

An Expanded Decision-Making Procedure for Examining Cross-Level Interaction Effects With Multilevel Modeling

Organizational Research Methods
2015, Vol. 18(2) 155-176
© The Author(s) 2015
Reprints and permission:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/1094428114563618
orm.sagepub.com



Herman Aguinis¹ and Steven Andrew Culpepper²

Abstract

Cross-level interaction effects lay at the heart of multilevel contingency and interactionism theories. Also, practitioners are particularly interested in such effects because they provide information on the contextual conditions and processes under which interventions focused on individuals (e.g., selection, leadership training, performance appraisal, and management) result in more or less positive outcomes. We derive a new intraclass correlation, ρ_β , to assess the degree of lower-level outcome variance that is attributed to higher-level differences in slope coefficients. We provide analytical and empirical evidence that ρ_β is an index of variance that differs from the traditional intraclass correlation ρ_α and use data from recently published articles to illustrate that ρ_α assesses differences across collectives and higher-level processes (e.g., teams, leadership styles, reward systems) but ignores the variance attributed to differences in lower-level relationships (e.g., individual level job satisfaction and individual level performance). Because ρ_α and ρ_β provide information on two different sources of variability in the data structure (i.e., differences in means and differences in relationships, respectively), our results suggest that researchers contemplating the use of multilevel modeling, as well those who suspect non-independence in their data structure, should expand the decision criteria for using multilevel approaches to include both types of intraclass correlations. To facilitate this process, we offer an illustrative data set and the `icc beta` R package for computing ρ_β in single- and multiple-predictor situations and make them available through the Comprehensive R Archive Network (i.e., CRAN).

Keywords

cross-level analysis, interactions, measurement models, quantitative multilevel research

¹Department of Management and Entrepreneurship, Kelley School of Business, Indiana University, Bloomington, IN, USA

²Department of Statistics, University of Illinois at Urbana-Champaign, Champaign, IL, USA

Corresponding Author:

Herman Aguinis, Department of Management and Entrepreneurship, Kelley School of Business, Indiana University, 1309 E. 10th Street, Bloomington, IN 47405, USA.

Email: haguinis@indiana.edu

Researchers in organizational behavior, human resource management, entrepreneurship, strategy, and many other fields now explicitly recognize that lower-level entities are usually nested within higher-level collectives. For example, employees are nested within jobs (e.g., Taylor, Li, Shi, & Borman, 2008) and teams (e.g., Kim, Bhawe, & Glomb, 2013), establishments within companies (e.g., Takeuchi, Chen, & Lepak, 2009), and firms within industries (e.g., Short, Ketchen, Bennett, & du Toit, 2006). Similarly, a nested data structure exists in studies involving longitudinal or repeated measures designs in which the lower level refers to observations and the higher level to the units (e.g., entrepreneurs, teams, firms) about which data have been collected over time (e.g., Uy, Foo, & Aguinis, 2010).

Covariation between higher-level variables and lower-level outcomes leads to errors of prediction if a researcher uses statistical approaches such as ordinary least squares (OLS) regression, which are not designed to model data structures that include dependence due to clustering (Aguinis, Gottfredson, & Culpepper, 2013; Heck, Thomas, & Tabata, 2010; Hox, 2010; Raudenbush & Bryk, 2002; Snijders & Bosker, 2012). In other words, dependence is “not adequately represented by the probability model of multiple linear regression analysis” (Snijders & Bosker, 2012, p. 3) and “the effect is generally *not* negligible” (Hox, 2010, p. 5).

Multilevel modeling, also referred to as hierarchical linear modeling (HLM) (Raudenbush & Bryk, 2002), mixed-effect models (Cao & Ramsay, 2010), random coefficient modeling (Longford, 1993), and covariance components models (e.g., Searle, Casella, & McCulloch, 1992), allows researchers to explicitly incorporate and model bias in standard errors and statistical tests resulting from the dependence of observations that occurs in nested data structures (Kenny, Korchmaros, & Bolger, 2003). Moreover, multilevel modeling allows researchers to assess three types of relationships (Mathieu, Aguinis, Culpepper, & Chen, 2012). First, it allows for tests of lower-level direct effects: whether a lower-level predictor X (i.e., Level 1 or L1 predictor) has an effect on a lower-level outcome variable Y (i.e., L1 outcome). For example, there may be an interest in assessing whether individual job satisfaction predicts individual job performance. Second, it allows for tests of cross-level direct effects: whether a higher-level predictor W (i.e., Level 2 or L2 predictor) is related to a L1 outcome variable Y . For example, a researcher may want to test whether team cohesion (an L2 variable) predicts individual job performance (an L1 outcome). Third, it allows for tests of cross-level interaction effects: whether the nature and/or strength of the relationship between two lower-level variables (e.g., L1 predictor X and L1 outcome Y) change as a function of a higher-level variable W . For example, a researcher may be interested in testing the hypothesis that the relationship between individual job satisfaction and individual performance may vary as a function of (i.e., is moderated by) the degree of team cohesion such that the relationship will be stronger for highly cohesive compared to less cohesive teams.

One of the three types of effects mentioned previously, cross-level interactions, is at the heart of modern-day contingency theories, person-environment fit models, and any theory that considers outcomes to be a result of combined influences emanating from different levels of analysis (Mathieu et al., 2012). In addition to their specific role in those theoretical models, cross-level interaction effects are important in general because they are indicative of the presence of moderator variables. Specifically, the extent to which we understand the presence of cross-level interactions is an indication of theoretical progress because such relationships inform us of the conditions under which relationships change in nature, strength, or both. Cross-level interaction effects are particularly useful for practice because they provide information on the situations when a given intervention may result in more or less positive outcomes. For example, practitioners are particularly interested in knowing whether pre-employment tests, leadership training and development programs, performance management and appraisal processes, and compensation systems are equally as effective in terms of improving individual performance across different types of jobs and occupations, units of a firm (e.g., branches of a bank), and geographic locations (e.g., subsidiaries in different countries).

There is a fundamental question that all substantive researchers face prior to embarking in the search for cross-level interaction effects. Moreover, this fundamental question has remained unchanged since the very inception of multilevel modeling (e.g., Burstein, Linn, & Capell, 1978; Robinson, 1950) and, simply put, is: What is the degree of variability of a lower-level relationship across higher-order units? This has been and continues to be a critical question because its answer will dictate whether one should proceed with a formal test of cross-level moderator hypotheses. Stated differently, variability in the relationship between two variables across higher-level units is a precondition for the presence of moderator variables that could possibly account for this variability.

The goal of our article is to offer an expanded and more comprehensive approach to answering the question of whether there is sufficient variability in a lower-level relationship across higher-level units to warrant the search for cross-level interaction effects. The remainder of our article is organized as follows. First, we describe how researchers typically assess variability across higher-level collectives or contexts and clarify that this usual procedure is not informative regarding the possible presence of cross-level interaction effects. Second, we offer a general variance decomposition of L2 variability in lower-level scores. This section includes a description of the multilevel model, the typical procedure for assessing the presence of variability based on the intraclass correlation (ICC) ρ_{α} , and the derivation of a new index of variability in lower-level relationships across higher levels of analysis, which we label intraclass correlation ρ_{β} . Third, we describe a Monte Carlo study complementing analytical material in the previous section to provide evidence that ρ_{α} and ρ_{β} are indexes of orthogonal sources of variance. Fourth, we use data from recently published articles to illustrate the need for our recommended expanded procedure that includes ρ_{β} —and also demonstrate how decisions regarding the use of multilevel modeling improve as a consequence. Fifth, we compare and show the superiority of our newly proposed ρ_{β} to other indicators of variability that, although available in the statistical and methodological literature, are not usually implemented by organizational science researchers. For example, we describe that these indicators rely on significance testing procedures that require a large number of L2 units that is not frequently observed in management and organizational studies research (Mathieu et al., 2012). Sixth, we offer an illustrative data set and the R function `icc_beta` for computing ρ_{β} in single- and multiple-predictor situations and also make this package available through the Comprehensive R Archive Network (CRAN; <http://cran.us.r-project.org>). Finally, we close with recommendations regarding the expanded decision-making procedure for examining cross-level interaction effects and the possible presence of nonindependence in future empirical research, even if the particular research design and hypotheses do not include multilevel considerations explicitly.

Assessing Cross-Level Dependence and Variability

As is the case in all empirical research, theory considerations dictate the appropriateness of a particular data-analytic approach. Specifically regarding the possibility of using multilevel modeling in general and testing hypotheses about cross-level interactions in particular, there may be theory-based considerations that lead a researcher to suspect that dependence may be present in the data (i.e., variability based on a higher-level context or process). Moreover, as noted by Kenny and Judd (1996), “observations may be dependent, for instance, because they share some common feature, come from some common source, are affected by social interaction, or are arranged spatially or sequentially in time” (p. 138). Stated differently, the resulting data structure may include dependence of observations due to shared experiences even if there is no formal hierarchical structure such as individuals formally belonging to different teams.

