigma^2}$*	Proportion of total variance explained by the level-2-varying portion of predictors via fixed components of slopes	Values can be either systematically too small or too large; magnitude relative to $R_t^{2(f_1)}$ driven exclusively by ICC_x
Within-cluster R-squared measures			
R^2_{L1} (Raudenbush & Bryk, 2002)	$\frac{\sigma^2_{null}-\sigma^2}{\sigma^2_{null}}$	Proportion of level-1 variance explained by the level-1-varying portion of predictors via fixed components of slopes†	Values will be systematically too small; negative conflated values when γ_w^2 is less than $(\gamma_w-\gamma_c)^2$
$R_w^{2(f_1)}$ (Rights & Sterba, 2019)	$\frac{\gamma_w^2\text{var}(x_{ij}-x_{\bullet j})}{\gamma_w^2\text{var}(x_{ij}-x_{\bullet j})+\sigma^2}$*	Proportion of level-1 variance explained by the level-1-varying portion of predictors via fixed components of slopes	Values can be either systematically too small or too large
Between-cluster R-squared measures			
R^2_{L2} (Raudenbush & Bryk, 2002)	$\frac{\tau_{00,null}-\tau_{00}}{\tau_{00,null}}$	Proportion of level-2 variance explained by the level-2-varying portion of predictors via fixed components of slopes†	Values will be systematically too small; negative conflated values when γ_b^2 is less than $(\gamma_b-\gamma_c)^2$
$R_b^{2(f_2)}$ (Rights & Sterba, 2019)	$\frac{\gamma_b^2\text{var}(x_{\bullet j})}{\gamma_b^2\text{var}(x_{\bullet j})+\tau_{00}}$*	Proportion of level-2 variance explained by the level-2-varying portion of predictors via fixed components of slopes	Values can be either systematically too small or too large

Note. See derivations and further detail in Appendix B. Note that the sum of $R_t^{2(f_1)}$ and $R_t^{2(f_2)}$ is termed $R_t^{2(f)}$ and is equivalent to the marginal measure from Nakagawa and Schielzeth (2013). *Total R-squared measures* quantify the proportion of the total (i.e., both within and between) variance that is explained. *Within-cluster R-squared measures* quantify the proportion of the within-cluster/level-1 variance that is explained. *Between-cluster R-squared measures* quantify the proportion of the between-cluster/level-2 variance that is explained.

* Equations provided here for the *conflated-x model* in Equation 1 (for general matrix-based formulas, see Rights & Sterba, 2019, 2021).

† Definition holds only for fixed slope models; in random slope models, these can quantify variance explained by predictors via both fixed and random components of slopes (Rights & Sterba, 2019, 2021).

Additionally, as noted in Column 4, under specific conditions, certain measures will have negative conflated population values, particularly the measures that are computed as a proportion reduction in residual variance going from the random-intercept-only to the full model (namely R^2_{SB}, R^2_{L1}, and R^2_{L2}), which follows from the earlier discussion of variances increasing when adding predictors.

Illustration #4: Distortion in R-Squared Measures

To illustrate the distortion in R-squared measures in conflated models, for each of the earlier-presented examples and models (*GPA* predicting *salary*, *mood* predicting *symptoms*, *yearsworked* predicting *authority*), Table 5 shows the estimates of the R-squared measures from Table 4. The discrepancy between the conflated and unconflated versions of these estimates are consistent with the analytic results noted in the fourth column of Table 4, and whereas the unconflated model yields sensible values for all measures, the conflated models give some values below 0 (e.g., an estimate of R^2_{L2} of −.480 and an estimate of R^2_{L1} of −.002). Note also that the total and between-cluster measures are markedly different across the conflated and unconflated models, whereas the within-cluster measures are fairly similar (this pattern will also be observed across repeated samples in the forthcoming simulation).

Issue 4: Increased Variance in Estimation for Other Fixed Components in the Model

Historically, even researchers who are aware that conflated slopes can reflect an ambiguous blend of within-cluster and between-cluster effects may still specify such slopes in several circumstances: (a) when there is no substantive interest in disaggregating level-specific effects of a certain level-1 variable; (b) when there is no theoretical reason to believe these effects differ; and (c) when certain level-1 variables are simply seen as control variables. This practice echoes many methodological recommendations (mentioned earlier). Recent work has clarified that specifying a conflated slope—even for a

Table 5

Illustration #4: R-Squared Distortion in Conflated Models

Measure	Unconflated *salary* model estimate	Conflated *salary* model estimate	Unconflated *symptoms* model estimate	Conflated *symptoms* model estimate	Unconflated *authority* model estimate	Conflated *authority* model estimate
$R_t^{2(f_1)}$	.303	.265	.001	.005	.401	.245
$R_t^{2(f_2)}$	.019	.035	.151	.004	<.001	.157
R_{SB}^2	.320	.209	.151	.043	.400	.172
$R_w^{2(f_1)}$	.386	.383	.004	.014	.401	.360
$R_b^{2(f_2)}$	.086	.113	.228	.006	1.000*	.491*
R_{L1}^2	.391	.391	.001	−.002	.400	.397
R_{L2}^2	.054	−.480	.229	.066	Undefined*	Undefined*

Note. See *illustrativeexamples.R* for R script. *Salary, symptoms*, and *authority* are dependent variables defined in the prior Illustrative Example sections.
* Values for these measures are atypical given that there was 0 estimated between-cluster variance in the null model; in practice, these measures wouldn't be useful in this circumstance.

control variable—can induce bias in the slopes of other predictors in the model (Rights et al., 2020). This, however, occurs only when the level-1 predictors with conflated slopes have correlation of sufficient magnitude with the other predictors of interest.

Here, I extend this recent work by clarifying that the distortion in variances induced by conflation can adversely impact estimation of slopes of other predictors in the model, even in cases in which there is no correlation between the predictors with conflated slopes and the other predictors. The reason this occurs is that, holding all else constant, increased magnitude of the variance components leads to increased variance in estimation of the fixed components (Raudenbush & Bryk, 2002; Snijders & Bosker, 2012). As I show in the upcoming simulation, a single conflated slope included in the model can lead to greatly increased variance in estimating slopes, and hence reduced power.

Illustration #5: Increased Estimation Variance for Other Model Terms

To illustrate, here I consider an analysis of patients nested within hospitals (this simulated dataset and associated R script is available in the online supplemental materials) in which the primary question of interest is whether or not some hypothetical hospital-level program designed to improve patient care (*program*; binary independent variable in which 1 = implements program and 0 = does not implement program) is actually associated with higher patient-perceived quality of treatment (*quality*). Here, this hospital-level effect is also conditioned on patient-level socioeconomic status (SES) to account for possible confounding. Table 6 provides results from an unconflated model that includes as predictors *program* as well as hospital-mean and hospital-mean-centered SES. The magnitude of the estimated within-hospital slope of SES is less than that of the between-hospital slope; in this case, hospital-level SES may serve as a proxy for the resources available for a given hospital, which may be more impactful on patient care than within-hospital SES differences. The effect of *program*—conditioned on hospital-level SES—is positive and significant at an alpha of .05 (0.215; $p = .022$). However, in the conflated model that includes just *program* and (uncentered) SES, because of the underlying disparate level-specific effects *of SES*, the estimated random intercept variation is much greater than that of the unconflated model (0.891 vs. 0.409). As a result, the standard error for the slope of *program* is much larger in the conflated model (0.135 vs. 0.093), and the treatment effect is no longer significant ($p = .085$ vs. $p = .022$), despite the fact that the conflated model even has a slightly more positive estimate of the effect (0.235 vs. 0.215; here simply a result of sampling variability). In this example, the data were generated with a positive effect of *program*, and hence, the conflated model yields a crucial Type II error resulting from the distortion in the estimated random intercept variation.

Simulation With Fixed Slopes

The results presented thus far have been based on hypothetical population quantities and single-sample illustrations; to supplement these,

Table 6

Illustration #5: Increased Estimation Variance and Standard Errors for Terms in Conflated Models (Predicting Quality of Care From Program Implementation, Conditioning on SES)

Parameter	Unconflated model estimate	Conflated model estimate
Intercept	6.394 (*SE* = 0.064; $p < .001$)	6.384 (*SE* = 0.093; $p < .001$)
Slope of *program*	0.215 (*SE* = 0.093; $p = .022$)	0.235 (*SE* = 0.135; $p = .085$)
Within-hospital slope of *SES*	0.266 (*SE* = 0.014; $p < .001$)	—
Between-hospital slope of *SES*	1.326 (*SE* = 0.070; $p < .001$)	—
Conflated slope of *SES*	—	0.286 (*SE* = 0.014; $p < .001$)
Level-1 error variance	1.075	1.076
Level-2 error variance	0.409	0.891

Note. See *quality_exdat.txt* for dataset and *illustrativeexamples.R* for R script. *SES* is standardized. *p*-values based on *t*-statistics with Satterthwaite degrees of freedom approximation.

in this section, I provide corroborative simulation results that demonstrate the distortion discussed throughout this article across repeated samples. Here, I will emphasize the point that conflating a *single* slope, even if it is not the slope of primary interest, can cause any of the aforementioned issues. Thus, here I generate data from a model in which there are two level-1 independent variables, and investigate the impact of specifying one conflated effect (for the control variable, x_1) while disaggregating level-specific effects of the other variable (the predictor of primary interest, x_2). Hence, the generating model is

$$y_{ij} = \gamma_{00} + u_{0j} + \gamma_b x_{1\bullet j} + \gamma_w (x_{1ij} - x_{1\bullet j}) + \gamma_{b2} x_{2\bullet j} + \gamma_{w2}(x_{2ij} - x_{2\bullet j}) + e_{ij}$$
$$e_{ij} \sim N(0, \sigma^2);\ u_{0j} \sim N(0, \tau_{00}) \tag{18}$$

Here, γ_w and γ_{w2} are the within-cluster effects of x_1 and x_2, respectively, and likewise γ_b and γ_{b2} are the between-cluster effects of x_1 and x_2, respectively. The fitted models then include (a) the unconflated model that disaggregates the level-specific effects of both x_1 and x_2 via cluster-mean-centering and adding the cluster-mean as a separate predictor and (b) the conflated model that estimates only a single slope for grand-mean-centered x_1 (and thus constrains $\gamma_w = \gamma_b$) but still disaggregates the effects of x_2 (and thus does *not* constrain $\gamma_{w2} = \gamma_{b2}$). Here, I specify fixed slopes and assume cluster means are measured without error, but soon address generalizations (e.g., to random slope models). All models were fit using *lmer* (Bates et al., 2004) in R with restricted maximum likelihood estimation.

For each condition, I generated 1,000 samples from the model in Equation 18. Across conditions, I manipulated both the within-cluster and between-cluster effects of x_1 across the following range: {−2.5, −2, −1.5, −1, −0.5, 0, 0.5, 1, 1.5, 2, 2.5}. This allowed for a variety of values quantifying the degree of conflation, in that the difference between γ_w and γ_b ranged from −5 to 5, including 0 wherein the conflated model is correctly specified. I held the within-cluster effect of x_2 at 1 and the between-cluster effect of x_2 at −1; hence, both fitted models correctly disaggregated the disparate level-specific effects of x_2. I additionally held the fixed component of the intercept at 1, the level-1 error variance at 14, and the random intercept variance at 7. Both level-1 predictors were generated such that their within-cluster component and their between-cluster component each were standard normally distributed. Importantly, x_1 and x_2 were uncorrelated in the population; this allows us to isolate the impact of the level-1 and level-2 variance distortion on the *estimation variance* (and the associated power for testing) of γ_{w2} and γ_{b2} (if the predictors were correlated, this could also induce *bias* in estimation of γ_{w2} and γ_{b2} for the conflated model, as mentioned earlier). In terms of sample size, I manipulated the average cluster size to be either 3 (discrete uniformly distributed from 2 to 4), 7 (ranging from 5 to 9), or 50 (ranging from 45 to 55), and the number of clusters to be either 50, 100, or 200.