Given theory-based considerations, there is a need to assess empirically the extent to which these shared experiences and context and, more generally, the clustering of entities within collectives have actually led to dependence. To do so, the consistent recommendation in the multilevel modeling literature is to assess the degree of dependence by computing the intraclass correlation ρ_{α} , which

assesses the proportion of between-group variance relative to total variance in an outcome variable and can be interpreted as the correlation between two randomly selected members of the same group. This same recommendation is offered in many of the most influential and established textbooks addressing multilevel modeling (e.g., Heck et al., 2010; Hox, 2010; Raudenbush & Bryk, 2002; Snijders & Bosker, 2012). As summarized by Heck et al. (2010),

The first step in a multilevel analysis is partitioning the variance in an outcome variable into its within- and between-group components. If it turns out that there is little or no variation (perhaps less than 5%) in outcomes between groups, there would be no compelling need for conducting a multilevel analysis. (p. 6)

Not surprisingly, given this consistent recommendation in the methodological literature, substantive researchers compute and report results regarding ρ_α as evidence regarding the presence or absence of dependence and for justifying using multilevel modeling (or not) and subsequently testing cross-level interaction hypotheses. This is a pervasive and common practice that is reported in virtually all articles addressing multilevel issues (e.g., Halbesleben, Wheeler, & Paustian-Underdahl, 2013; Hu & Liden, 2013; Hülshager, Alberts, Feinholdt, & Lang, 2013). Although some methodological sources have raised concerns about the sole reliance on ρ_α (e.g., Snijders & Bosker, 2012), and as we describe later in our article, some indicators of sources of L2 variance exist, using ρ_α is a well-established procedure for determining the possible presence of dependence and deciding whether the data structure requires the use of multilevel modeling. Next, we offer a general variance decomposition of L2 variability, which leads to the derivation of a new intraclass correlation ρ_β and analytical and empirical (i.e., Monte Carlo) evidence that this new index accounts for variance that is orthogonal to the variance assessed by ρ_α , which refers to differences across collectives and higher-level processes but ignores the variance attributed to differences in lower-level relationships.

General Variance Decomposition of Level 2 Variability in y_{ij} Scores

Multilevel Model With a Single Predictor

The relationship between a predictor and a criterion at the lower level of a multilevel study is (Enders & Tofighi, 2007; Raudenbush & Bryk, 2002):

$$y_{ij} = \beta_{0j} + \beta_{1j}x_{ij} + r_{ij}, \quad (1)$$

where y_{ij} is the criterion score for the i^{th} person in group j , β_{0j} is the intercept value for group j , β_{1j} is the slope for group j , x_{ij} is the predictor score for the i^{th} person in group j , and r_{ij} is the L1 residual term such that $r_{ij} \sim N(0, \sigma^2)$. The single predictor case involves estimating the following L2 intercept-only models (i.e., models that exclude L2 predictors) (Enders & Tofighi, 2007; Hofmann & Gavin, 1998; Hox, 2010):

$$\beta_{0j} = \gamma_{00} + u_{0j}, \quad (2)$$

where the intercept γ_{00} and the residual term u_{0j} describe how group-level intercepts deviate from the grand-mean intercept, respectively, and

$$\beta_{1j} = \gamma_{10} + u_{1j}, \quad (3)$$

where the intercept γ_{10} and the residual term u_{1j} describe how group-level slopes deviate from the grand-mean slope, respectively.

The usual assumption is that regression coefficients are distributed jointly as random normal variables (Hox, 2010; Raudenbush & Bryk, 2002),

$$\begin{bmatrix} \beta_{0j} \\ \beta_{1j} \end{bmatrix} \sim N_2 \left(\begin{bmatrix} \gamma_{00} \\ \gamma_{10} \end{bmatrix}, \begin{bmatrix} \tau_{00} & \tau_{01} \\ \tau_{01} & \tau_{11} \end{bmatrix} \right). \tag{4}$$

That is, τ_{00} and τ_{11} are the variances of β_{0j} and β_{1j} , respectively; τ_{01} is the covariance between β_{0j} and β_{1j} ; and u_{0j} and u_{1j} are residuals, or random effects, that capture group differences as mentioned previously. The average y_{ij} across L2 units is γ_{00} . Additionally, as mentioned previously, γ_{10} is the grand-mean slope of y_{ij} on x_{ij} across L2 units. Equations 2 and 3 do not include L2 predictors to be able to quantify the total variance attributed to group differences in intercepts (Equation 2) and slopes (Equation 3). Thus, the L2 Equations 2 and 3 can be substituted into the L1 Equation 1 to yield the mixed-model version of the multilevel linear model:

$$y_{ij} = \gamma_{00} + \gamma_{10}x_{ij} + u_{0j} + u_{1j}x_{ij} + r_{ij}. \tag{5}$$

General Multilevel Model

In matrix notation, the multilevel model with more than one predictor is the following (Hox, 2010; Raudenbush & Bryk, 2002):

$$y_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta}_j + r_{ij}, \tag{6}$$

where $\mathbf{x}'_{ij} = (x_{ij0}, x_{ij1}, \dots, x_{ijp})$ is a $p + 1$ dimensional row vector of predictors for person i in group j , $x_{ij0} = 1$ for the intercept, $\boldsymbol{\beta}_j$ is a $p + 1$ dimensional vector of coefficients (the first element is the intercept and the remaining are slopes), and r_{ij} is an error term. Let $\bar{\mathbf{x}}' = (\bar{x}_0, \bar{x}_1, \dots, \bar{x}_p)$ be a vector of grand means and $\bar{\mathbf{x}}'_j = (\bar{x}_{j0}, \bar{x}_{j1}, \dots, \bar{x}_{jp})$ a vector of predictor averages for group j , $N = \sum_{j=1}^J N_j$ where N_j is the sample size in group j ($j = 1, \dots, J$), and the overall sample mean is $\bar{y} = \sum_{j=1}^J N_j \bar{y}_j / N$. Also, let $\mathbf{X}'_j = (\mathbf{x}'_{1j}, \dots, \mathbf{x}'_{N_j j})$ be a $(p + 1) \times N_j$ matrix of predictors for group j and $\mathbf{X}' = (\mathbf{X}'_1, \dots, \mathbf{X}'_J)$ a $(p + 1) \times N$ matrix of predictors across the J groups.

Expanding on Equation 6, the typical multilevel model assumes that

$$r_{ij} \sim N(0, \sigma^2), \tag{7}$$

$$y_{ij} | \mathbf{x}_{ij}, \boldsymbol{\beta}_j, \sigma^2 \sim N(\mathbf{x}'_{ij}\boldsymbol{\beta}_j, \sigma^2), \tag{8}$$

$$\boldsymbol{\beta}_j \sim N_{p+1}(\boldsymbol{\gamma}, \mathbf{T}), \tag{9}$$

where σ^2 is the error variance conditioned on $\mathbf{x}'_{ij}\boldsymbol{\beta}_j$, $\boldsymbol{\gamma}$ is a vector of fixed-effects, and \mathbf{T} is the variance-covariance matrix of the regression coefficients. Note that deviation from the overall outcome variable mean is denoted by

$$y_{ij} - \bar{y} = (y_{ij} - \bar{y}_j) + (\bar{y}_j - \bar{y}), \tag{10}$$

where $y_{ij} - \bar{y}_j$ captures within group-variability whereas $\bar{y}_j - \bar{y}$ measures between-group variance. Accordingly, each source of variance is independent of the other.

In the next section, we derive ρ_β using the general variance decomposition of y_{ij} in Equation 10. Before we do so, however, we note that rescaling predictors is common in multilevel modeling to help in the interpretation of results (Aguinis et al., 2013; Dalal & Zickar, 2012; Enders & Tofighi, 2007). The two main rescaling approaches are group-mean centering and grand-mean centering, but other options include a hybrid approach (i.e., group-mean centering for L1 predictors and using group-level means for L2 predictors) as well as no rescaling at all. In the context of offering best-practice recommendations for assessing cross-level interaction effects, Aguinis et al. (2013) suggested using group-mean centering in most cases because the use of grand-mean centering conflates the between-L2 and within-L2 effects. Moreover, Aguinis et al. (2013), Enders and Tofighi (2007), and Hofmann and Gavin (1998) concluded that if a researcher uses grand-mean centering for the L1 predictor, it is not possible to make an accurate, or even meaningful, interpretation of the cross-level interaction effect. However, because there are several rescaling approaches available, the model in Equation 6 is general and can accommodate any centering strategy by, for example, defining the vector of predictors as $\mathbf{x}_{ij} - \bar{\mathbf{x}}$ for grand-mean centering or $\mathbf{x}_{ij} - \bar{\mathbf{x}}_j$ for group-mean centering. Moreover, as we describe later in our article, the computation of ρ_β remains the same regardless of the particular rescaling strategy.

Analytical Evidence of Differences between ρ_α and ρ_β

As shown in the Appendix, the sample variance S^2 can be partitioned as follows:

$$S^2 = \frac{1}{N-1} \sum_{i=1}^N (y_{ij} - \bar{y})^2 = S_W^2 + S_B^2, \quad (11)$$

where the between-group variance is $S_B^2 = \frac{1}{N-1} \sum_{j=1}^J N_j (\bar{y}_j - \bar{y})^2$ and the within-group variance is $S_W^2 = \frac{1}{N-1} \sum_{j=1}^J \sum_{i=1}^{N_j} (y_{ij} - \bar{y}_j)^2$. The traditional intraclass correlation ρ_α is defined as the portion of variance in the criterion (i.e., outcome variable) attributed to grouping or nesting (i.e., L2 variability), which is the relative size of S_B^2 to the total variance,

$$\rho_\alpha = \frac{S_B^2}{S_B^2 + S_W^2}. \quad (12)$$

To compute ρ_α , we first estimate the following random-intercepts model, which is equivalent to a one-way random effects analysis of variance (Aguinis et al., 2013; Hox, 2010; Snijders & Bosker, 2012):

$$y_{ij} = \gamma_{00} + u_{0j} + r_{0ij}, \quad (13)$$

where, as noted earlier, γ_{00} is the grand mean of y_{ij} , u_{0j} is the corresponding random effect, and r_{0ij} is the L1 error term. Based on the model in Equation 13, ρ_α is computed as follows (Raudenbush & Bryk, 2002):

$$\rho_\alpha = \frac{\tau_{00}}{\sigma_Y^2}, \quad (14)$$

where τ_{00} is the variance of u_{0j} and σ_Y^2 is the variance of y_{ij} .

Equation 10 shows that ρ_α is orthogonal to within-group deviations from the mean, $y_{ij} - \bar{y}_j$, and Equation 12 demonstrates that ρ_α measures the share of criterion variance attributed to between-

group differences in criterion means. In fact, derivations in the Appendix demonstrate that the expected value of S_B^2 is a function of both between-group intercept and slope differences as follows:

$$E[S_B^2|\mathbf{X}] = \frac{1}{N-1} \sum_{j=1}^J N_j tr \left\{ \left[\left(1 - \frac{N_j}{N} \right) \mathbf{T} + \boldsymbol{\gamma}\boldsymbol{\gamma}' \right] \bar{\mathbf{x}}_j \bar{\mathbf{x}}_j' \right\} - \frac{N}{N-1} tr \left[\boldsymbol{\gamma}\boldsymbol{\gamma}' \bar{\mathbf{x}}\bar{\mathbf{x}}' \right] + \frac{J-1}{N-1} \sigma^2, \quad (15)$$

where tr indicates a matrix trace (i.e., the sum of the elements along the diagonal).

In short, ρ_α is a function of between-group variability in intercepts and slopes as quantified by \mathbf{T} . However, the extent to which slope differences affect variability in $\bar{y}_j - \bar{y}$ is a function of group predictor mean differences. For simplicity of exposition, consider the case where $\bar{\mathbf{x}}_j = \bar{\mathbf{x}}$ for all j and the last p elements of $\bar{\mathbf{x}}$ are zero (e.g., grand-mean centered). One implication is that when groups are equivalent on the predictors, all but the first element of $\bar{\mathbf{x}}$ equals zero and $E[S_B^2|\mathbf{X}]$ simplifies significantly. For this special case, if $N_j = N_*$ and $\bar{\mathbf{x}}_j = \bar{\mathbf{x}}$ then $E[S_B^2|\mathbf{X}]$ reduces to

$$\frac{N - N_*}{N - 1} \mathbf{T}_{00} + \frac{J - 1}{N - 1} \sigma^2, \quad (16)$$

where $\mathbf{T}_{00} = Var(\beta_{0j}|\mathbf{X})$ and $N = JN_*$. Consequently, researchers can expect ρ_α to include some variance attributed to group slope differences in cases where groups vary in average predictor values.

The following derivations show that a portion of slope variability across groups also contributes to variability in $y_{ij} - \bar{y}_j$, which implies that ρ_α accounts for only a portion of variability attributed to slope differences. This finding is important because variability of $y_{ij} - \bar{y}_j$ is orthogonal to the traditional between group variability as indexed by $\bar{y}_j - \bar{y}$.

Consider S_{Wj}^2 and note that the average criterion score for group j is,

$$\bar{y}_j = \frac{1}{N_j} \sum_{i=1}^{N_j} (\mathbf{x}'_{ij} \boldsymbol{\beta}_j + r_{ij}) = \bar{\mathbf{x}}_j' \boldsymbol{\beta}_j + \bar{r}_j, \quad (17)$$

where $\bar{r}_j = \frac{1}{N_j} \sum_{i=1}^{N_j} r_{ij}$. Consequently, the within-group deviation is,

$$y_{ij} - \bar{y}_j = \mathbf{x}'_{ij} \boldsymbol{\beta}_j + r_{ij} - \bar{\mathbf{x}}_j' \boldsymbol{\beta}_j - \bar{r}_j = (\mathbf{x}_{ij} - \bar{\mathbf{x}}_j)' \boldsymbol{\beta}_j + r_{ij} - \bar{r}_j. \quad (18)$$

There are several important observations regarding Equation 18. First, partitioning variance in $y_{ij} - \bar{y}_j$ naturally leads to group-mean centering predictors because the within-group average on the outcome, \bar{y}_j , is a function of $\bar{\mathbf{x}}_j$. This implication is consistent with the consensual recommendation mentioned earlier that group-mean centering be the preferred rescaling strategy within the specific context of assessing cross-level interaction effects and quantifying outcome variance attributed to slope heterogeneity. Second, in Equation 18 the first element of $\mathbf{x}_{ij} - \bar{\mathbf{x}}_j$ is zero because by definition $x_{ij0} = \bar{x}_{j0} = 1$. Consequently, group j 's intercept does not influence $y_{ij} - \bar{y}_j$. This result reinforces the finding in Equation 10 that the sources of variance captured by ρ_α and ρ_β are indeed orthogonal.

In order to find an estimator for ρ_β , we first must identify the expected value of the within-group variance, S_{Wj}^2 , given a matrix \mathbf{X} of predictors. In fact, the Appendix includes derivations that imply that $E[S_{Wj}^2|\mathbf{X}]$ is defined as

$$E[S_{Wj}^2|\mathbf{X}] = tr \left[\left(\mathbf{T} + \boldsymbol{\gamma}\boldsymbol{\gamma}' \right) \frac{\mathbf{X}'_c \mathbf{X}_c}{N-1} \right] + \frac{N-J}{N-1} \sigma^2, \quad (19)$$

where, as shown in Equation 9, \mathbf{T} is the variance-covariance matrix of the random effects and $\boldsymbol{\gamma}$ is a vector of fixed effects. Furthermore, $\mathbf{X}'_c \mathbf{X}_c = \sum_{j=1}^J \sum_{i=1}^{N_j} (\mathbf{x}_{ij} - \bar{\mathbf{x}}_j) (\mathbf{x}_{ij} - \bar{\mathbf{x}}_j)'$, which by definition implies

that \mathbf{X}_c is a $N \times (p + 1)$ matrix of group-mean centered predictors for all groups and individuals. Note that the first column of \mathbf{X}_c includes zeroes for the intercept and $\mathbf{X}'_c\mathbf{X}_c/(N - 1)$ represents the average within-group relationship among the predictors over groups. Equation 19 shows that $E[S^2_{\beta'}|\mathbf{X}]$ is a function of the random effects variance-covariance matrix, \mathbf{T} , the relationships among group-mean centered predictors $\mathbf{X}'_c\mathbf{X}_c/(N - 1)$, the fixed effects, $\boldsymbol{\gamma}$, and within-group error variance σ^2 . Accordingly, the portion of $E[S^2_{\beta'}|\mathbf{X}]$ that is attributed to group slope differences is $tr\left[\mathbf{T}\frac{\mathbf{X}'_c\mathbf{X}_c}{N-1}\right]$ and the intraclass correlation that quantifies the share of within-group outcome variance attributed to slope differences is

$$\rho_{\beta} = tr\left[\mathbf{T}\frac{\mathbf{X}'_c\mathbf{X}_c}{N-1}\right]S^{-2}. \tag{20}$$

One observation from Equation 20 is that the first row and column of $\mathbf{X}'_c\mathbf{X}_c$ includes zeros, which has the effect of removing the covariances between intercepts and slopes when computing $tr\left[\mathbf{T}\frac{\mathbf{X}'_c\mathbf{X}_c}{N-1}\right]$. Stated differently, covariances between group intercepts and slopes do not contribute to variance quantified by ρ_{β} .

Monte Carlo Empirical Evidence of Differences Between ρ_{α} and ρ_{β}

We conducted a Monte Carlo simulation as a follow-up to the analytical results. We implemented 1,000 replications with number of groups $J = 30$ and group size $N_j = 30$ for all groups for each condition and then calculated the empirically derived ρ_{α} and ρ_{β} values. We manipulated two factors that, based on the derivations in the previous section, affect ρ_{α} and ρ_{β} : (a) the variability of group intercepts and slopes as indicated by \mathbf{T} and (b) the extent to which groups differ in averages on the predictor variables. We generated the outcome variable as normal with unit variance when conditioned on the predictors and regression coefficients,

$$y_{ij}|\mathbf{x}_{ij}, \boldsymbol{\beta}_j \sim N\left(\mathbf{x}'_{ij}\boldsymbol{\beta}_j, 1\right). \tag{21}$$

In addition, we considered two scenarios for \mathbf{x}_{ij} and the between-group predictor means $\bar{\mathbf{x}}_j$. The first scenario reflected the circumstance where groups did not differ in the population on predictor averages,

$$\mathbf{x}_{ij}|\bar{\mathbf{x}}_j \sim N_3\left(\bar{\mathbf{x}}_j, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\right), \bar{\mathbf{x}}_j \sim N_3\left(\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}\right). \tag{22}$$

That is, $\bar{\mathbf{x}}_j$ was assumed to originate from a population with constant values in each group where the first element equals 1 to represent the intercept and, consequently, all predictor variance is attributed to within-group differences. The second scenario for \mathbf{x}_{ij} and $\bar{\mathbf{x}}_j$ incorporates some between-group predictor variance. Namely, the distributions for \mathbf{x}_{ij} and $\bar{\mathbf{x}}_j$ are

$$\mathbf{x}_{ij}|\bar{\mathbf{x}}_j \sim N_3\left(\bar{\mathbf{x}}_j, \frac{4}{5}\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\right), \bar{\mathbf{x}}_j \sim N_3\left(\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \frac{1}{5}\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\right), \tag{23}$$

which reflects a circumstance where 20% of the variability in \mathbf{x}_{ij} is due to between-group variability. For both scenarios, the distribution of \mathbf{x}_{ij} conditioned on $\bar{\mathbf{x}}_j$ is trivariate normal.

Table 1. Monte Carlo Simulation Results: Average ρ_α and ρ_β for Different Scenarios Manipulating Between-Group Predictor Differences and Intercept and Slope Differences.