Note that, in the forthcoming results, there was virtually zero impact of the number of clusters. This is not surprising given that changing the number of clusters (but not average cluster size) does not change the amount of within-cluster versus between-cluster information, and hence does not affect the weighting described earlier (Equation 4). I hence here supply results for 200 clusters (of varying cluster size), but provide a full depiction of results in Supplemental Appendix A.

In Figure 3, Panel A, the *x*-axis denotes the population difference in the within-cluster and between-cluster effects of x_1 (i.e., $\gamma_w - \gamma_b$) and the *y*-axis the average estimated random intercept variance across 1,000 repeated samples. The horizontal line at *y* = 7 denotes the generating level-2 variance, τ_{00}. From this plot, it is clear that, across all cluster sizes, the conflated model overestimates the level-2 variance whenever $\gamma_w \neq \gamma_b$, and the degree to which it is overestimated is greater whenever there is greater difference between γ_w and γ_b. The extent to which there is overestimation, however, is mitigated when the cluster size is smaller. This is because, as cluster size increases, there is more within-cluster information relative to the amount of between-cluster information, causing the conflated effect to be weighted more toward the within-cluster effect relative to the between-cluster effect, and hence the degree of distortion (see Equation 8) is greater for the level-2 variance at larger cluster sizes. In contrast, the unconflated model provided accurate estimation across all conditions.

The Figure 3, Panel B plot is similar to that in Panel A, but here showing results for the estimated level-1 variance. Again, provided $\gamma_w \neq \gamma_b$, the level-1 variance will be overestimated in the conflated model, whereas it is accurately estimated in the unconflated model. In contrast to the results for the level-2 variance, however, the distortion induced in estimating the level-1 variance is mitigated at *larger* cluster sizes (and is virtually nonexistent at the largest cluster size), because increasing cluster size leads to the conflated effect being more similar to the within-cluster effect and hence yields less distortion (see Equation 7). Hence, Figure 3 demonstrates the trade-off mentioned earlier—the less that conflation distorts the level-1 variance, the more it will distort the level-2 variance, and vice versa. Furthermore, I note that the overall impact on the level-1 variance is much less than that of the level-2 variance, which is obvious when comparing the scale of the *y*-axes for both plots. The level-2 variance is wildly distorted (e.g., at the extremes, the conflated model yields values over three times the generating variance), whereas the level-1 variance only marginally so.[8]

In Supplemental Appendix B, I discuss the full simulation results, including an investigation of the extent to which the conflated model is subject to the aforementioned Issues 1–4. In short, it is shown that, as expected, these issues are avoided across all conditions for the unconflated model, and are generally worsened in the conflated model when level-specific effects differ to a greater extent.

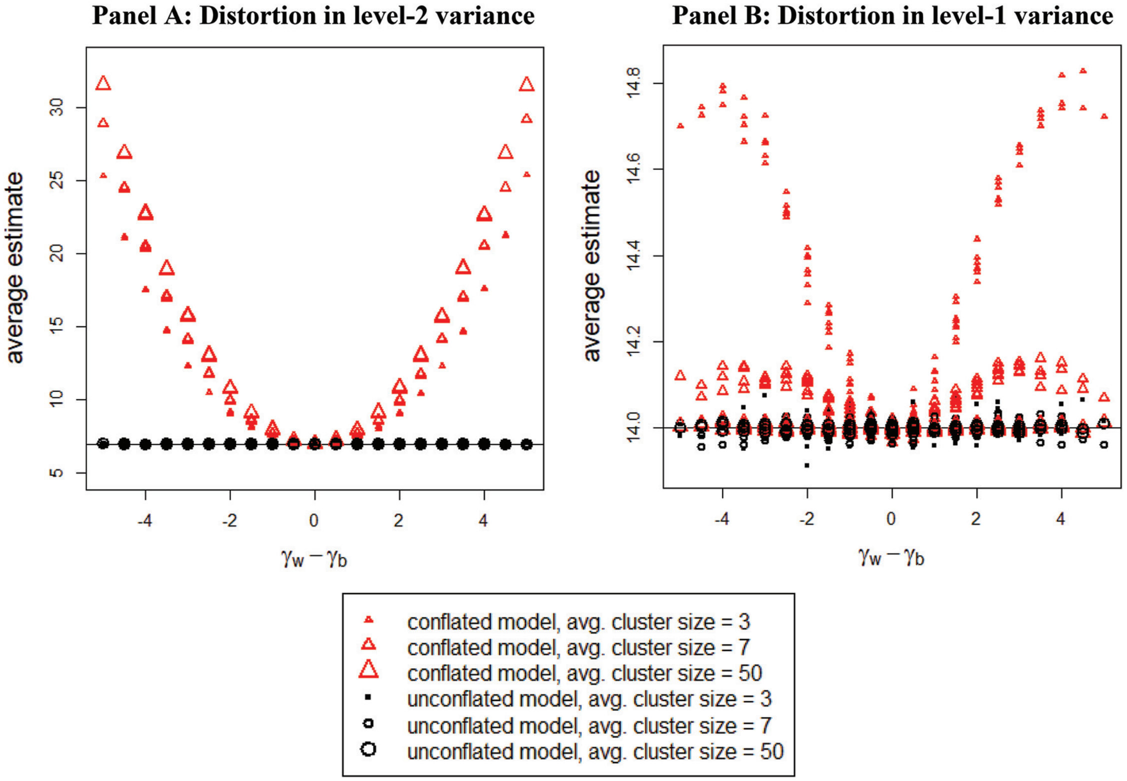
Impact of Conflation in Random Slope Models

Thus far, I have made the simplifying assumption that all slopes were fixed, so as to isolate the impact that *fixed conflation* (i.e., implicitly setting $\gamma_w = \gamma_b$) has on distorting variances. However, in random slope models, there is also the potential for *random conflation*, defined as placing an implicit equality

[8] A nuance to the results in Figure 3, Panel B is that the overestimation is less severe at the extremes of the *x*-axis. What is happening here is that these generating conditions yield a large degree of within-cluster explained variance (because the within-cluster effect is large at these conditions) relative to the between-cluster explained variance, leading to more within-cluster precision, and hence, the conflated effect being closer to the within-cluster effect. Thus, though not a specific focus here, it is important to note that the degree of within- vs. between-cluster explained variation can impact the extent to which the conflated effect is weighted towards the within-cluster vs. between-cluster effect (Raudenbush & Bryk, 2002).

Figure 3

Simulation Results: Distortion in Level-1 and Level-2 Variances in the Conflated Model



Note. Horizontal line depicts the correct population level-1/level-2 variance. Number of clusters here is set at 200. See the online article for the color version of this figure.

constraint on the within-cluster and between-cluster random components (u's) of the level-1 variable. Such random conflation is a new concept that is not well recognized or understood, but has been shown to result in erroneous interpretation and inferences regarding across-cluster slope heterogeneity, as well as biased standard errors for fixed effects (Rights & Sterba, 2020). In this section, I newly show that random conflation additionally distorts all other variance components—both at level-1 and level-2—similar to the distortion caused by fixed conflation.

To illustrate random conflation, here I present what I term the *fully conflated-x model*, which is similar to the earlier *conflated-x* model in Equation 3, but also includes a random component of x:

$$\begin{aligned} y_{ij} &= \gamma^*_{00} + u^*_{0j} + \gamma_c x_{ij} + u_{cj} x_{ij} + e^*_{ij} \\ &= \gamma^*_{00} + u^*_{0j} + \gamma_c (x_{ij} - x_{\bullet j} + x_{\bullet j}) + u_{cj}(x_{ij} - x_{\bullet j} + x_{\bullet j}) + e^*_{ij} \\ &= \gamma^*_{00} + u^*_{0j} + \gamma_c (x_{ij} - x_{\bullet j}) + \gamma_c x_{\bullet j} + u_{cj}(x_{ij} - x_{\bullet j}) + u_{cj} x_{\bullet j} + e^*_{ij} \end{aligned} \tag{19}$$

This re-expression in Equation 19 demonstrates both the *fixed conflation* described earlier (in that the fixed components for $x_{ij} - x_{\bullet j}$ and $x_{\bullet j}$ are implicitly constrained equal to γ_c) as well as *random conflation* (in that random components for $x_{ij} - x_{\bullet j}$ and $x_{\bullet j}$ are implicitly constrained equal to u_{cj}). This is thus a more constrained version of the following unconflated model with separate random components for both $x_{ij} - x_{\bullet j}$ and $x_{\bullet j}$:

$$y_{ij} = \gamma_{00} + u_{0j} + \gamma_w (x_{ij} - x_{\bullet j}) + \gamma_b x_{\bullet j} + u_{wj}(x_{ij} - x_{\bullet j}) + u_{bj} x_{\bullet j} + e_{ij} \tag{20}$$

where u_{wj} is the random component of $x_{ij} - x_{\bullet j}$ and u_{bj} that of $x_{\bullet j}$. It is worth noting that researchers are largely unfamiliar with this idea of including a random slope for a level-2 variable—doing so amounts to modeling heteroscedasticity at level-2 in allowing the across-cluster variance of the model-implied cluster means of y to follow a quadratic pattern as a function of the cluster means of x (Goldstein, 2010; Rights & Sterba, 2020; Snijders & Bosker, 2012). Here, I thus term the model in Equation 20 the *heteroscedastic unconflated-x model*. This atypical specification is necessary to present in order to define the implicit assumptions made by the *fully conflated-x model*; specifically, the conflated model is nested within this unconflated model, placing one constraint on the fixed portion, $\gamma_w = \gamma_b$, and three constraints on the random portion, $\text{var}(u_{wj}) = \text{var}(u_{bj})$, $\text{corr}(u_{wj}, u_{bj}) = 1$, and $\text{corr}(u_{0j}, u_{wj}) = \text{corr}(u_{0j}, u_{bj})$ (Rights & Sterba, 2020).

Though fitting a model subject to both fixed and random conflation (i.e., that in Equation 19) is common, another typical practice is to fit models that properly disaggregate the fixed component of x but *fail to disaggregate the random component*. In particular, the conventional *random-conflated contextual effect model*, which adds a fixed slope of $x_{\bullet j}$ to the model in Equation 19, is given as:

$$\begin{aligned} y_{ij} &= \gamma^*_{00} + u^*_{0j} + \gamma_w x_{ij} + \gamma_{bc} x_{\bullet j} + u_{cj} x_{ij} + e^*_{ij} \\ &= \gamma^*_{00} + u^*_{0j} + \gamma_w(x_{ij} - x_{\bullet j} + x_{\bullet j}) + \gamma_{bc} x_{\bullet j} + u_{cj}(x_{ij} - x_{\bullet j} + x_{\bullet j}) + e^*_{ij} \\ &= \gamma^*_{00} + u^*_{0j} + \gamma_w(x_{ij} - x_{\bullet j}) + (\gamma_{bc} + \gamma_w) x_{\bullet j} + u_{cj}(x_{ij} - x_{\bullet j}) + u_{cj} x_{\bullet j} + e^*_{ij} \end{aligned} \tag{21}$$

This re-expression in Equation 21 shows that the fixed component is disaggregated (noting that γ_{bc} is called the *contextual effect* that represents the difference in the between-cluster and within-cluster effects of x; Raudenbush & Bryk, 2002), but that the random component is still conflated. In other words, this model allows $x_{ij} - x_{\bullet j}$ and $x_{\bullet j}$ to have distinct fixed effects, but (contradictorily) defines their individual influence via random effects using the exact same random term (Rights & Sterba, 2020). This model is similarly nested within the *heteroscedastic unconflated-x model*, now placing only the three aforementioned constraints on the random portion.

It is important to note that: (1) the same exact distortion in variance components resulting from fixed conflation (discussed throughout the current article) still occurs in random slope models; and (2) random conflation can cause additional distortion in these variance components. That is, discrepant level-specific fixed components (wherein $\gamma_w \neq \gamma_b$) will yield distortion in the level-1 variance and the random intercept variance of the *fully conflated-x model*, and discrepant level-specific random components (wherein $u_{wj} \neq u_{bj}$) will yield distortion in these variances of both the *fully conflated-x model* and the *random-conflated contextual effect model*. The mathematical basis of this distortion induced by random conflation is discussed further in Appendix C.