T Scenarios	\mathbf{x}_{ij} Scenario 1						\mathbf{x}_{ij} Scenario 2							
	$\hat{\rho}_\alpha$	$\hat{\rho}_\beta$	ρ_α	ρ_β	LRT	$\bar{\lambda}_1$	$\bar{\lambda}_2$	$\hat{\rho}_\alpha$	$\hat{\rho}_\beta$	ρ_α	ρ_β	LRT	$\bar{\lambda}_1$	$\bar{\lambda}_2$
1	.238	.004	.242	.000	.012	.097	.096	.287	.004	.292	.000	.010	.098	.103
2	.003	.373	.000	.400	1.000	.922	.920	.114	.297	.117	.320	1.000	.905	.904
3	.160	.312	.161	.333	1.000	.921	.922	.254	.252	.259	.267	1.000	.904	.903

Note: $\hat{\rho}_\alpha$ and $\hat{\rho}_\beta$ are empirically derived Monte Carlo estimates of the theoretical values ρ_α and ρ_β using 1,000 replications. \mathbf{x}_{ij} Scenario 1 corresponds to no between-group differences, and scenario 2 represents the case of 20% variance in predictors attributed to group membership. The **T** scenarios reflect scenario 1: only intercept differences, scenario 2: only slope differences, and scenario 3: intercept and slope differences. “LRT” denotes the probability of rejecting the likelihood ratio test where the observed deviance was compared with a χ^2_5 critical value (there are 5 degrees of freedom: two slope variances, one covariance between random slopes, and two covariances between slopes and the random intercept). The full model “LRT” denotes Type I error rates for **T** scenario 1 and statistical power for **T** scenarios 2 and 3. $\bar{\lambda}_1$ and $\bar{\lambda}_2$ are the average least squares slope reliabilities for the two simulated continuous predictors.

Also, we examined the effect of different values of **T** on ρ_α and ρ_β across three situations in which we fixed the means of β_j in the population to $\gamma = (0, 2^{-1}, 2^{-1})$ and specified **T** so that β_j is distributed as follows:

$$\beta_j \sim N_3 \left(\frac{1}{2} \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \mathbf{T} \right). \tag{24}$$

The three scenarios for **T** are as follows:

$$\mathbf{T} = \frac{1}{2} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \mathbf{T} = \frac{1}{2} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \mathbf{T} = \frac{1}{2} \mathbf{I}_3, \tag{25}$$

where \mathbf{I}_3 denotes a three-dimensional identity matrix. In other words, the first scenario for **T** represents the case of no variability in slopes, the second scenario includes between-group variance only in group slopes, and the third scenario considers variability in intercepts and slopes.

Table 1 reports values for $\hat{\rho}_\alpha$ and $\hat{\rho}_\beta$, which are averages of Monte Carlo estimates of ρ_α and ρ_β from 1,000 replications for the six conditions full crossing (a) two conditions for between-group predictor differences and (b) three conditions for intercept and slope differences. For the sake of completeness, Table 1 also reports results regarding the likelihood ratio test (LRT) of non-zero τ_{11} as well as $\bar{\lambda}_1$ and $\bar{\lambda}_2$ —average least squares slope reliabilities for the two simulated continuous predictors (we describe results pertaining to LRT and reliabilities later in our article).

For scenario 1 for **T** and scenario 1 for \mathbf{x}_{ij} , all between-group variance is attributed to variability in intercepts only and groups are equivalent on the predictors. Table 1 reports Monte Carlo values of $\hat{\rho}_\alpha = 0.238$ and $\hat{\rho}_\beta = 0.004$ and their theoretical counterparts are quite similar: $\rho_\alpha = 0.242$ and $\rho_\beta = 0.000$. Introducing group predictor differences in scenario 2 results in a larger $\hat{\rho}_\alpha = 0.287$, which corresponds closely with the theoretical value of $\rho_\alpha = 0.292$.

The second **T** scenario includes group slope variance and no intercept variance with average Monte Carlo values of $\hat{\rho}_\alpha = 0.003$ and $\hat{\rho}_\beta = 0.373$ for the \mathbf{x}_{ij} scenario 1. The value of $\hat{\rho}_\alpha = 0.003$ mirrors the analytic value of $\rho_\alpha = 0.000$ and demonstrates that ρ_β accounts for variance that is orthogonal to the variance captured by ρ_α .

Results regarding scenario 2 for \mathbf{x}_{ij} demonstrates another contribution of our study, which is that ρ_α includes variance attributed to slopes only when groups differ in average predictor values. The

second scenario for \mathbf{x}_{ij} represents a situation where 20% of predictor variability is attributed to group mean differences, and in this case $\hat{\rho}_\alpha = 0.114$ (with a theoretical value of $\rho_\alpha = 0.117$), which reflects the variance in criterion means that is attributed to slope differences. Table 1 also shows a decrease in ρ_β in the case of group differences in the predictors with $\hat{\rho}_\beta = 0.297$ and $\rho_\beta = 0.320$.

In the third and final **T** scenario, groups differ in both intercepts and slopes with average Monte Carlo values of $\hat{\rho}_\alpha = 0.160$ and $\hat{\rho}_\beta = 0.312$ with corresponding theoretical values of $\rho_\alpha = 0.161$ and $\rho_\beta = 0.333$, respectively. As expected, introducing group predictor heterogeneity in scenario 2 for \mathbf{x}_{ij} results in a larger $\hat{\rho}_\alpha = 0.254$ and a smaller $\hat{\rho}_\beta = 0.252$. This third **T** scenario further demonstrates that ρ_α and ρ_β capture two different sources of criterion variance.

In short, Monte Carlo simulation results confirmed the analytical evidence presented earlier: ρ_α and ρ_β are orthogonal and capture two different sources of between-group criterion variance. In addition, ρ_α includes variance attributed to slopes only when groups differ in average predictor values.

Implications for Substantive Research: Different Conclusions Based on the Use of ρ_α Versus ρ_β

Mathieu et al. (2012, Table 1) reported summary statistics for a set of articles addressing cross-level interaction effects. Mathieu et al. reported data for Studies 1 and 2 (based on Chen, Kirkman, Kanfer, Allen, & Rosen, 2007), Study 3 (based on Hofmann, Morgeson, & Gerras, 2003), Studies 4 and 5 (based on Liao & Rupp, 2005), and Studies 6 and 7 (based on Mathieu, Ahearne, & Taylor, 2007). We used those data from Mathieu et al. to compare and contrast the magnitude of ρ_β to the traditional ICC ρ_α .

Table 2 includes ρ_β values, which we calculated for each study. Across the seven studies, ρ_β ranged from 0.00 to 0.111. Stated differently, group slope differences accounted for a range of 0.00% to 11.1% of the variance in the outcome variables, and the average variance attributed to between-group differences in slopes was 3.5%.

Table 2 also shows that for 2 of the 7 studies, the variance attributed to slope differences was larger than the variance associated with differences in outcome variable means as assessed by the traditional ICC (i.e., $\rho_\beta > \rho_\alpha$). The values of ρ_α and ρ_β have implications for how researchers proceed with data analysis and model selection. Specifically, those who base judgment about the need to use multilevel modeling solely on ρ_α are likely to mistakenly ignore important between-group slope variance. For instance, consider Study 1 in Table 2. For Study 1, $\rho_\alpha = 0.00$ whereas $\rho_\beta = 0.057$, which implies that groups do not differ in averages on the outcome variable and 5.7% of criterion variable variance is attributed to group slope differences. Consequently, researchers who rely solely on ρ_α may incorrectly conclude that it is appropriate to use OLS regression as opposed to multilevel modeling. Moreover, this decision would also result in not estimating cross-level interaction effects.

An additional consideration regarding results in Table 2 concerns the magnitude of ρ_β . Specifically, many of the values may seem to be small—perhaps leading to the conclusion that the search for cross-level interaction effects may be futile. However, this conclusion is not warranted for the following reasons. First, LeBreton and Senter (2008) argued that ICC values of about .05 represent a small to medium effect and “values as small as .05 may provide prima facie evidence of a group effect” (p. 838). Reinforcing this recommendation, evidence made available recently demonstrates that the effect size guidelines reported by Cohen (1988) are overestimates of the types of effects usually reported in management and organizational studies research (Bosco, Aguinis, Singh, Field, & Pierce, in press). Second, ICC values should be considered within specific contexts because small observed effects may result from inauspicious designs, studies that involve phenomena leading to obscure consequences, and studies that challenge fundamental assumptions (Cortina & Landis, 2009). Third, there are numerous methodological and statistical artifacts that decrease observed effect sizes compared to their true population counterparts, and this is particularly true for interaction effects (Aguinis, 2004; Aguinis, Beaty, Boik, & Pierce, 2005; Aguinis & Stone-Romero, 1997). Accordingly, it is likely that in many

Table 2. Illustration of Differences Between Traditional (i.e., ρ_α) and Newly Derived Intraclass Correlation (i.e., ρ_β) as Indicators of Higher-Level Variability.

Study	Variables	ρ_α	ρ_x	τ_{11}	ρ_β
1: Chen, Kirkman, Kanfer, Allen, and Rosen (2007)	X: Leader-member exchange W: Empowering leadership Y: Individual empowerment	0.000	0.040	0.064	0.057
2: Chen et al. (2007)	X: Individual empowerment W: Team empowerment Y: Individual performance	0.176	0.000	0.002	0.002
3: Hofmann, Morgeson, and Gerras (2003)	X: Leader-member exchange W: Safety climate Y: Safety role definitions	0.000	0.409	0.210	0.111
4: Liao and Rupp (2005)	X: Justice orientation W: Org.-focus PJ climate Y: Satisfaction with organization	0.107	0.111	0.011	0.008
5: Liao and Rupp (2005)	X: Justice orientation W: Sup.-focus PJ climate Y: Supervisor commitment	0.310	0.111	0.021	0.015
6: Mathieu, Ahearne, and Taylor (2007)	X: Work experience W: Empowering leadership Y: Technology self-efficacy	0.000	0.022	0.000	0.000
7: Mathieu et al. (2007)	X: Technology use W: Empowering leadership Y: Individual performance	0.326	0.037	0.114	0.066
Average		0.131	0.104	0.060	0.037

Note: ρ_α = traditional intraclass correlation, ρ_x = intraclass correlation for x_{ij} , τ_{11} = variance of β_{1j} , and ρ_β = intraclass correlation indexing the proportion of between-group variance in an outcome variable due to differences in slopes across groups. Data source: Mathieu, Aguinis, Culpepper, and Chen's (2012) Table 1, who calculated τ_{11} after transforming the predictor and criterion to have unit variance. X: L1 predictor, W: L2 moderator, Y: L1 outcome, Org.-focus PJ: organization-focused procedural justice, Sup.-focus PJ: supervisor-focused procedural justice.

cases, and due to methodological and statistical artifacts, observed variability in slopes as indicated by ρ_β is actually larger in the population than has been estimated. Fourth, in some cases, small effect sizes may be practically significant (Aguinis et al., 2010). Thus, when the phenomena of interest has important implications for theory or practice, even small values for ρ_β may be considered as an indication for the need to assess particular cross-level interaction hypotheses.