To avoid random conflation, one option is to fit the *unconflated heteroscedastic random-slope model* as written in Equation 20 (which is also analytically equivalent to a contextual effect model that adds a random component of the cluster mean of x; Rights & Sterba, 2020). However, conflation is also avoided in the more typical, simpler model with a random slope of cluster-mean-centered x that excludes the random slope of the cluster mean of x. That is, one can assume homoscedasticity by fitting what I term here the *homoscedastic unconflated-x model*, which is equal to Equation 20 when excluding the u_{bj} term (i.e., setting all $u_{bj} = 0$). This model is simpler and thus less likely to run into convergence issues than the heteroscedastic model in applied practice, but if there is heteroscedasticity at level-2, this model would be underspecified (and would yield biased standard errors for slopes of the cluster mean of $x_{\bullet j}$; Rights & Sterba, 2020). However, this does not imply that variances would otherwise be distorted—conceptually, in this underspecified model, the intercept variance (assuming all predictors have a mean of 0) would represent the remaining between-cluster variance that is not accounted for by the cluster mean of x via its fixed component, and should thus be equal to $\tau_{00} + \text{var}(u_{bj})\text{var}(x_{\bullet j})$ (Rights & Sterba, 2021).

Simulation With Random Slopes

Noting that the distortion induced by random conflation is less mathematically apparent and intuitive than that induced by fixed conflation (see Appendix C), here I investigate the former via a simulation in which data are generated from a model similar to the *heteroscedastic unconflated-x model* in Equation 20, but with also a second level-1 predictor, x_2, including a fixed slope of $x_{2ij} - x_{2\bullet j}$ and of $x_{2\bullet j}$.[9] The generating conditions are identical to the earlier fixed slope simulation, but here including the random effect variances $\text{var}(u_{wj})$ and $\text{var}(u_{bj})$, holding constant the fixed components of x_1 ($\gamma_w = 1$, $\gamma_b = -1$), and increasing the smallest cluster size condition to range from 3 to 4 instead of 2 to 4 (to prevent unidentifiability when there are many cluster sizes of 2 along with the random effects). I compare three broad conditions of $\text{var}(u_{wj}) = \text{var}(u_{bj}) = 4$, $\text{var}(u_{wj}) = 6 > \text{var}(u_{bj}) = 2$, and $\text{var}(u_{wj}) = 2 < \text{var}(u_{bj}) = 6$ while, for each of these, manipulating the correlation between u_{wj} and u_{bj}.[10] As for fitted models, I compare the *random-conflated contextual effect model* to the *heteroscedastic unconflated-x model* to the *homoscedastic unconflated-x model*, but with each also including a fixed slope of $x_{2\bullet j}$ and $x_{2ij} - x_{2\bullet j}$. To ensure comparability across models, all predictors are centered to have a mean of 0, so that for each model the random intercept variance, in theory, reflects the between-cluster variance not accounted for by predictors via either fixed or random effects (Rights & Sterba, 2021). As a clarification on how this simulation differs from prior work, Rights and Sterba (2019) looked at similar sets of conditions, but focused on distortion in the random slope variance itself. Here, instead, I show how random conflation additionally distorts the random intercept variance, the level-1 variance, and estimation of slopes of additional predictors in the model.

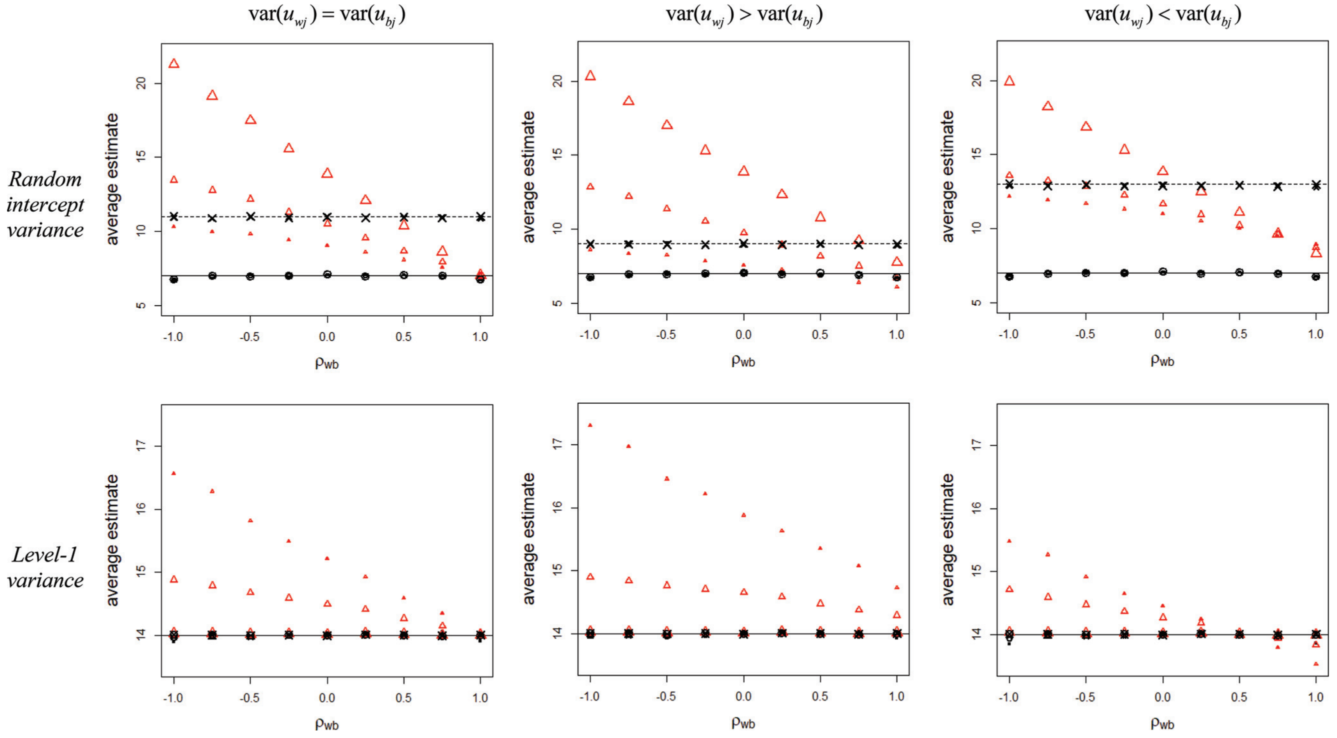
Figure 4 provides results for the across-1,000-sample average random intercept variance and level-1 variance for the three models (holding number of clusters constant at 200, given that results were virtually identical across the level-2 sample sizes; full results provided in Supplemental Appendix A), with the y-axis denoting the average estimate, the solid horizontal lines denoting the population variance, and the x-axis denoting the correlation between u_{wj} and u_{bj}. Focusing first on the $\text{var}(u_{wj}) = \text{var}(u_{bj})$ conditions and on the comparison between the *heteroscedastic unconflated-x model* and the *random-conflated contextual effect model*, it is evident that, when the assumptions of the latter are met (i.e., $\text{var}(u_{wj}) = \text{var}(u_{bj})$ and $\text{corr}(u_{wj}, u_{bj}) = 1$), both models adequately recover the generating level-1 and level-2 variance. However, the more incorrect the assumption of $\text{corr}(u_{wj}, u_{bj}) = 1$ (i.e., as the correlation decreases going right-to-left), the more upward bias there is for the random-conflated model. The bias is much more pronounced for the level-2 variance at *larger* cluster sizes, and more pronounced for the level-1 variance at *smaller* cluster sizes, mirroring the pattern found resulting from fixed conflation. Additionally, noting the scale of

[9] If one were to also add random components to x_2 for both the generating and fitted models, this would be investigating the *joint* impact of random conflation for the variable itself, x_2, as well as random conflation of the other variable, x_1, on estimation of the slopes for x_2 (the former of which was investigated in Rights & Sterba, 2020, and the latter of which is the focus here).

[10] Note that there being distortion induced by random conflation is not predicated on there being heteroscedasticity at level-2; if $\text{var}(u_{bj})$ is excluded from the generating model (the more typical assumption), distortion is still observed.

Figure 4

Distortion of Level-1 and Random Intercept Variances in the Random-Conflated Contextual Effect Model



Note. See legend in Figure 3, noting that the X's represent results from the *homoscedastic unconflated model* and smallest cluster size = average of 3.5 (number of clusters is 200). Solid horizontal lines denote the population variance; dashed horizontal lines denote $\tau_{00} + \text{var}(u_{bj})\text{var}(x_{\bullet j})$, i.e., the portion of the population level-2 variance that is not accounted by the cluster-mean of x via its fixed component. See the online article for the color version of this figure.

the y-axis, the distortion that can occur in the level-2 variance is more dramatic than that of the level-1 variance.

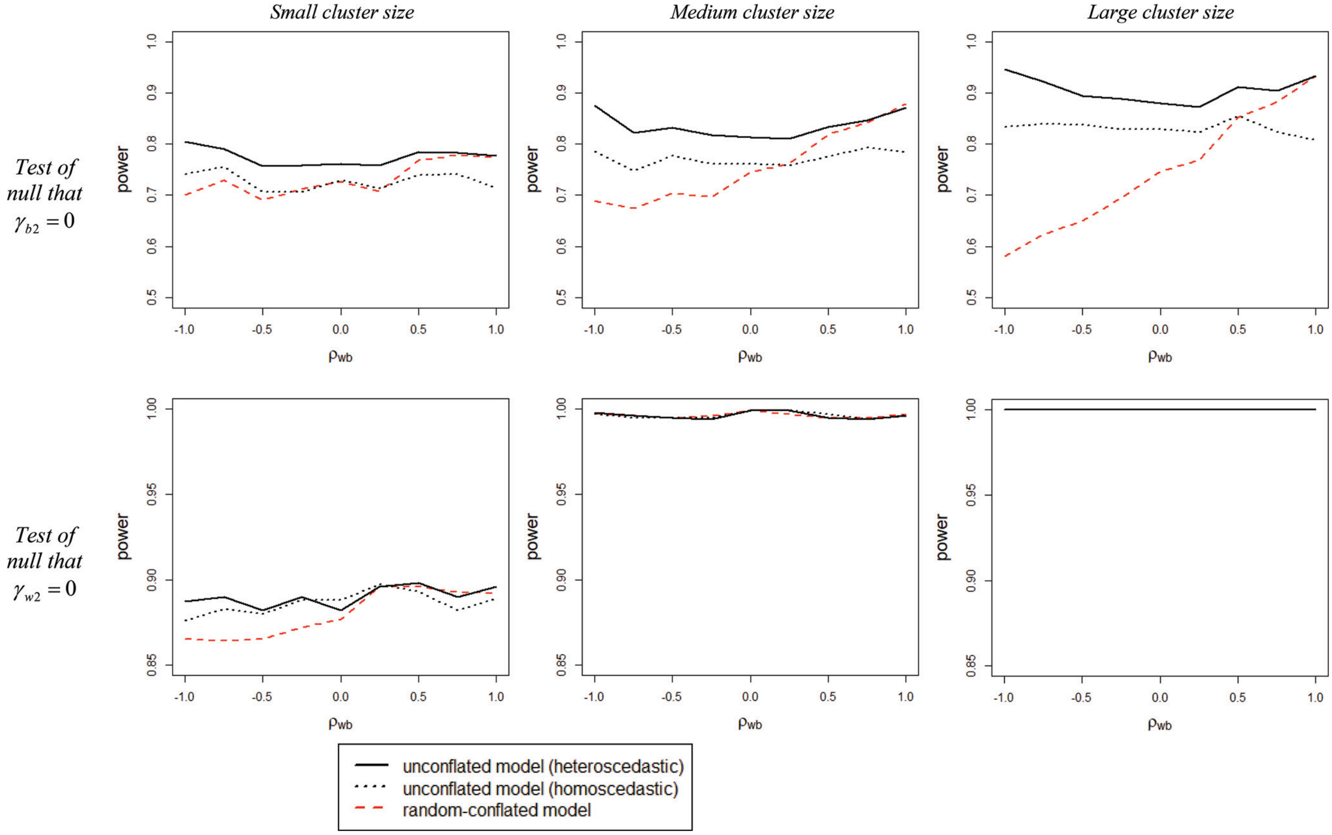
For the conditions in which $\text{var}(u_{wj}) > \text{var}(u_{bj})$ and $\text{var}(u_{wj}) < \text{var}(u_{bj})$, the pattern of distortion is very similar, with the exception that there is still distortion in level-1 and level-2 variance when $\text{corr}(u_{wj}, u_{bj}) = 1$. In these conditions, the equal variance assumption is never met, and when $\text{var}(u_{wj}) > \text{var}(u_{bj})$, the random intercept variance can actually be slightly underestimated, and when $\text{var}(u_{wj}) < \text{var}(u_{bj})$, the level-1 variance can be slightly underestimated. Hence, though random conflation generally appears to cause distortion in terms of upward bias, it can, in some specific circumstances, induce downward bias.

Next, for the *homoscedastic unconflated model*, Figure 4 reveals that, despite incorrectly modeling the random effect structure by failing to include a random component for $x_{\bullet j}$, this model still accurately recovers the *overall* amount of variance at level-1 and level-2. Though this result is immediately evident for the level-1 variance (all conditions overlap with the solid line), for the level-2 variance, note that (as discussed previously) the random intercept variance of the homoscedastic model is conceptually an estimate of $\tau_{00} + \text{var}(u_{bj})\text{var}(x_{\bullet j})$. This population value is given by the dashed horizontal line, and it is clear that the homoscedastic model accurately reflects this value across all conditions. Hence, though this model could yield inaccurate standard errors for level-2 predictors when there is heteroscedasticity, it does not result in the type of interpretational distortion discussed in this article—that is, the level-1 and random intercept variances still accurately represent the portion of level-specific variance that is not accounted for by the included predictors.

Last, to investigate the impact of random conflation on estimating the slope of a different predictor, in Figure 5, I provide the empirical power in testing the slopes associated with x_2. To provide a discernable visual, the results are presented for the $\text{var}(u_{wj}) = \text{var}(u_{bj})$ conditions, broken into separate plots for each cluster size (full results available in Supplemental Appendix A). Focusing first on the test of $\gamma_{b2} = 0$ for small cluster sizes, the properly specified *heteroscedastic unconflated model* has a little more power than the underspecified *homoscedastic unconflated model*, and the power for each remains fairly consistent across the range of $\text{corr}(u_{wj}, u_{bj})$. In contrast, the power for the *random-conflated contextual effect model* decreases as $\text{corr}(u_{wj}, u_{bj})$ decreases—when $\text{corr}(u_{wj}, u_{bj})$ is high (i.e., the assumption of $\text{corr}(u_{wj}, u_{bj}) = 1$ is met or nearly met), the power is comparable to that of the *heteroscedastic unconflated model*, but when $\text{corr}(u_{wj}, u_{bj})$ is low, the power is lowest of the three models. This same exact pattern holds for each cluster size condition, but is much more pronounced at the largest cluster size wherein the random intercept variance is more highly distorted in the random-conflated model. Interestingly, whereas increasing the cluster size always increases power for both unconflated models, increasing cluster size actually *decreases* power in the random-conflated model when $\text{corr}(u_{wj}, u_{bj})$ is not close to 1. This can be explained by the variance distortion being worse with increasing cluster size, as shown in Figure 4.

Figure 5

Comparing Power in Random-Conflated Model Versus Heteroscedastic and Homoscedastic Unconflated Models



Note. Results presented for conditions in which $\text{var}(u_{wj}) = \text{var}(u_{bj})$. Row 1 displays results with number of clusters = 100, row 2 displays results with number of clusters = 50 (see Supplemental Appendix A for full results). See the online article for the color version of this figure.

The power in testing $\gamma_{w2} = 0$ shown in the second row of Figure 5, focusing first on the small cluster size conditions, mirrors the general pattern found for the power in testing $\gamma_{b2} = 0$. However, in general there is little difference between the three models, and the difference disappears when increasing the cluster size (a supplemental simulation with a smaller effect size such that power did not approach 1 showed this same result). Hence, detriments in power of testing slopes of level-1 predictors resulting from conflation will likely only happen at small cluster sizes, and even under such conditions, it is much less dramatic than that for testing slopes of level-2 predictors.

Overall, these simulation results show that random conflation results in not only distortion in the random slope variance (as shown in Rights & Sterba, 2020), but also of level-1 variance and of random intercept variance. When avoiding conflation by cluster-mean-centering level-1 predictors, the consideration of whether or not to model heteroscedasticity at level-2 is a separate issue than that discussed in the current article, as the distortion discussed herein is automatically avoided even when this term is omitted. Nonetheless, researchers might wish to at least consider the possibility of heteroscedastic variances and investigate it (e.g., via diagnostic plots; Snijders & Berkhof, 2008), although it should be noted that there are additional methods to account for such heteroscedasticity in cluster-mean-centered models beyond adding a random component for the cluster-mean of *x* (e.g., Hedeker et al., 2012).

Discussion

Summary

In this article, I provided analytic derivations, illustrative demonstrations, and corroborative simulations to clarify the way in which conflating level-specific effects in multilevel models can lead to highly distorted variance components. I clarified several key implications of such distortion, namely that it can lead to: (1) both level-1 and level-2 variance components increasing (in the population) after adding predictors; (2) a large observed degree of between-cluster random-effect variance in cases in which there is actually no between-cluster variance; (3) biased and uninterpretable R-squared measures; and (4) increased estimation variance and reduced power in testing fixed components of the model. I further showed how these issues are avoided by fitting models that properly disaggregate level-specific effects.

Recommendations for Practice

Given the results presented in this article, the first and most obvious recommendation is that researchers not fit conflated models. A straightforward way to avoid both fixed and random conflation is to ensure level-1 predictors are cluster-mean-centered to estimate within-cluster effects and, if between-cluster effects are also of interest, to add the cluster-mean as a separate predictor. Though this recommendation is consistent with much established literature, it is at odds with the notion that specifying conflated slopes is defensible when there is no interest in separately considering within-cluster versus between-cluster effects, there is no reason to expect level-specific effects to differ, or the level-1 variables are primarily thought to be control variables—importantly, the issues discussed in this article can occur under each of these three circumstances. Second, I recommend that researchers use caution in interpreting results from prior studies that fit conflated models. Importantly, however, one must be cautious not only of the conflated slopes themselves, but also of the variance components and any metrics that include these in their computation (e.g., measures of intraclass correlation and R-squared), as well as any inference made regarding the fixed components in the model (particularly those of level-2 predictors).

There are, however, several caveats worth noting. The first is that the distortion discussed in this article only occurs when level-specific effects actually differ, and when level-specific effects are similar, the distortion can be fairly small. Hence, if a researcher had a very strong theoretical reason to believe level-specific effects are equivalent (and perhaps also tested this empirically), fitting a conflated model might not be too problematic. The second caveat is that conflation will cause much greater distortion in level-2 variance components than level-1 variance components. Hence, in interpreting results from prior research, metrics involving only the level-1 variance (e.g., within-cluster R-squared measures) will likely not have been distorted too greatly, and inferential testing of fixed components associated with level-1 predictors that varied exclusively within-cluster was likely not overly compromised. A third caveat is that, in certain cases, the between-cluster variance in a level-1 predictor might be incredibly small (e.g., in longitudinal settings, the *time* variable might be nearly balanced), and for such predictors, conflation is unlikely to have much of an impact. In such cases, the conflated effect would be nearly identical to the within-cluster effect (Raudenbush & Bryk, 2002) and the conflated variances would be similar to the unconflated variances.

As a final caveat/consideration, in this article, I focused on situations in which the level-2 sample size (i.e., number of clusters) allowed for reasonable estimation for slopes of cluster means of each level-1 independent variable. In practice, however, the level-2 sample size may sometimes be low in comparison with the number of level-1 variables. In such cases, researchers may be accustomed to fitting conflated models for parsimony (i.e., estimating a single slope per level-1 predictor) while retaining the between-cluster variance of the level-1 predictors so that they can serve as controls for other level-2 predictors of interest (Enders & Tofighi, 2007). When the level-specific effects of level-1 variables are exactly equivalent, such models give the most efficient estimation of both the effect of the level-1 variable (Raudenbush & Bryk, 2002) as well as the slope of other level-2 predictors (Rights et al., 2020). However, when these level-specific effects do differ, the bias for the slopes of additional level-2 predictors can actually be much worse than the bias induced by failing to include a confounding level-1 variable altogether (Rights et al., 2020), and, as discussed in this article, the increased random intercept variance induced by conflating can further adversely impact estimation.[11] I thus recommend that even when the level-2 sample size is small, it is better to still cluster-mean-center level-1 variables rather than conflating their effects. If cluster means of level-1 predictors cannot reasonably be included as well, and there is risk of the cluster means confounding other between-cluster relationships, researchers can note this as an inherent limitation of making between-cluster inferences with small level-2 sample sizes.

Impact of Using Latent Versus Observed Cluster Means

An assumption made in all of the simulations was that the cluster means of the level-1 independent variable were measured without error, as is standard in traditional multilevel modeling. That is, I assumed that the cluster mean of x could be accurately represented by the observed sample mean. Though this is sometimes a reasonable assumption, other times it is more appropriate to assume that some underlying *latent* cluster mean—of which the observed mean is an imperfect representation—is responsible for the between-cluster effect (for further detail, see, e.g., Lüdtke et al., 2008). In the latter case, one can model the cluster mean of x as a latent variable using multilevel structural equation modeling (MSEM). I underscore, however, that the general ideas and the population-level derivations provided in this article hold in either case. For instance, in Equations 7 and 8, γ_w and γ_b could theoretically represent the fixed components associated with either observed or latent predictors. If one were to fit an unconflated model while erroneously assuming cluster means were measured without error, this would avoid the issues noted in the current article related to conflation, but could induce a separate source of bias in the slopes (Asparouhov & Muthén, 2019; Lüdtke et al., 2008).

Future Directions

Future research can more thoroughly investigate the impact of the aforementioned complexities, for example, the choice between using latent versus observed cluster means, and the recovery of different forms of heteroscedastic variances. Additionally, future work can examine the generality of the current results to a broader class of latent variable models, outside the scope of multilevel modeling specifically. Though the general logic in the current article would apply in any context in which errors are specified at the cluster-level and at the observation-level, future work can explicitly examine how conflation might impact estimation in structural equation models (such as latent

[11] As a quick test, I simulated data using the same conditions of the original simulation but with only 20 clusters of size 10 and 15 independent level-1 predictors (each with an associated within-cluster effect of 1 and between-cluster effect of -1), plus an additional (uncorrelated) level-2 predictor with variance 1 and slope of 1. The across-1,000-sample average intercept variance was 65.097 for the conflated model, 7.286 for the unconflated model with cluster means of all level-1 variables (the correct population value was 7), and 22.161 for the unconflated model with no cluster means of x (correct population value, given the lack of cluster means to explain level-2 variance, was 22). Though the power in testing if the slope of the level-2 predictor was different than 0 at alpha = .05 was (unsurprisingly) very low for all models, it was worst for the conflated model (.077 vs. .097 and .149, respectively).

growth curve models) which are also used to accommodate multilevel data structures.

As a final noted future direction, I highlight an important question that is difficult to assess given the current state of the literature—*how much* has this distortion in variance components compromised results in published research? From the results of this article, it is clear that the distortion will be most pronounced when level-specific effects of level-1 variables differ meaningfully, but what is less clear is the degree to which researchers can expect such effects to differ in practice. Though certain examples lend themselves to a theoretical expectation of disparate level-specific effects, others might be less obvious. Future work can systematically investigate substantive contexts in which level-specific effects are more likely to differ, for instance by reviewing published literature that has decomposed these effects, to determine the cases in which variance distortion has historically been most problematic.