Comparison of ρ_β With Existing Tests and Statistics

Although the recommendation in most textbooks on multilevel modeling is to rely on the traditional intraclass correlation ρ_α and management and organizational studies researchers seem to follow this recommendation, there are tests and indices described in the statistical and methodological literature that researchers might employ for understanding the nature of group slope differences (LaHuis, Hartman, Hakoyama, & Clark, 2014). However, ρ_β offers a unique value-added contribution as an index for how group variability in slopes translates to differences in the outcome variable y_{ij} . Furthermore, ρ_β provides researchers insight regarding the amount of variability that exists in y_{ij} due to group differences in slopes and can guide researchers in the theory-based search for cross-level moderators. As such, ρ_β can be used to produce preliminary estimates of group slope differences based on pilot data or exploratory analyses. Next, we compare and contrast ρ_β to tests based on model likelihood and statistics based on the variance of group slopes and reliability indices.

Likelihood Ratio Tests of Non-Zero τ_{11}

Researchers could employ LRTs to statistically evaluate whether group slopes differ. Specifically, LRTs are defined as the difference between -2 log-likelihood values ($-2LL$) for a full and reduced model. For the current context, $-2LL_{null}$ represents a random intercepts model with nonrandom slopes whereas $-2LL_{full}$ allows group slopes to differ. LRTs use deviance as a test statistic where deviance = $2LL_{full} - 2LL_{null}$, which is asymptotically distributed as a χ^2 random variable (Aguinis et al., 2013).

It is possible to test the significance of adding a random slope to an existing random intercept model with a chi-square distribution with 2 degrees of freedom (i.e., $\chi^2_{1-\alpha,2}$): one for the new random slope and a second for the covariance between group slopes and intercepts. However, using $\chi^2_{1-\alpha,2}$ to test for group slope differences is asymptotically too conservative. Specifically, Self and Liang (1987) and Stram and Lee (1994) offered theoretical results showing that the preferred option is a chi-square distribution that is a 50:50 mixture of $\chi^2_{1-\alpha,1}$ and $\chi^2_{1-\alpha,2}$ distributions. Additional research has proposed a Monte Carlo procedure for comparing a calculated $-2LL$ to an appropriate null distribution (Crainiceanu & Ruppert, 2004). Theoretical results hold for large sample sizes, but these are only rarely observed in typical management and organizational studies research, as documented by Mathieu et al.'s (2012) review.

An additional drawback is that the LRT test does not involve a meaningful effect size to assess the extent to which group slopes differ. Namely, LRTs serve as an omnibus test for the presence or absence of group slope differences and only provide researchers with guidance as to whether group slopes might differ in the population. In contrast, ρ_β quantifies the amount of group differences in y_{ij} that is attributed to group differences in slopes. Consequently, in contrast to LRT, researchers can use ρ_β as an index concerning the effect size of group differences in slopes on y_{ij} .

We illustrate some of the aforementioned limitations regarding the use of null hypothesis significance testing to assess the presence of variability of slopes with data made available by Hofmann, Griffin, and Gavin (2000). We refer to this particular illustration because their chapter and data set are used by many instructors teaching multilevel modeling. In particular, a reanalysis of the Hofmann et al. data suggests a statistically significant slope variance with $p < 10^{-6}$ for mood whereas $\hat{\rho}_\beta = 0.003$ (note that $\hat{\rho}_\alpha = 0.770$). In other words, ρ_β seems to suggest that the variance in the outcome attributed to group-based slope differences in mood is relatively small whereas employing an LRT (with a Monte Carlo procedure using the LRTSim R function) indicates the slope differs from zero (and it is very unlikely that it is zero). This apparent discrepancy in conclusions based on the LRT, and ρ_β is due to the fact that LRTs are tests that become more statistically powerful for larger sample sizes. In contrast, ρ_β is an effect size estimate that is independent of sample size. Specifically, the Hofmann et al. data set includes 1,000 observations (i.e., 50 groups and 20 individuals per group), which results in a statistically powerful test to find non-zero differences in slopes—even if they are small. On the other hand, ρ_β indicates that the variance in the outcome attributed to these slope differences is substantively small. In short, these results offer a good illustration regarding fundamental differences between null hypothesis statistical tests and effect size estimates as they pertain particularly to the assessment of slope variance across groups.

Finally, results included in Table 1 pertaining to LRT also offer additional insights regarding this test in relationship with ρ_β . Specifically, Table 1 includes results based on a comparison of a full model with random intercepts and slopes to a model with only random intercepts. Statistical power for the LRT (i.e., T scenarios 2 and 3) was 1.00 across the three scenarios for x_{ij} . Clearly, knowing the precise amount of variance due to slope differences is more informative than concluding, across all of these conditions, that such variance is unlikely to be zero. In addition, Table 1 shows that using a critical value from a chi-square distribution with 5 degrees of freedom led to a statistically conservative test as only 1.2% and 1.0% of replications were rejected in comparison to the true 5% rejection (i.e., a priori Type I error) level.

Statistics Describing Group Slope Differences

There are additional strategies for interpreting the size and nature of group slope differences. These include confidence intervals and reliability coefficients for group slopes.

First, the variance of slopes could be used to gauge the degree of group slope differences in the population. Specifically, τ_{11} is the variance of the slope random effect, and larger values imply greater group-based differences. Furthermore, it is possible to use τ_{11} to construct a confidence interval or range of plausible values for group slopes in repeated sampling from the population (Raudenbush & Bryk, 2002). For instance, in larger samples it is possible to compute $\gamma_{10} \pm 2\tau_{11}$ to indicate the approximately 95% range of slope values in the population. That is, larger plausible values ranges are associated with greater group differences in slopes.

Confidence intervals involving τ_{11} provide a practical understanding of group differences in the population. However, τ_{11} and any computed confidence intervals do not translate how variance in y_{ij} is attributed to group differences in slopes. Specifically, τ_{11} is on the metric of the relationship between x_{ij} and y_{ij} rather than a metric for the variance of y_{ij} . Instead, ρ_{β} is a function of τ_{11} and offers an index for translating observed slope differences into the degree to which group slope differences contribute towards differences in y_{ij} .

As an additional approach, Raudenbush and Bryk (2002) described equations for estimating the reliability of group slopes as an index for the extent to which groups differ in β_{1j} , but not y_{ij} . Specifically, the reliability of β_{1j} across groups is defined as,

$$\lambda_1 = \frac{1}{J} \sum_{j=1}^J \frac{\tau_{11}}{\tau_{11} + v_{1j}}, \quad (26)$$

where J is the number of groups and v_{1j} is the sampling error for β_{1j} . Note that $\tau_{11} + v_{1j}$ represents the “total variance” of β_{1j} , so λ_1 represents the average proportion of true slope variability across J groups and, more precisely, quantifies the proportion of variance in the least squares group slopes that are attributed to true between-group slope differences. Although λ_1 is a valuable index, it does not indicate how group slope differences relate to variability in y_{ij} . Furthermore, λ_1 provides information about a single predictor only and does not yield information concerning the overall reliability of group slope differences. In contrast, ρ_{β} is an index of criterion variance attributed to group slope differences.

Referring back to Table 1, results offer additional insights regarding the reliability of slopes as an index of slope variability in relationship with ρ_{β} . For example, Table 1 shows that the average slope reliability was approximately 0.1 in the absence of slope differences (i.e., **T** scenario 1) and 0.9 in the presence of slope differences (i.e., **T** scenarios 2 and 3). Clearly, λ_1 and λ_2 provide an indication as to the share of slope variance that is true rather than sampling error, but slope reliability does not explicitly describe criterion variance attributed to slope differences.

Illustrative Data Set for Computing ρ_{β}

We created a data set to illustrate the computation of ρ_{β} . The data include the following simulated variables: “1,” which is a column of 1s, “X1” (L1 predictor), “X2” (L2 predictor), and “Y” (criterion). The R code we used for all calculations is included as a vignette in the documentation for the `icc_beta` R package, which can be freely downloaded through the CRAN. We make this code and data available so that they can be used for calculating ρ_{β} for substantive as well as instructional purposes. Note that the code can be used in situations including any number of predictors. The data were simulated using the same code as for the simulation with the population model defined as:

$$y_{ij} | \mathbf{x}_{ij}, \boldsymbol{\beta}_j \sim N(\mathbf{x}'_{ij} \boldsymbol{\beta}_j, 1) \quad (27)$$

$$\mathbf{x}_{ij}|\bar{\mathbf{x}}_j \sim N_3 \left(\bar{\mathbf{x}}_j, \frac{49}{50} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right), \bar{\mathbf{x}}_j \sim N_3 \left(\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \frac{1}{50} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right), \boldsymbol{\beta}_j \sim N_3 \left(\frac{1}{2} \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \frac{1}{20} \mathbf{I}_3 \right). \quad (28)$$

Implementation of the R code using the illustrative data set leads to the following results and conclusions (note that we used a random-intercepts model to calculate ρ_α). First, $\hat{\rho}_\alpha = 0.033$, which implies that 3.3% of the variability in the criterion is associated with group mean differences. Second, we estimated a random-slopes multilevel model that included both X1 and X2 as predictors of Y. Results show that $\hat{\rho}_\beta = 0.075$, which means that 7.5% of the variability in Y is accounted for by group differences in slopes. In conclusion, the sole reliance on ρ_α , which is small, would have led to the conclusion that multilevel modeling may not be necessary and, consequently, a missed opportunity to investigate cross-level interaction effects. In contrast, the result based on ρ_β that about 7% of variance is attributed to slope differences across groups leads to the conclusion that there is a need to understand which particular higher-level moderators explain this variance.