References

Asparouhov, T., & Muthén, B. (2019). Latent variable centering of predictors and mediators in multilevel and time-series models. *Structural Equation Modeling*, *26*(1), 119–142. https://doi.org/10.1080/10705511.2018.1511375

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2004). lme4: Linear mixed-effects models using Eigen and S4 (R package version 1.0–6) [Computer software]. https://cran.rproject.org/web/packages/lme4/index.html

Bell, A., Jones, K., & Fairbrother, M. (2018). Understanding and misunderstanding group mean centering: A commentary on Kelley et al.'s dangerous practice. *Quality & Quantity: International Journal of Methodology*, *52*(5), 2031–2036. https://doi.org/10.1007/s11135-017-0593-5

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. https://doi.org/10.1037/met0000159

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

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*(1), 583–619. https://doi.org/10.1146/annurev.psych.093008.100356

Dalal, D. K., & Zickar, M. J. (2012). Some common myths about centering predictor variables in moderated multiple regression and polynomial regression. *Organizational Research Methods*, *15*(3), 339–362. https://doi.org/10.1177/1094428111430540

Enders, C. K. (2013). Centering predictors and contextual effects. In M. A. Scott, J. S. Simonoff, & B. D. Mark (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

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

Gelman, A., Shor, B., Bafumi, J., & Park, D. (2007). Rich state, poor state, red state, blue state: What's the matter with Connecticut? *Quarterly Journal of Political Science*, *2*(4), 345–367. https://doi.org/10.1561/100.00006026

Goldstein, H. (2010). *Multilevel statistical models* (4th ed.). Wiley. https://doi.org/10.1002/9780470973394

Gottard, A., Grilli, L., & Rampichini, C. (2007). A chain graph multilevel model for the analysis of graduates' employment. In L. Fabbris (Ed.), *Effectiveness of university education in Italy: Employability, competencies, human capital* (pp. 169–182). Physica-Verlag.

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. (2020). 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

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

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

Hoffman, L. (2019). On the interpretation of parameters in multivariate multilevel models across different combinations of model specification and estimation. *Advances in Methods and Practices in Psychological Science*, *2*(3), 288–311. https://doi.org/10.1177/2515245919842770

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., Moerbeek, M., & van de Schoot, R. (2018). *Multilevel analysis: Techniques and applications* (3rd ed.). Routledge.

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

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

Krull, J. L., & MacKinnon, D. P. (1999). Multilevel mediation modeling in group-based intervention studies. *Evaluation Review*, *23*(4), 418–444. https://doi.org/10.1177/0193841X9902300404

Kuo, Y. F., Loresto, F. L., Jr., Rounds, L. R., & Goodwin, J. S. (2013). States with the least restrictive regulations experienced the largest increase in patients seen by nurse practitioners. *Health Affairs*, *32*(7), 1236–1243. https://doi.org/10.1377/hlthaff.2013.0072

LaHuis, D. M., Hartman, M. J., Hakoyama, S., & Clark, P. C. (2014). Explained variance measures for multilevel models. *Organizational Research Methods*, *17*(4), 433–451. https://doi.org/10.1177/1094428114541701

Lai, M. H., & Kwok, O. M. (2015). Examining the rule of thumb of not using multilevel modeling: The "design effect smaller than two" rule. *Journal of Experimental Education*, *83*(3), 423–438. https://doi.org/10.1080/00220973.2014.907229

Lüdtke, O., Marsh, H. W., Robitzsch, A., Trautwein, U., Asparouhov, T., & Muthén, B. (2008). The multilevel latent covariate model: A new, more reliable approach to group-level effects in contextual studies. *Psychological Methods*, *13*(3), 203–229. https://doi.org/10.1037/a0012869

Marsh, H. W., Seaton, M., Trautwein, U., Lüdtke, O., Hau, K. T., O'Mara, A. J., & Craven, R. G. (2008). The big-fish–little-pond-effect stands up to critical scrutiny: Implications for theory, methodology, and future research. *Educational Psychology Review*, *20*(3), 319–350. https://doi.org/10.1007/s10648-008-9075-6

McCoach, D. B. (2010). Hierarchical linear modeling. In G. R. Hancock & R. O. Mueller (Eds.), *The reviewer's guide to quantitative methods in the social sciences* (pp. 123–140). Routledge.

Murayama, K., Goetz, T., Malmberg, L.-E., Pekrun, R., Tanaka, A., & Martin, A. J. (2017). Within-person analysis in educational psychology: Importance and illustrations. *British Journal of Educational Psychology Monograph Series II*, *12*, 71–87.

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, *4*(2), 133–142. https://doi.org/10.1111/j.2041-210x.2012.00261.x

Nakagawa, S., Johnson, P. C. D., & Schielzeth, H. (2017). The coefficient of determination R^2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *Journal of the Royal Society, Interface*, *14*(134), 20170213. https://doi.org/10.1098/rsif.2017.0213

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

Peugh, J. L. (2010). A practical guide to multilevel modeling. *Journal of School Psychology*, *48*(1), 85–112. https://doi.org/10.1016/j.jsp.2009.09.002

Preacher, K. J., & Sterba, S. K. (2019). Aptitude-by-treatment interactions in research on educational interventions. *Exceptional Children*, *85*(2), 248–264. https://doi.org/10.1177/0014402918802803

Preacher, K. J., Zyphur, M. J., & Zhang, Z. (2010). A general multilevel SEM framework for assessing multilevel mediation. *Psychological Methods*, *15*(3), 209–233. https://doi.org/10.1037/a0020141

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

Rencher, A. C., & Schaalje, G. B. (2007). *Linear models in statistics* (2nd ed.). Wiley. https://doi.org/10.1002/9780470192610

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). *On the common but problematic specification of conflated random slopes in multilevel models* [Manuscript submitted for publication]. Department of Psychology, University of British Columbia.

Rights, J. D., & Sterba, S. K. (2021). Effect size measures for longitudinal growth analyses: Extending a framework of multilevel model R-squared 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

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 & Statistical Psychology*, *73*(S1), 194–211. https://doi.org/10.1111/bmsp.12194

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., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling* (2nd ed.). Sage.

Snijders, T. A. B., & Berkhof, J. (2008). Diagnostic checks for multilevel models. In J. de Leeuw & E. Meijer (Eds.), *Handbook of multilevel analysis* (pp. 141–175). Springer.

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

Appendix A

Derivation of Variance Distortion Caused by Conflation in Random Intercept Models

In this appendix, I mathematically compare the level-1 and level-2 error terms from the *unconflated-x model* in Equation 1 and the *conflated-x model* in Equation 2, and use this to, in turn, show how the *conflated-x model* distorts both the level-1 and level-2 error variances. All terms herein are defined in the article text. Here I assume all slopes are fixed, but provide subsequent derivations for random slope models in Appendix C.

I start first with the comparison of level-2 errors, and preliminarily note that, in the population, the model-implied cluster means of the outcome for the *unconflated-x model* are:

$$\begin{aligned} y_{\bullet j} &= E_{i|j}[y_{ij}] \\ &= E_{i|j}[\gamma_{00} + \gamma_b x_{\bullet j} + u_{0j} + \gamma_w(x_{ij} - x_{\bullet j}) + e_{ij}] \\ &= E_{i|j}[\gamma_{00}] + E_{i|j}[\gamma_b x_{\bullet j}] + E_{i|j}[u_{0j}] + E_{i|j}[\gamma_w(x_{ij} - x_{\bullet j})] + E_{i|j}[e_{ij}] \\ &= \gamma_{00} + \gamma_b x_{\bullet j} + u_{0j} \end{aligned} \tag{A1}$$

Note that the "$i \mid j$" subscript refers to taking the expectation across level-1 units (i.e., across i), holding level-2 unit constant (i.e., conditioning on j). Taking this same expectation for the *conflated-x model* yields:

$$\begin{aligned} y_{\bullet j} &= E_{i|j}[y_{ij}] \\ &= E_{i|j}[\gamma^*_{00} + \gamma_c x_{\bullet j} + u^*_{0j} + \gamma_c(x_{ij} - x_{\bullet j}) + e^*_{ij}] \\ &= E_{i|j}[\gamma^*_{00}] + E_{i|j}[\gamma_c x_{\bullet j}] + E_{i|j}[u^*_{0j}] + E_{i|j}[\gamma_c(x_{ij} - x_{\bullet j})] + E_{i|j}[e^*_{ij}] \\ &= \gamma^*_{00} + \gamma_c x_{\bullet j} + u^*_{0j} \end{aligned} \tag{A2}$$

Hence, the following equality holds in the population (i.e., assuming both models recover the underlying cluster means):

$$\gamma^*_{00} + \gamma_c x_{\bullet j} + u^*_{0j} = \gamma_{00} + \gamma_b x_{\bullet j} + u_{0j} \tag{A3}$$

Which implies that

$$\begin{aligned} u^*_{0j} &= u_{0j} + \gamma_{00} - \gamma^*_{00} + \gamma_b x_{\bullet j} - \gamma_c x_{\bullet j} \\ &= u_{0j} + \gamma_{00} - \gamma^*_{00} + (\gamma_b - \gamma_c) x_{\bullet j} \end{aligned} \tag{A4}$$

Which thus yields the following expression for the level-2 error variance for the *conflated-x model*:

(*Appendices continue*)

$$\begin{aligned}\tau_{00}^{*} &= \mathrm{var}(u_{0j}^{*})\\ &= \mathrm{var}(u_{0j} + \gamma_{00} - \gamma_{00}^{*} + (\gamma_b - \gamma_c)x_{\bullet j})\\ &= \mathrm{var}(u_{0j} + (\gamma_b - \gamma_c)x_{\bullet j})\\ &= \mathrm{var}(u_{0j}) + \mathrm{var}((\gamma_b - \gamma_c)x_{\bullet j}) + 2\mathrm{cov}(u_{0j}, (\gamma_b - \gamma_c)x_{\bullet j})\\ &= \tau_{00} + (\gamma_b - \gamma_c)^2 x_{\bullet j} + 0\\ &= \tau_{00} + (\gamma_b - \gamma_c)^2 x_{\bullet j}\end{aligned} \tag{A5}$$

Note that, across observations, $\gamma_{00} - \gamma_{00}^{*}$ is a constant value, and hence can be removed from the variance expression.

I next consider the level-1 errors, and note that the observation-level outcome can be defined for the *unconflated-x model* as Equation 1, and for the *conflated-x model* as Equation 2, implying the following to hold in the population:

$$\begin{aligned}&\gamma_{00}^{*} + u_{0j}^{*} + \gamma_c x_{\bullet j} + \gamma_c(x_{ij} - x_{\bullet j}) + e_{ij}^{*}\\ &\quad = \gamma_{00} + \gamma_b x_{\bullet j} + u_{0j} + \gamma_w(x_{ij} - x_{\bullet j}) + e_{ij}\end{aligned} \tag{A6}$$

Which implies that

$$\begin{aligned}e_{ij}^{*} &= e_{ij} + \gamma_{00} - \gamma_{00}^{*} + \gamma_b x_{\bullet j} - \gamma_c x_{\bullet j} + u_{0j} - u_{0j}^{*} + \gamma_w(x_{ij} - x_{\bullet j}) - \gamma_c(x_{ij} - x_{\bullet j})\\ &= e_{ij} + \gamma_{00} - \gamma_{00}^{*} + \gamma_b x_{\bullet j} - \gamma_c x_{\bullet j} + u_{0j} - (u_{0j} + \gamma_{00} - \gamma_{00}^{*} + (\gamma_b - \gamma_c)x_{\bullet j})\\ &\quad + \gamma_w(x_{ij} - x_{\bullet j}) - \gamma_c(x_{ij} - x_{\bullet j})\\ &= e_{ij} + \gamma_{00} - \gamma_{00}^{*} + (\gamma_b - \gamma_c)x_{\bullet j} + u_{0j}\\ &\quad - (u_{0j} + \gamma_{00} - \gamma_{00}^{*} + (\gamma_b - \gamma_c)x_{\bullet j}) + (\gamma_w - \gamma_c)(x_{ij} - x_{\bullet j})\\ &= e_{ij} + (\gamma_w - \gamma_c)(x_{ij} - x_{\bullet j})\end{aligned} \tag{A7}$$

Which thus yields the following expression for the level-1 error variance for the *conflated-x model*:

$$\begin{aligned}\sigma^{2*} &= \mathrm{var}(e_{ij}^{*})\\ &= \mathrm{var}(e_{ij} + (\gamma_w - \gamma_c)(x_{ij} - x_{\bullet j}))\\ &= \mathrm{var}(e_{ij}) + \mathrm{var}((\gamma_w - \gamma_c)(x_{ij} - x_{\bullet j})) + 2\mathrm{cov}(e_{ij}, (\gamma_w - \gamma_c)(x_{ij} - x_{\bullet j}))\\ &= \sigma^2 + (\gamma_w - \gamma_c)^2 \mathrm{var}(x_{ij} - x_{\bullet j}) + 0\\ &= \sigma^2 + (\gamma_w - \gamma_c)^2 \mathrm{var}(x_{ij} - x_{\bullet j})\end{aligned} \tag{A8}$$

Appendix B

Distortion in R-Squared Measures Induced by Conflating Level-Specific Effects

Here, I show mathematically how conflation can distort several popular R-squared measures for MLM, focusing on those that quantify variance attributable to the fixed components of the model. In each section, I provide the formulas for both the *unconflated-x model* in Equation 1 and the *conflated-x model* in Equation 2, and take the difference of these expressions to show the distortion induced by conflating level-specific effects.

Snijders and Bosker (2012) Total R-Squared

This measure was defined for the *unconflated-x model* in Table 4, and in the population can be further written as:

$$\begin{aligned}R_{SB}^2 &= 1 - \frac{\tau_{00} + \sigma^2}{\tau_{00,null} + \sigma_{null}^2}\\ &= 1 - \frac{\tau_{00} + \sigma^2}{\tau_{00} + \gamma_b^2\mathrm{var}(x_{\bullet j}) + \sigma^2 + \gamma_w^2\mathrm{var}(x_{ij} - x_{\bullet j})}\\ &= \frac{\tau_{00} + \gamma_b^2\mathrm{var}(x_{\bullet j}) + \sigma^2 + \gamma_w^2\mathrm{var}(x_{ij} - x_{\bullet j})}{\tau_{00} + \gamma_b^2\mathrm{var}(x_{\bullet j}) + \sigma^2 + \gamma_w^2\mathrm{var}(x_{ij} - x_{\bullet j})}\\ &\quad - \frac{\tau_{00} + \sigma^2}{\tau_{00} + \gamma_b^2\mathrm{var}(x_{\bullet j}) + \sigma^2 + \gamma_w^2\mathrm{var}(x_{ij} - x_{\bullet j})}\\ &= \frac{\gamma_b^2\mathrm{var}(x_{\bullet j}) + \gamma_w^2\mathrm{var}(x_{ij} - x_{\bullet j})}{\tau_{00} + \gamma_b^2\mathrm{var}(x_{\bullet j}) + \sigma^2 + \gamma_w^2\mathrm{var}(x_{ij} - x_{\bullet j})}\end{aligned} \tag{B1}$$

For the *conflated-x model* this measure can be written in the population as

$$\begin{aligned}R_{SB}^{2*} &= 1 - \frac{\tau_{00}^{*} + \sigma^{2*}}{\tau_{00,null} + \sigma_{null}^2}\\ &= 1 - \frac{\tau_{00} + \sigma^2 + (\gamma_b - \gamma_c)^2\mathrm{var}(x_{\bullet j}) + (\gamma_w - \gamma_c)^2\mathrm{var}(x_{ij} - x_{\bullet j})}{\tau_{00} + \gamma_b^2\mathrm{var}(x_{\bullet j}) + \sigma^2 + \gamma_w^2\mathrm{var}(x_{ij} - x_{\bullet j})}\\ &= 1 - \frac{\tau_{00} + \sigma^2}{\tau_{00} + \gamma_b^2\mathrm{var}(x_{\bullet j}) + \sigma^2 + \gamma_w^2\mathrm{var}(x_{ij} - x_{\bullet j})}\\ &\quad - \left(\frac{(\gamma_b - \gamma_c)^2\mathrm{var}(x_{\bullet j}) + (\gamma_w - \gamma_c)^2\mathrm{var}(x_{ij} - x_{\bullet j})}{\tau_{00} + \gamma_b^2\mathrm{var}(x_{\bullet j}) + \sigma^2 + \gamma_w^2\mathrm{var}(x_{ij} - x_{\bullet j})}\right)\\ &= R_{SB}^2 - \left(\frac{(\gamma_b - \gamma_c)^2\mathrm{var}(x_{\bullet j}) + (\gamma_w - \gamma_c)^2\mathrm{var}(x_{ij} - x_{\bullet j})}{\tau_{00} + \gamma_b^2\mathrm{var}(x_{\bullet j}) + \sigma^2 + \gamma_w^2\mathrm{var}(x_{ij} - x_{\bullet j})}\right)\end{aligned} \tag{B2}$$

Hence, the expression in the paratheses of Equation B2 represents the distortion induced by conflating. This expression in paratheses is always going to be positive, implying that conflating will cause this measure to be systematically too small.

Here, I additionally consider conditions that would yield a negative value for this measure in the population:

(*Appendices continue*)

$$\begin{aligned}
&R_{SB}^{2*} < 0\\
&\Rightarrow 1 - \frac{\tau_{00} + \sigma^2 + (\gamma_b - \gamma_c)^2 \mathrm{var}(x_{\bullet j}) + (\gamma_w - \gamma_c)^2 \mathrm{var}(x_{ij} - x_{\bullet j})}{\tau_{00} + \gamma_b^2 \mathrm{var}(x_{\bullet j}) + \sigma^2 + \gamma_w^2 \mathrm{var}(x_{ij} - x_{\bullet j})} < 0\\
&\Rightarrow 1 < \frac{\tau_{00} + \sigma^2 + (\gamma_b - \gamma_c)^2 \mathrm{var}(x_{\bullet j}) + (\gamma_w - \gamma_c)^2 \mathrm{var}(x_{ij} - x_{\bullet j})}{\tau_{00} + \gamma_b^2 \mathrm{var}(x_{\bullet j}) + \sigma^2 + \gamma_w^2 \mathrm{var}(x_{ij} - x_{\bullet j})}\\
&\Rightarrow \tau_{00} + \gamma_b^2 \mathrm{var}(x_{\bullet j}) + \sigma^2 + \gamma_w^2 \mathrm{var}(x_{ij} - x_{\bullet j}) < \tau_{00} + \sigma^2\\
&\quad + (\gamma_b - \gamma_c)^2 \mathrm{var}(x_{\bullet j}) + (\gamma_w - \gamma_c)^2 \mathrm{var}(x_{ij} - x_{\bullet j})\\
&\Rightarrow \gamma_b^2 \mathrm{var}(x_{\bullet j}) + \gamma_w^2 \mathrm{var}(x_{ij} - x_{\bullet j}) < (\gamma_b - \gamma_c)^2 \mathrm{var}(x_{\bullet j})\\
&\quad + (\gamma_w - \gamma_c)^2 \mathrm{var}(x_{ij} - x_{\bullet j})\\
&\Rightarrow \gamma_b^2 \mathrm{var}(x_{\bullet j}) - (\gamma_b - \gamma_c)^2 \mathrm{var}(x_{\bullet j}) + \gamma_w^2 \mathrm{var}(x_{ij} - x_{\bullet j})\\
&\quad - (\gamma_w - \gamma_c)^2 \mathrm{var}(x_{ij} - x_{\bullet j}) < 0\\
&\Rightarrow (\gamma_b^2 - (\gamma_b - \gamma_c)^2) \mathrm{var}(x_{\bullet j})\\
&\quad + (\gamma_w^2 - (\gamma_w - \gamma_c)^2) \mathrm{var}(x_{ij} - x_{\bullet j}) < 0\\
&\Rightarrow (\gamma_b^2 - (\gamma_b - \gamma_c)^2) \frac{\mathrm{var}(x_{\bullet j})}{\mathrm{var}(x_{ij})}\\
&\quad + (\gamma_w^2 - (\gamma_w - \gamma_c)^2) \frac{\mathrm{var}(x_{ij} - x_{\bullet j})}{\mathrm{var}(x_{ij})} < 0\\
&\Rightarrow (\gamma_b^2 - (\gamma_b - \gamma_c)^2) ICC_x\\
&\quad + (\gamma_w^2 - (\gamma_w - \gamma_c)^2)(1 - ICC_x) < 0\\
&\Rightarrow (\gamma_b^2 - (\gamma_b - \gamma_c)^2) ICC_x <\\
&\quad - (\gamma_w^2 - (\gamma_w - \gamma_c)^2)(1 - ICC_x)\\
&\Rightarrow (\gamma_b^2 - (\gamma_b - \gamma_c)^2) ICC_x < (\gamma_w^2 - (\gamma_w - \gamma_c)^2)(ICC_x - 1)
\end{aligned} \tag{B3}$$

Thus, if $(\gamma_b^2 - (\gamma_b - \gamma_c)^2) ICC_x$ is less than $(\gamma_w^2 - (\gamma_w - \gamma_c)^2)(ICC_x - 1)$ (where ICC_x is the ratio of $\mathrm{var}(x_{\bullet j})$ to $\mathrm{var}(x_{ij})$), then R_{SB}^{2*} will be negative. Note also that this is guaranteed whenever both $\gamma_b^2 < (\gamma_b - \gamma_c)^2$ and $\gamma_w^2 < (\gamma_w - \gamma_c)^2$.

Rights and Sterba (2019) Total R-Squared Measures

The first total Rights and Sterba (2019) measure, $R_t^{2(f_1)}$, was defined in the population in Table 4 for the *unconflated-x model*, and can be defined in the population for the *conflated-x model* as:

$$\begin{aligned}
R_t^{2(f_1)*} &= \frac{\gamma_c^2 \mathrm{var}(x_{ij} - x_{\bullet j})}{\gamma_c^2 \mathrm{var}(x_{ij} - x_{\bullet j}) + \gamma_c^2 \mathrm{var}(x_{\bullet j}) + \tau_{00}^* + \sigma^{2*}}\\
&= \frac{\gamma_c^2 \mathrm{var}(x_{ij} - x_{\bullet j})}{\gamma_c^2 \mathrm{var}(x_{ij} - x_{\bullet j}) + \gamma_c^2 \mathrm{var}(x_{\bullet j}) + \tau_{00} + (\gamma_b - \gamma_c)^2 \mathrm{var}(x_{\bullet j}) + \sigma^2 + (\gamma_w - \gamma_c)^2 \mathrm{var}(x_{ij} - x_{\bullet j})}
\end{aligned} \tag{B4}$$

Similarly, $R_t^{2(f_2)}$ was defined in the population in Table 4 for the *unconflated-x model* and can be defined for the *conflated-x model* as:

$$\begin{aligned}
R_t^{2(f_2)*} &= \frac{\gamma_c^2 \mathrm{var}(x_{\bullet j})}{\gamma_c^2 \mathrm{var}(x_{ij} - x_{\bullet j}) + \gamma_c^2 \mathrm{var}(x_{\bullet j}) + \tau_{00}^* + \sigma^{2*}}\\
&= \frac{\gamma_c^2 \mathrm{var}(x_{\bullet j})}{\gamma_c^2 \mathrm{var}(x_{ij} - x_{\bullet j}) + \gamma_c^2 \mathrm{var}(x_{\bullet j}) + \tau_{00} + (\gamma_b - \gamma_c)^2 \mathrm{var}(x_{\bullet j}) + \sigma^2 + (\gamma_w - \gamma_c)^2 \mathrm{var}(x_{ij} - x_{\bullet j})}
\end{aligned} \tag{B5}$$

Note that, given the potential distortion in both the numerator and denominator, conflation can cause these measures to be either systematically too small or too large (as demonstrated via simulation).