Discussion

There is an increased awareness regarding the need to understand the nature of cross-level interaction effects—the extent to which relationships at a lower-level of analysis (e.g., two individual-level variables) vary across higher-level units (e.g., groups, units, firms). This need is central for making progress in modern-day contingency theories, person-environment fit models, and any theory that considers outcomes to be a result of combined influences emanating from different levels of analysis. In addition, a better understanding of cross-level interaction effects offers information that practitioners can use in planning and implementing interventions in specific contexts because such knowledge allows them to anticipate the relative effectiveness of such interventions given certain contextual (i.e., higher-level) factors. Thus, knowledge about cross-level interaction effects allows practitioners to enhance the effectiveness of interventions. Because of these reasons, there has been an exponential growth in the literature on multilevel modeling (Aguinis, Pierce, Bosco, & Muslin, 2009; Mathieu et al., 2012), which is a data-analytic approach that considers and models data dependence and such cross-level interaction effects explicitly.

In spite of the increased diversity and complexity in the methodological literature, there is a common challenge that permeates all types of multilevel modeling: the need to understand the degree of variability of a lower-level relationship across higher-order units, processes, or contexts. Although theory-based considerations initially dictate whether multilevel modeling may be the preferred data-analytic approach, the consistent recommendation in the methodological literature is that researchers first compute the intraclass correlation, ρ_α , as a criterion in deciding whether to use multilevel modeling (e.g., Aguinis et al., 2013). Not surprisingly, researchers use this criterion for deciding whether the use of multilevel modeling is appropriate. If the intraclass correlation is not sufficiently high, then multilevel modeling is not considered necessary and there is also not sufficient justification for assessing cross-level interaction effect hypotheses. Inversely, multilevel modeling is the preferred approach if the intraclass correlation ρ_α is sufficiently high. The reason for this recommendation, which is offered in most major textbooks on multilevel modeling, is that the intraclass correlation ρ_α assesses the proportion of between-group variance relative to total variance in an outcome variable, and therefore, it signals the presence of nonindependence in the data structure.

Our article showed analytically and via simulation that the current conceptualization and estimation of the intraclass correlation captures across-group variability due to intercept differences and

only a portion of variability attributed to slope differences. Thus, the intraclass correlation ρ_α is, using psychometric terminology, a deficient index of dependence. In other words, across-group variability may also exist due to slope differences across groups, but this is not reflected in the intraclass correlation as currently conceptualized and calculated. In contrast, the newly derived intraclass correlation ρ_β is an index of proportion of variance in criterion scores due to group differences, but the source of this variability is group difference in slopes.

We used data reported in several articles addressing substantive theories and research domains to illustrate that using the traditional intraclass correlation ρ_α as an index of group differences ignores the variance attributed to group slope differences and reduces the total reported variance attributed to group differences. In some cases, using ρ_α as the sole criterion for understanding the degree of across-group variability may lead researchers to miss an opportunity to study cross-level interaction effects that, as noted by Mathieu et al. (2012), “lay at the heart of modern-day contingency theories, person-environment fit models, and any theory that considers outcomes to be a result of combined influences emanating from different levels of analysis” (p. 952).

There is an additional use for ρ_β that has implications for future theory development. Because ρ_β is expressed in standardized metric, its value is not dependent on the particular scales used in a particular study. Accordingly, similar to a Pearson’s correlation coefficient, ρ_β can be used in meta-analytic reviews, and such research can open up new lines of investigation. For example, assume that a meta-analysis of the literature on the relationship between job satisfaction and task performance results in a larger mean value for ρ_β compared to the estimate based on the relationship between job satisfaction and organizational citizenship behavior (OCB). This result implies that there is greater cross-level heterogeneity for the satisfaction–task performance relationship compared to the satisfaction-OCB relationship. Accordingly, this result indicates that it would be more fruitful to conduct primary-level research investigating cross-level moderating effects of the satisfaction–task performance compared to the satisfaction-OCB relationship. Alternatively, conducting meta-analyses based on ρ_β values can also result in information that would be useful in terms of deciding to not search for cross-level interaction effects in certain domains. Given the proliferation of management and organizational studies theories and the need to engage in theory pruning (Leavitt, Mitchell, & Peterson, 2010), using ρ_β as an index of where one should *not* search for cross-level interaction effects could be just as useful, or even more useful, than using it as an index of areas where such effects are more likely to be found.

In terms of yet additional uses of ρ_β , our discussion thus far focused on research designs in which units are nested within collectives such as individuals within groups, groups within firms, or firms within industries. However, ρ_β can also be computed within the context of longitudinal designs where repeated measurements are collected for units (e.g., individuals, firms) and time as well as time-varying predictors are included in the model. In terms of a multilevel model conceptualization, the lower level refers to observations and the higher level to the units (e.g., entrepreneurs, teams, firms) about which data have been collected over time. Referring back to Equation 10, in these types of designs, within-unit variability is captured by $y_{it} - \bar{y}_i$ where i indexes units and t indexes time. ρ_β can be particularly useful in studies adopting a longitudinal designs because it quantifies variance in an outcome over time that is attributed to unit-based differences in slopes. The presence of such variance can then lead to testing specific hypotheses about moderator variables that may account for slope differences.

Conclusion

ρ_α and ρ_β index different sources of variance in y_{ij} . Because ρ_α and ρ_β reflect two different sources of group-based differences, we suggest that researchers contemplating the use of multilevel modeling, as

well those who suspect nonindependence in their data structure, expand the decision criteria for using such data-analytic approach to include both types of intraclass correlations. Continued use of ρ_α as the sole decision criterion may lead to the inappropriate use of data-analytic approaches that require independence among observations and also lead to opportunity cost in terms of testing precise and specific cross-level interaction effect hypotheses. In contrast, using both ρ_α and ρ_β improves the decision-making procedures for using multilevel modeling and assessing of cross-level interaction effects.

Appendix

Decomposition of Sample Variance S^2 and Derivation of Expected Between-Group Variance $E[S_B^2|\mathbf{X}]$, Expected Within-Group Variance $E[S_W^2|\mathbf{X}]$, and Expected Sample Variance $E(S^2|\mathbf{X})$

Decomposition of S^2 . The sample variance S^2 can be partitioned as follows:

$$\begin{aligned}
 S^2 &= \frac{1}{N-1} \sum_{i=1}^N (y_{ij} - \bar{y})^2 = \frac{1}{N-1} \sum_{j=1}^J \sum_{i=1}^{N_j} (y_{ij} - \bar{y}_j + \bar{y}_j - \bar{y})^2 \\
 &= \frac{1}{N-1} \sum_{j=1}^J \sum_{i=1}^{N_j} \left[(y_{ij} - \bar{y}_j)^2 + 2(y_{ij} - \bar{y}_j)(\bar{y}_j - \bar{y}) + (\bar{y}_j - \bar{y})^2 \right] \quad (A1) \\
 &= \frac{1}{N-1} \sum_{j=1}^J \sum_{i=1}^{N_j} (y_{ij} - \bar{y}_j)^2 + \frac{1}{N-1} \sum_{j=1}^J N_j (\bar{y}_j - \bar{y})^2,
 \end{aligned}$$

where $\sum_{i=1}^{N_j} (y_{ij} - \bar{y}_j) = 0$ was used to obtain the last equality.

Derivation of $E[S_B^2|\mathbf{X}]$. Note that S_B^2 can be written as follows:

$$S_B^2 = \frac{1}{N-1} \sum_{j=1}^J N_j (\bar{y}_j - \bar{y})^2 = \frac{1}{N-1} \sum_{j=1}^J N_j \bar{y}_j^2 - \frac{N}{N-1} \bar{y}^2 \quad (A2)$$

Consequently, $E[S_B^2|\mathbf{X}]$ is defined as

$$E[S_B^2|\mathbf{X}] = \frac{1}{N-1} \sum_{j=1}^J N_j E(\bar{y}_j^2|\mathbf{X}_j) - \frac{N}{N-1} E(\bar{y}^2|\mathbf{X}) \quad (A3)$$

where \mathbf{X}_j is an $N_j \times (p + 1)$ matrix of predictors for group j . The following material derives expressions for both expectations in Equation A3.