Additionally, under conflation, the degree to which x explains variance via fixed components at the within-cluster versus at the between-cluster level will be driven exclusively by the degree of within-cluster versus between-cluster variation in x. Hence, the difference between $R_t^{2(f_1)*}$ and $R_t^{2(f_2)*}$ will not at all reflect the underlying strength of the within-cluster and between-cluster slopes, γ_w and γ_b. I can show this mathematically by computing the ratio of $R_t^{2(f_1)*}$ to $R_t^{2(f_2)*}$ (with the assumption that neither are 0):

$$\begin{aligned}
\frac{R_t^{2(f_1)*}}{R_t^{2(f_2)*}} &= \frac{\gamma_c^2 \mathrm{var}(x_{ij} - x_{\bullet j})}{\gamma_c^2 \mathrm{var}(x_{ij} - x_{\bullet j}) + \gamma_c^2 \mathrm{var}(x_{\bullet j}) + \tau_{00}^* + \sigma^{2*}} \Big/ \frac{\gamma_c^2 \mathrm{var}(x_{\bullet j})}{\gamma_c^2 \mathrm{var}(x_{ij} - x_{\bullet j}) + \gamma_c^2 \mathrm{var}(x_{\bullet j}) + \tau_{00}^* + \sigma^{2*}}\\
&= \frac{\gamma_c^2 \mathrm{var}(x_{ij} - x_{\bullet j})}{\gamma_c^2 \mathrm{var}(x_{\bullet j})}\\
&= \frac{\mathrm{var}(x_{ij} - x_{\bullet j})}{\mathrm{var}(x_{\bullet j})}
\end{aligned} \tag{B6}$$

Hence, assuming $\mathrm{var}(x_{ij} - x_{\bullet j})$, $\mathrm{var}(x_{\bullet j})$, and γ_c are nonzero: (a) if $\mathrm{var}(x_{ij} - x_{\bullet j}) > \mathrm{var}(x_{\bullet j})$, then $R_t^{2(f_1)*} > R_t^{2(f_2)*}$; (b) if $\mathrm{var}(x_{ij} - x_{\bullet j}) < \mathrm{var}(x_{\bullet j})$, then $R_t^{2(f_1)*} < R_t^{2(f_2)*}$; and (c) if $\mathrm{var}(x_{ij} - x_{\bullet j}) = \mathrm{var}(x_{\bullet j})$, then $R_t^{2(f_1)*} = R_t^{2(f_2)*}$. In contrast, for the *unconflated-x model*:

$$\begin{aligned}
\frac{R_t^{2(f_1)}}{R_t^{2(f_2)}} &= \frac{\gamma_w^2 \mathrm{var}(x_{ij} - x_{\bullet j})}{\gamma_w^2 \mathrm{var}(x_{ij} - x_{\bullet j}) + \gamma_b^2 \mathrm{var}(x_{\bullet j}) + \tau_{00} + \sigma^2} \Big/ \frac{\gamma_b^2 \mathrm{var}(x_{\bullet j})}{\gamma_w^2 \mathrm{var}(x_{ij} - x_{\bullet j}) + \gamma_b^2 \mathrm{var}(x_{\bullet j}) + \tau_{00} + \sigma^2}\\
&= \frac{\gamma_w^2 \mathrm{var}(x_{ij} - x_{\bullet j})}{\gamma_b^2 \mathrm{var}(x_{\bullet j})}
\end{aligned} \tag{B7}$$

Hence, the relative magnitude of $R_t^{2(f_1)}$ versus $R_t^{2(f_2)}$ will be based on both the variance of x at the within-cluster versus between-cluster levels as well as the strength of the underlying within-cluster versus between-cluster effects, γ_w and γ_b.

Raudenbush and Bryk (2002) Between-Cluster R-Squared

This measure was defined for the *unconflated-x model* in Table 4, and in the population can be further written as:

(*Appendices continue*)

$$
\begin{aligned}
R^2_{L2} &= \frac{\tau_{00,null} - \tau_{00}}{\tau_{00,null}} \\
&= \frac{\tau_{00} + \gamma_b^2 \text{var}(x_{\bullet j}) - \tau_{00}}{\tau_{00} + \gamma_b^2 \text{var}(x_{\bullet j})} \\
&= \frac{\gamma_b^2 \text{var}(x_{\bullet j})}{\tau_{00} + \gamma_b^2 \text{var}(x_{\bullet j})}
\end{aligned} \tag{B8}
$$

For the *conflated-x model* this measure can be written in the population as:

$$
\begin{aligned}
R^{2*}_{L2} &= \frac{\tau_{00,null} - \tau^*_{00}}{\tau_{00,null}} \\
&= \frac{(\tau_{00} + \gamma_b^2 \text{var}(x_{\bullet j})) - (\tau_{00} + (\gamma_b - \gamma_c)^2 \text{var}(x_{\bullet j}))}{\tau_{00} + \gamma_b^2 \text{var}(x_{\bullet j})} \\
&= \frac{\gamma_b^2 \text{var}(x_{\bullet j}) - (\gamma_b - \gamma_c)^2 \text{var}(x_{\bullet j})}{\tau_{00} + \gamma_b^2 \text{var}(x_{\bullet j})} \\
&= \frac{\gamma_b^2 \text{var}(x_{\bullet j})}{\tau_{00} + \gamma_b^2 \text{var}(x_{\bullet j})} - \left(\frac{(\gamma_b - \gamma_c)^2 \text{var}(x_{\bullet j})}{\tau_{00} + \gamma_b^2 \text{var}(x_{\bullet j})}\right) \\
&= R^2_{L2} - \left(\frac{(\gamma_b - \gamma_c)^2 \text{var}(x_{\bullet j})}{\tau_{00} + \gamma_b^2 \text{var}(x_{\bullet j})}\right)
\end{aligned} \tag{B9}
$$

The expression in the paratheses of Equation B9 will always be positive, implying that conflation will cause this measure to be systematically too small. In terms of when this measure will be negative:

$$
\begin{aligned}
&R^{2*}_{L2} < 0 \\
&\Rightarrow \frac{\gamma_b^2 \text{var}(x_{\bullet j}) - (\gamma_b - \gamma_c)^2 \text{var}(x_{\bullet j})}{\tau_{00} + \gamma_b^2 \text{var}(x_{\bullet j})} < 0 \\
&\Rightarrow \gamma_b^2 \text{var}(x_{\bullet j}) - (\gamma_b - \gamma_c)^2 \text{var}(x_{\bullet j}) < 0 \\
&\Rightarrow \gamma_b^2 \text{var}(x_{\bullet j}) < (\gamma_b - \gamma_c)^2 \text{var}(x_{\bullet j}) \\
&\Rightarrow \gamma_b^2 < (\gamma_b - \gamma_c)^2
\end{aligned} \tag{B10}
$$

Hence, R^{2*}_{L2} is negative whenever γ_b^2 is less than $(\gamma_b - \gamma_c)^2$.

Rights and Sterba (2019) Between-Cluster R-Squared

The between-cluster Rights and Sterba (2019) measure, $R_b^{2(f_2)}$, was defined in the population in Table 4 for the *unconflated-x model*. This measure can be expressed in the population for the *conflated-x model* as:

$$
\begin{aligned}
R_b^{2(f_2)*} &= \frac{\gamma_c^2 \text{var}(x_{\bullet j})}{\gamma_c^2 \text{var}(x_{\bullet j}) + \tau^*_{00}} \\
&= \frac{\gamma_c^2 \text{var}(x_{\bullet j})}{\gamma_c^2 \text{var}(x_{\bullet j}) + \tau_{00} + (\gamma_b - \gamma_c)^2 \text{var}(x_{\bullet j})}
\end{aligned} \tag{B11}
$$

Given the potential distortion in both the numerator and denominator, conflating can cause this measure to be either systematically too small or too large (as demonstrated via simulation).

Raudenbush and Bryk (2002) Within-Cluster R-Squared

This measure was defined for the *unconflated-x model* in Table 4, and in the population can be further written as:

$$
\begin{aligned}
R^2_{L1} &= \frac{\sigma^2_{null} - \sigma^2}{\sigma^2_{null}} \\
&= \frac{(\sigma^2 + \gamma_w^2 \text{var}(x_{ij} - x_{\bullet j})) - \sigma^2}{\sigma^2 + \gamma_w^2 \text{var}(x_{ij} - x_{\bullet j})} \\
&= \frac{\gamma_w^2 \text{var}(x_{ij} - x_{\bullet j})}{\sigma^2 + \gamma_w^2 \text{var}(x_{ij} - x_{\bullet j})}
\end{aligned} \tag{B12}
$$

For the *conflated-x model* this measure can be written in the population as

$$
\begin{aligned}
R^{2*}_{L1} &= \frac{\sigma^2_{null} - \sigma^{2*}}{\sigma^2_{null}} \\
&= \frac{\sigma^2 + \gamma_w^2 \text{var}(x_{ij} - x_{\bullet j}) - (\sigma^2 + (\gamma_w - \gamma_c)^2 \text{var}(x_{ij} - x_{\bullet j}))}{\sigma^2 + \gamma_w^2 \text{var}(x_{ij} - x_{\bullet j})} \\
&= \frac{\gamma_w^2 \text{var}(x_{ij} - x_{\bullet j}) - (\gamma_w - \gamma_c)^2 \text{var}(x_{ij} - x_{\bullet j})}{\sigma^2 + \gamma_w^2 \text{var}(x_{ij} - x_{\bullet j})} \\
&= \left(\frac{\gamma_w^2 \text{var}(x_{ij} - x_{\bullet j})}{\sigma^2 + \gamma_w^2 \text{var}(x_{ij} - x_{\bullet j})}\right) - \left(\frac{(\gamma_w - \gamma_c)^2 \text{var}(x_{ij} - x_{\bullet j}))}{\sigma^2 + \gamma_w^2 \text{var}(x_{ij} - x_{\bullet j})}\right) \\
&= R^2_{L1} - \left(\frac{(\gamma_w - \gamma_c)^2 \text{var}(x_{ij} - x_{\bullet j}))}{\sigma^2 + \gamma_w^2 \text{var}(x_{ij} - x_{\bullet j})}\right)
\end{aligned} \tag{B13}
$$

The expression in the paratheses of Equation B13 will always be positive, implying that conflation can cause this measure to be systematically too small. In terms of when this measure will be negative:

$$
\begin{aligned}
&R^{2*}_{L1} < 0 \\
&\Rightarrow \frac{\gamma_w^2 \text{var}(x_{ij} - x_{\bullet j}) - (\gamma_w - \gamma_c)^2 \text{var}(x_{ij} - x_{\bullet j})}{\sigma^2 + \gamma_w^2 \text{var}(x_{ij} - x_{\bullet j})} < 0 \\
&\Rightarrow \gamma_w^2 \text{var}(x_{ij} - x_{\bullet j}) - (\gamma_w - \gamma_c)^2 \text{var}(x_{ij} - x_{\bullet j}) < 0 \\
&\Rightarrow \gamma_w^2 \text{var}(x_{ij} - x_{\bullet j}) < (\gamma_w - \gamma_c)^2 \text{var}(x_{ij} - x_{\bullet j}) \\
&\Rightarrow \gamma_w^2 < (\gamma_w - \gamma_c)^2
\end{aligned} \tag{B14}
$$

Hence, R^{2*}_{L1} is negative whenever γ_w^2 is less than $(\gamma_w - \gamma_c)^2$.