First, the average criterion for group j is

$$\bar{y}_j = \frac{1}{N_j} \sum_{i=1}^{N_j} (\mathbf{x}'_{ij} \boldsymbol{\beta}_j + r_{ij}) = \bar{\mathbf{x}}'_j \boldsymbol{\beta}_j + \bar{r}_j \quad (A4)$$

where $\bar{\mathbf{x}}'_j = (\bar{x}_{j0}, \bar{x}_{j1}, \dots, \bar{x}_{jp})$ is a vector of predictor averages for group j . The expected value of group j 's criterion mean is $E(\bar{y}_j|\mathbf{X}_j) = \bar{\mathbf{x}}'_j \boldsymbol{\gamma}$ and the variance of \bar{y}_j given \mathbf{X}_j is $Var(\bar{y}_j|\mathbf{X}_j) = \bar{\mathbf{x}}'_j \mathbf{T} \bar{\mathbf{x}}_j + \sigma^2/N_j$. Consequently, $E(\bar{y}_j^2|\mathbf{X}_j)$ is

$$E[\bar{y}_j^2|\mathbf{X}_j] = \text{Var}(\bar{y}_j|\mathbf{X}_j) + [E(\bar{y}_j|\mathbf{X}_j)]^2 = \bar{\mathbf{x}}_j'(\mathbf{T} + \boldsymbol{\gamma}\boldsymbol{\gamma}')\bar{\mathbf{x}}_j + \frac{\sigma^2}{N_j}. \tag{A5}$$

Furthermore, the first sum in Equation A3 is

$$\begin{aligned} \frac{1}{N-1} \sum_{j=1}^J N_j E(\bar{y}_j^2|\mathbf{X}_j) &= \frac{1}{N-1} \sum_{j=1}^J N_j \left[\bar{\mathbf{x}}_j'(\mathbf{T} + \boldsymbol{\gamma}\boldsymbol{\gamma}')\bar{\mathbf{x}}_j + \frac{\sigma^2}{N_j} \right] \\ &= \frac{1}{N-1} \sum_{j=1}^J N_j \text{tr} \left[(\mathbf{T} + \boldsymbol{\gamma}\boldsymbol{\gamma}')\bar{\mathbf{x}}_j\bar{\mathbf{x}}_j' \right] + \frac{J}{N-1} \sigma^2. \end{aligned} \tag{A6}$$

For $E(\bar{y}^2|\mathbf{X})$, note that $E(\bar{y}^2|\mathbf{X}) = \text{Var}(\bar{y}|\mathbf{X}) + [E(\bar{y}|\mathbf{X})]^2$ where $E(\bar{y}|\mathbf{X})$ is defined as

$$E(\bar{y}|\mathbf{X}) = \frac{1}{N} \sum_{j=1}^J N_j E(\bar{y}_j|\mathbf{X}) = \frac{1}{N} \sum_{j=1}^J N_j \bar{\mathbf{x}}_j' \boldsymbol{\gamma} = \bar{\mathbf{x}}' \boldsymbol{\gamma} \tag{A7}$$

because $\bar{\mathbf{x}} = \frac{1}{N} \sum_{j=1}^J N_j \bar{\mathbf{x}}_j$. The law of total variance can be used to find an expression for $\text{Var}(\bar{y}|\mathbf{X})$ as

$$\begin{aligned} \text{Var}(\bar{y}|\mathbf{X}) &= E \left[\text{Var}(\bar{y}|\mathbf{X}, \boldsymbol{\beta}_j) \mid \mathbf{X} \right] + \text{Var} \left(E(\bar{y}|\mathbf{X}, \boldsymbol{\beta}_j) \mid \mathbf{X} \right) \\ &= E \left[\text{Var} \left(\frac{1}{N} \sum_i y_{ij} \mid \mathbf{X}, \boldsymbol{\beta}_j \right) \mid \mathbf{X} \right] + \text{Var} \left(E \left(\frac{1}{N} \sum_i y_{ij} \mid \mathbf{X}, \boldsymbol{\beta}_j \right) \mid \mathbf{X} \right) \\ &= E \left(\frac{\sigma^2}{N} \mid \mathbf{X} \right) + \text{Var} \left(\frac{1}{N} \sum_i \mathbf{x}'_{ij} \boldsymbol{\beta}_j \mid \mathbf{X} \right) = \frac{\sigma^2}{N} + \text{Var} \left(\frac{1}{N} \sum_{j=1}^J \sum_{i=1}^{N_j} \mathbf{x}'_{ij} \boldsymbol{\beta}_j \mid \mathbf{X} \right) \\ &= \frac{\sigma^2}{N} + \frac{1}{N^2} \sum_{j=1}^J N_j^2 \bar{\mathbf{x}}_j' \mathbf{T} \bar{\mathbf{x}}_j \\ &= \frac{1}{N^2} \sum_{j=1}^J N_j^2 \text{tr} \left[\mathbf{T} \bar{\mathbf{x}}_j \bar{\mathbf{x}}_j' \right] + \frac{\sigma^2}{N} \end{aligned} \tag{A8}$$

where the assumption of independence of $\boldsymbol{\beta}_j$ and the fact that $\sum_{i=1}^{N_j} \mathbf{x}'_{ij} = N_j \bar{\mathbf{x}}_j'$ was used to obtain the second to last equality. Consequently,

$$E[\bar{y}^2|\mathbf{X}] = \frac{1}{N^2} \text{tr} \left[\mathbf{T} \sum_{j=1}^J N_j^2 \bar{\mathbf{x}}_j \bar{\mathbf{x}}_j' \right] + \frac{\sigma^2}{N} + \text{tr} \left[\boldsymbol{\gamma}\boldsymbol{\gamma}' \bar{\mathbf{x}} \bar{\mathbf{x}}' \right]. \tag{A9}$$

Therefore, $E[S_B^2|\mathbf{X}]$ is defined as

$$\begin{aligned} E[S_B^2|\mathbf{X}] &= \frac{1}{N-1} \left[\sum_{j=1}^J N_j \text{tr} \left[(\mathbf{T} + \boldsymbol{\gamma}\boldsymbol{\gamma}')\bar{\mathbf{x}}_j\bar{\mathbf{x}}_j' \right] - \frac{1}{N} \sum_{j=1}^J N_j^2 \text{tr} \left[\mathbf{T} \bar{\mathbf{x}}_j \bar{\mathbf{x}}_j' \right] - N \text{tr} \left[\boldsymbol{\gamma}\boldsymbol{\gamma}' \bar{\mathbf{x}} \bar{\mathbf{x}}' \right] \right] + \frac{J-1}{N-1} \sigma^2 \\ &= \frac{1}{N-1} \sum_{j=1}^J N_j \text{tr} \left\{ \left[\left(1 - \frac{N_j}{N} \right) \mathbf{T} + \boldsymbol{\gamma}\boldsymbol{\gamma}' \right] \bar{\mathbf{x}}_j \bar{\mathbf{x}}_j' \right\} - \frac{N}{N-1} \text{tr} \left[\boldsymbol{\gamma}\boldsymbol{\gamma}' \bar{\mathbf{x}} \bar{\mathbf{x}}' \right] + \frac{J-1}{N-1} \sigma^2. \end{aligned} \tag{A10}$$

Derivation of $E[S_{\bar{w}}^2|\mathbf{X}]$. Note that $E[S_{\bar{w}}^2|\mathbf{X}]$ can be written as

$$\begin{aligned}
 E[S_{\bar{w}}^2|\mathbf{X}] &= E\left[\frac{1}{N-1}\sum_{j=1}^J\sum_{i=1}^{N_j}(y_{ij}-\bar{y}_j)^2\middle|\mathbf{X}\right] = \frac{1}{N-1}\sum_{j=1}^J\sum_{i=1}^{N_j}E\left[(y_{ij}-\bar{y}_j)^2\middle|\mathbf{X}\right] \\
 &= \frac{1}{N-1}\sum_{j=1}^J\sum_{i=1}^{N_j}E\left\{E\left[(y_{ij}-\bar{y}_j)^2\middle|\mathbf{X},\boldsymbol{\beta}_j\right]\middle|\mathbf{X}\right\},
 \end{aligned}
 \tag{A11}$$

where the last equality was obtained using the law of iterated expectations. Squaring $y_{ij} - \bar{y}_j$ implies that

$$(y_{ij} - \bar{y}_j)^2 = (\mathbf{x}_{ij} - \bar{\mathbf{x}}_j)' \boldsymbol{\beta}_j \boldsymbol{\beta}_j' (\mathbf{x}_{ij} - \bar{\mathbf{x}}_j) + (r_{ij} - \bar{r}_j)^2 + 2(\mathbf{x}_{ij} - \bar{\mathbf{x}}_j)' \boldsymbol{\beta}_j (r_{ij} - \bar{r}_j).
 \tag{A12}$$

Accordingly, the interior expectation in Equation A11 is

$$\begin{aligned}
 E\left[(y_{ij} - \bar{y}_j)^2\middle|\mathbf{X},\boldsymbol{\beta}_j\right] &= (\mathbf{x}_{ij} - \bar{\mathbf{x}}_j)' \boldsymbol{\beta}_j \boldsymbol{\beta}_j' (\mathbf{x}_{ij} - \bar{\mathbf{x}}_j) + E\left[(r_{ij} - \bar{r}_j)^2\middle|\mathbf{X},\boldsymbol{\beta}_j\right] \\
 &= (\mathbf{x}_{ij} - \bar{\mathbf{x}}_j)' \boldsymbol{\beta}_j \boldsymbol{\beta}_j' (\mathbf{x}_{ij} - \bar{\mathbf{x}}_j) + \sigma^2 \left(\frac{N_j - 1}{N_j}\right)
 \end{aligned}
 \tag{A13}$$

because the r_{ij} are independent with $E(r_{ij}r_{i'j}|\mathbf{X},\boldsymbol{\beta}_j) = 0$ for all $i \neq i'$ and,

$$\begin{aligned}
 E\left[(r_{ij} - \bar{r}_j)^2\middle|\mathbf{X},\boldsymbol{\beta}_j\right] &= E\left[\left(\frac{N_j - 1}{N_j}r_{ij} - \frac{1}{N_j}\sum_{i' \neq i} r_{i'j}\right)^2\middle|\mathbf{X},\boldsymbol{\beta}_j\right] = \left(\frac{N_j - 1}{N_j}\right)^2 E(r_{ij}^2|\mathbf{X},\boldsymbol{\beta}_j) \\
 &\quad + \frac{1}{N_j^2}\sum_{i' \neq i} E(r_{i'j}^2|\mathbf{X},\boldsymbol{\beta}_j) = \left(\frac{N_j - 1}{N_j}\right)^2 \sigma^2 + \frac{1}{N_j^2}\sum_{i' \neq i} \sigma^2 \\
 &= \left(\frac{N_j - 1}{N_j}\right)^2 \sigma^2 + \frac{N_j - 1}{N_j^2} \sigma^2 = \frac{N_j - 1}{N_j} \sigma^2.
 \end{aligned}
 \tag{A14}$$