(*Appendices continue*)

Rights and Sterba (2019) Within-Cluster R-Squared

The within-cluster Rights and Sterba (2019) measure, $R_w^{2(f_1)}$, was defined in Table 4 for the *unconflated-x model* and can be written in the population for the *conflated-x model* as:

$$\begin{aligned} R_w^{2(f_1)*} &= \frac{\gamma_c^2 \text{var}(x_{ij} - x_{\bullet j})}{\gamma_c^2 \text{var}(x_{ij} - x_{\bullet j}) + \sigma^{2*}} \\ &= \frac{\gamma_c^2 \text{var}(x_{ij} - x_{\bullet j})}{\gamma_c^2 \text{var}(x_{ij} - x_{\bullet j}) + \sigma^2 + (\gamma_w - \gamma_c)^2 \text{var}(x_{ij} - x_{\bullet j})} \end{aligned} \tag{B15}$$

Given the potential distortion in both the numerator and demonstration, conflation can cause this measure to be either systematically too small or too large (as demonstrated via simulation).

Appendix C

Derivation of Variance Distortion Caused by Both Fixed and Random Conflation in Random Slope Models

Here, I mathematically compare the level-1 errors terms and the level-2 random intercept error terms from the *heteroscedastic unconflated-x model* (Equation 20) and the *fully conflated-x model* (Equation 19). This expands upon the derivation in Appendix A by adding random slopes, and shows how these error terms can be distorted not only as a function of *fixed* conflation (i.e., constraining the fixed components associated with $x_{ij} - x_{\bullet j}$ and $x_{\bullet j}$ to equality) but also *random* conflation (i.e., constraining the random components associated $x_{ij} - x_{\bullet j}$ and $x_{\bullet j}$ to equality; Rights & Sterba, 2020). To ensure comparability between the models, I will assume all predictors are centered such that they have a mean of 0, and hence for both models the random intercept variance can, in theory, be interpreted as the between-cluster variance that is not accounted for × predictors × either fixed or random effects (Rights & Sterba, 2021).

The model-implied cluster-means of the outcome for the *heteroscedastic unconflated-x model* is

$$\begin{aligned} y_{\bullet j} &= E_{i|j}[y_{ij}] \\ &= E_{i|j}[\gamma_{00} + \gamma_b x_{\bullet j} + u_{0j} + \gamma_w (x_{ij} - x_{\bullet j}) + u_{wj}(x_{ij} - x_{\bullet j}) + u_{bj} x_{\bullet j} + e_{ij}] \\ &= E_{i|j}[\gamma_{00}] + E_{i|j}[\gamma_b x_{\bullet j}] + E_{i|j}[u_{0j}] + E_{i|j}[\gamma_w (x_{ij} - x_{\bullet j})] \\ &\quad + E_{i|j}[u_{wj}(x_{ij} - x_{\bullet j})] + E_{i|j}[u_{bj} x_{\bullet j}] + E_{i|j}[e_{ij}] \\ &= \gamma_{00} + \gamma_b x_{\bullet j} + u_{bj} x_{\bullet j} + u_{0j} \end{aligned} \tag{C1}$$

And those of the *fully conflated-x model* is

$$\begin{aligned} y_{\bullet j} &= E_{i|j}[y_{ij}] \\ &= E_{i|j}[\gamma_{00}^* + \gamma_c x_{\bullet j} + u_{cj} x_{\bullet j} + u_{0j}^* + \gamma_c (x_{ij} - x_{\bullet j}) + u_{cj}(x_{ij} - x_{\bullet j}) + e_{ij}^*] \\ &= E_{i|j}[\gamma_{00}^*] + E_{i|j}[\gamma_c x_{\bullet j}] + E_{i|j}[u_{cj} x_{\bullet j}] + E_{i|j}[u_{0j}^*] + E_{i|j}[\gamma_c (x_{ij} - x_{\bullet j})] \\ &\quad + E_{i|j}[u_{cj}(x_{ij} - x_{\bullet j})] + E_{i|j}[e_{ij}^*] \\ &= \gamma_{00}^* + \gamma_c x_{\bullet j} + u_{cj} x_{\bullet j} + u_{0j}^* \end{aligned} \tag{C2}$$

Hence, if both models recover the population cluster means, the following equality holds in the population:

$$\gamma_{00}^* + \gamma_c x_{\bullet j} + u_{cj} x_{\bullet j} + u_{0j}^* = \gamma_{00} + \gamma_b x_{\bullet j} + u_{bj} x_{\bullet j} + u_{0j} \tag{C3}$$

Which implies that

$$\begin{aligned} u_{0j}^* &= \gamma_{00} - \gamma_{00}^* + \gamma_b x_{\bullet j} - \gamma_c x_{\bullet j} + u_{0j} + u_{bj} x_{\bullet j} - u_{cj} x_{\bullet j} \\ &= \gamma_{00} - \gamma_{00}^* + (\gamma_b - \gamma_c) x_{\bullet j} + u_{0j} + (u_{bj} - u_{cj}) x_{\bullet j} \end{aligned} \tag{C4}$$

Which yields the following expression for the level-2 error variance for the *fully conflated-x model*:

$$\begin{aligned} \tau_{00}^* &= \text{var}(u_{0j}^*) \\ &= \text{var}(\gamma_{00} - \gamma_{00}^* + (\gamma_b - \gamma_c) x_{\bullet j} + u_{0j} + (u_{bj} - u_{cj}) x_{\bullet j}) \\ &= \text{var}(u_{0j} + (\gamma_b - \gamma_c) x_{\bullet j} + (u_{bj} - u_{cj}) x_{\bullet j}) \\ &= \text{var}(u_{0j}) + (\gamma_b - \gamma_c)^2 \text{var}(x_{\bullet j}) + \text{var}((u_{bj} - u_{cj}) x_{\bullet j}) \\ &\quad + 2(\gamma_b - \gamma_c)\text{cov}(u_{0j}, x_{\bullet j}) + 2\text{cov}(u_{0j}, (u_{bj} - u_{cj}) x_{\bullet j}) \\ &\quad + 2(\gamma_b - \gamma_c)\text{cov}(x_{\bullet j}, (u_{bj} - u_{cj}) x_{\bullet j}) \\ &= \text{var}(u_{0j}) + (\gamma_b - \gamma_c)^2 \text{var}(x_{\bullet j}) + \text{var}((u_{bj} - u_{cj}) x_{\bullet j}) \\ &\quad + 2\text{cov}(u_{0j}, (u_{bj} - u_{cj}) x_{\bullet j}) \\ &\quad + 2(\gamma_b - \gamma_c)\text{cov}(x_{\bullet j}, (u_{bj} - u_{cj}) x_{\bullet j}) \end{aligned} \tag{C5}$$

For the level-1 errors, note first that the observation-level outcome is defined for the *heteroscedastic unconflated-x model* as Equation 19, and for the *fully conflated-x model* as Equation 20, implying the following to hold if both models recover y_{ij}:

$$\begin{aligned} &\gamma_{00}^* + u_{0j}^* + \gamma_c (x_{ij} - x_{\bullet j}) + \gamma_c x_{\bullet j} + u_{cj}(x_{ij} - x_{\bullet j}) + u_{cj} x_{\bullet j} + e_{ij}^* \\ &\quad = \gamma_{00} + u_{0j} + \gamma_w (x_{ij} - x_{\bullet j}) + \gamma_b x_{\bullet j} + u_{wj}(x_{ij} - x_{\bullet j}) + u_{bj} x_{\bullet j} + e_{ij} \end{aligned} \tag{C6}$$

Which implies that

$$\begin{aligned} e_{ij}^* &= e_{ij} + \gamma_{00} - \gamma_{00}^* + u_{0j} - u_{0j}^* + \gamma_w (x_{ij} - x_{\bullet j}) - \gamma_c (x_{ij} - x_{\bullet j}) \\ &\quad + \gamma_b x_{\bullet j} - \gamma_c x_{\bullet j} + u_{bj} x_{\bullet j} - u_{cj} x_{\bullet j} + u_{wj}(x_{ij} - x_{\bullet j}) - u_{cj}(x_{ij} - x_{\bullet j}) \\ &= e_{ij} + \gamma_{00} - \gamma_{00}^* + u_{0j} \\ &\quad - (\gamma_{00} - \gamma_{00}^* + (\gamma_b - \gamma_c) x_{\bullet j} + u_{0j} + (u_{bj} - u_{cj}) x_{\bullet j}) \\ &\quad + (\gamma_w - \gamma_c)(x_{ij} - x_{\bullet j}) + (\gamma_b - \gamma_c) x_{\bullet j} + (u_{bj} - u_{cj}) x_{\bullet j} \\ &\quad + (u_{wj} - u_{cj})(x_{ij} - x_{\bullet j}) \\ &= e_{ij} + (\gamma_w - \gamma_c)(x_{ij} - x_{\bullet j}) + (u_{wj} - u_{cj})(x_{ij} - x_{\bullet j}) \end{aligned} \tag{C7}$$

(*Appendices continue*)

Which thus yields the following expression for the level-1 error variance for the *fully conflated-x model*:

$$
\begin{aligned}
\sigma^{2*} &= \mathrm{var}(e^{*}_{ij}) \\
&= \mathrm{var}(e_{ij} + (\gamma_w - \gamma_c)(x_{ij} - x_{\bullet j}) + (u_{wj} - u_{cj})(x_{ij} - x_{\bullet j})) \\
&= \mathrm{var}(e_{ij}) + \mathrm{var}((\gamma_w - \gamma_c)(x_{ij} - x_{\bullet j})) + \mathrm{var}((u_{wj} - u_{cj})(x_{ij} - x_{\bullet j})) \\
&\quad + 2(\gamma_w - \gamma_c)\mathrm{cov}(e_{ij}, (x_{ij} - x_{\bullet j})) \\
&\quad + 2\mathrm{cov}(e_{ij}, (u_{wj} - u_{cj})(x_{ij} - x_{\bullet j})) \\
&\quad + 2(\gamma_w - \gamma_c)\mathrm{cov}((x_{ij} - x_{\bullet j}), (u_{wj} - u_{cj})(x_{ij} - x_{\bullet j})) \\
&= \sigma^2 + (\gamma_w - \gamma_c)^2\mathrm{var}(x_{ij} - x_{\bullet j}) + \mathrm{var}(u_{wj} - u_{cj})\mathrm{var}(x_{ij} - x_{\bullet j}) \\
&\quad + \mathrm{var}(u_{wj} - u_{cj})E[x_{ij} - x_{\bullet j}]^2 \\
&\quad + \mathrm{var}(x_{ij} - x_{\bullet j})E[u_{wj} - u_{cj}]^2 \\
&\quad + 2(\gamma_w - \gamma_c)(E[(x_{ij} - x_{\bullet j})^2(u_{wj} - u_{cj})] \\
&\quad - E[x_{ij} - x_{\bullet j}]E[(x_{ij} - x_{\bullet j})(u_{wj} - u_{cj})]) \\
&= \sigma^2 + (\gamma_w - \gamma_c)^2\mathrm{var}(x_{ij} - x_{\bullet j}) + \mathrm{var}(u_{wj} - u_{cj})\mathrm{var}(x_{ij} - x_{\bullet j}) \\
&\quad + 2(\gamma_w - \gamma_c)(E[(x_{ij} - x_{\bullet j})^2]E[u_{wj} - u_{cj}]) \\
&= \sigma^2 + (\gamma_w - \gamma_c)^2\mathrm{var}(x_{ij} - x_{\bullet j}) + \mathrm{var}(u_{wj} - u_{cj})\mathrm{var}(x_{ij} - x_{\bullet j})
\end{aligned}
\tag{C8}
$$

When there is only random (and not fixed) conflation, that is, in the *random-conflated contextual effect model* (Equation 21), these expressions simplify in that the γ terms drop out.

The expressions derived here are notably less clean than those derived for the fixed slope model (i.e., Equations 7 and 8). This is driven primarily by the fact that the slopes associated with x are no longer just fixed quantities, but instead random quantities that follow an assumed distributional form. It is not mathematically clear, for instance, the extent to which conflated random terms might be correlated with the unconflated random terms or the predictors, nor how well the conflated model would actually recover the population outcome means in finite samples. Hence, the specific patterns of distortion are primarily investigated via simulation (see article text).

Received January 7, 2021
Revision received March 29, 2022
Accepted May 23, 2022 ■