Accordingly,

$$\begin{aligned}
 E[S_{\bar{w}}^2|\mathbf{X}] &= \frac{1}{N-1}\sum_{j=1}^J\sum_{i=1}^{N_j}E\left\{(\mathbf{x}_{ij} - \bar{\mathbf{x}}_j)' \boldsymbol{\beta}_j \boldsymbol{\beta}_j' (\mathbf{x}_{ij} - \bar{\mathbf{x}}_j) + \frac{N_j - 1}{N_j} \sigma^2\middle|\mathbf{X}\right\} \\
 &= \frac{1}{N-1}\sum_{j=1}^J\sum_{i=1}^{N_j}(\mathbf{x}_{ij} - \bar{\mathbf{x}}_j)' (\mathbf{T} + \boldsymbol{\gamma}\boldsymbol{\gamma}') (\mathbf{x}_{ij} - \bar{\mathbf{x}}_j) + \frac{1}{N-1}\sum_{j=1}^J\sum_{i=1}^{N_j}\frac{N_j - 1}{N_j} \sigma^2 \\
 &= \frac{1}{N-1}\sum_{j=1}^J\sum_{i=1}^{N_j}tr\left\{(\mathbf{T} + \boldsymbol{\gamma}\boldsymbol{\gamma}') (\mathbf{x}_{ij} - \bar{\mathbf{x}}_j) (\mathbf{x}_{ij} - \bar{\mathbf{x}}_j)'\right\} + \frac{N - J}{N - 1} \sigma^2 \\
 &= tr\left\{(\mathbf{T} + \boldsymbol{\gamma}\boldsymbol{\gamma}') \sum_{j=1}^J\sum_{i=1}^{N_j}\frac{(\mathbf{x}_{ij} - \bar{\mathbf{x}}_j) (\mathbf{x}_{ij} - \bar{\mathbf{x}}_j)'}{N - 1}\right\} + \frac{N - J}{N - 1} \sigma^2 \\
 &= tr\left\{(\mathbf{T} + \boldsymbol{\gamma}\boldsymbol{\gamma}') \sum_{j=1}^J\frac{\mathbf{X}'_{cj}\mathbf{X}_{cj}}{N - 1}\right\} + \frac{N - J}{N - 1} \sigma^2 \\
 &= tr\left\{(\mathbf{T} + \boldsymbol{\gamma}\boldsymbol{\gamma}') \frac{\mathbf{X}'_c\mathbf{X}_c}{N - 1}\right\} + \frac{N - J}{N - 1} \sigma^2
 \end{aligned}
 \tag{A15}$$

where $\mathbf{X}'_{cj}\mathbf{X}_{cj} = \sum_{i=1}^{N_j} (\mathbf{x}_{ij} - \bar{\mathbf{x}}_j)(\mathbf{x}_{ij} - \bar{\mathbf{x}}_j)'$, which by definition implies that \mathbf{X}_{cj} is a matrix of group-mean centered predictors for group j with zeroes for the intercept, and $\mathbf{X}'_c\mathbf{X}_c = \sum_{j=1}^J \mathbf{X}'_{cj}\mathbf{X}_{cj}$ is the group-mean centered sums of squared predictors across all groups.

Derivation of $E(S^2|\mathbf{X})$. This section confirms the results by showing that $E(S^2|\mathbf{X}) = E(S^2_B|\mathbf{X}) + E(S^2_W|\mathbf{X})$. Note that S^2 can be written as

$$S^2 = \frac{1}{N-1} \sum_{j=1}^J \sum_{i=1}^{N_j} (y_{ij} - \bar{y})^2 = \frac{1}{N-1} \sum_{j=1}^J \sum_{i=1}^{N_j} y_{ij}^2 - \frac{N}{N-1} \bar{y}^2. \tag{A16}$$

Accordingly, the expected sample variance is

$$E(S^2|\mathbf{X}) = \frac{1}{N-1} \sum_{j=1}^J \sum_{i=1}^{N_j} E(y_{ij}^2|\mathbf{X}) - \frac{N}{N-1} E(\bar{y}^2|\mathbf{X}). \tag{A17}$$

By definition, $E(y_{ij}^2|\mathbf{X}) = \text{Var}(y_{ij}|\mathbf{x}_{ij}) + [E(y_{ij}|\mathbf{x}_{ij})]^2$. We can show that $E(y_{ij}|\mathbf{x}_{ij}) = \mathbf{x}'_{ij}\boldsymbol{\gamma}$ and $\text{Var}(y_{ij}|\mathbf{x}_{ij}) = \mathbf{x}'_{ij}\mathbf{T}\mathbf{x}_{ij} + \sigma^2 = \text{tr}[\mathbf{T}\mathbf{x}_{ij}\mathbf{x}'_{ij}] + \sigma^2$. Using the previous results concerning $E(\bar{y}^2|\mathbf{X})$ we see that,

$$\begin{aligned} E(S^2|\mathbf{X}) &= \frac{1}{N-1} \sum_{j=1}^J \sum_{i=1}^{N_j} \text{tr}[(\mathbf{T} + \boldsymbol{\gamma}\boldsymbol{\gamma}')\mathbf{x}_{ij}\mathbf{x}'_{ij}] + \frac{N}{N-1} \sigma^2 \\ &\quad - \frac{N}{N-1} \left[\frac{1}{N^2} \sum_{j=1}^J N_j^2 \text{tr}[\mathbf{T}\bar{\mathbf{x}}_j\bar{\mathbf{x}}'_j] + \text{tr}[\boldsymbol{\gamma}\boldsymbol{\gamma}'\bar{\mathbf{x}}\bar{\mathbf{x}}'] + \frac{\sigma^2}{N} \right] \\ &= \text{tr} \left\{ (\mathbf{T} + \boldsymbol{\gamma}\boldsymbol{\gamma}') \frac{\sum_{j=1}^J \sum_{i=1}^{N_j} \mathbf{x}_{ij}\mathbf{x}'_{ij}}{N-1} \right\} + \sigma^2 - \frac{1}{N-1} \frac{1}{N} \sum_{j=1}^J N_j^2 \text{tr}[\mathbf{T}\bar{\mathbf{x}}_j\bar{\mathbf{x}}'_j] - \frac{N}{N-1} \text{tr}[\boldsymbol{\gamma}\boldsymbol{\gamma}'\bar{\mathbf{x}}\bar{\mathbf{x}}'] \\ &= \text{tr} \left\{ (\mathbf{T} + \boldsymbol{\gamma}\boldsymbol{\gamma}') \frac{\mathbf{X}'\mathbf{X}}{N-1} \right\} + \sigma^2 - \frac{1}{N-1} \frac{1}{N} \sum_{j=1}^J N_j^2 \text{tr}[\mathbf{T}\bar{\mathbf{x}}_j\bar{\mathbf{x}}'_j] - \frac{N}{N-1} \text{tr}[\boldsymbol{\gamma}\boldsymbol{\gamma}'\bar{\mathbf{x}}\bar{\mathbf{x}}'] \end{aligned} \tag{A18}$$

Note that by definition, $\mathbf{X}'\mathbf{X} = \sum_{j=1}^J \sum_{i=1}^{N_j} \mathbf{x}_{ij}\mathbf{x}'_{ij} = \sum_{j=1}^J \mathbf{X}'_j\mathbf{X}_j$. Let $\mathbf{1}_{N_j}$ be a N_j vector of 1s and note that $\mathbf{X}_{cj} = \mathbf{X}_j - \mathbf{1}_{N_j}\bar{\mathbf{x}}'_j$ and $\mathbf{X}'_{cj}\mathbf{1}_{N_j} = \mathbf{0}$. Accordingly, adding and subtracting $\mathbf{1}_{N_j}\bar{\mathbf{x}}'_j$ implies that $\mathbf{X}'\mathbf{X}$ can be further partitioned as

$$\mathbf{X}'\mathbf{X} = \sum_{j=1}^J \mathbf{X}'_j\mathbf{X}_j = \sum_{j=1}^J (\mathbf{X}_{cj} + \mathbf{1}_{N_j}\bar{\mathbf{x}}'_j)' (\mathbf{X}_{cj} + \mathbf{1}_{N_j}\bar{\mathbf{x}}'_j) = \sum_{j=1}^J \mathbf{X}'_{cj}\mathbf{X}_{cj} + \sum_{j=1}^J N_j\bar{\mathbf{x}}_j\bar{\mathbf{x}}'_j. \tag{A19}$$

Substituting the result from A19 into A18 confirms that $E(S^2|\mathbf{X}) = E(S^2_B|\mathbf{X}) + E(S^2_W|\mathbf{X})$,

$$\begin{aligned} E(S^2|\mathbf{X}) &= \text{tr} \left\{ (\mathbf{T} + \boldsymbol{\gamma}\boldsymbol{\gamma}') \frac{\mathbf{X}'_c\mathbf{X}_c + \sum_{i=1}^{N_j} N_j\bar{\mathbf{x}}_j\bar{\mathbf{x}}'_j}{N-1} \right\} \\ &\quad + \sigma^2 - \frac{1}{N-1} \frac{1}{N} \sum_{j=1}^J N_j^2 \text{tr}[\mathbf{T}\bar{\mathbf{x}}_j\bar{\mathbf{x}}'_j] - \frac{N}{N-1} \text{tr}[\boldsymbol{\gamma}\boldsymbol{\gamma}'\bar{\mathbf{x}}\bar{\mathbf{x}}'] \\ &= E(S^2_B|\mathbf{X}) + E(S^2_W|\mathbf{X}). \end{aligned} \tag{A20}$$

Acknowledgment

We thank James LeBreton and two *Organizational Research Methods* anonymous reviewers for providing us with highly constructive and useful feedback that allowed us to improve our manuscript substantially. Also, we thank Kyle Bradley and Harry Joo for their assistance testing the `icc_beta` R package described in our article.

Authors' Note

Both authors contributed equally to this research.

Declaration of Conflicting Interests

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

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

- Aguinis, H. (2004). *Regression analysis for categorical moderators*. New York, NY: Guilford.
- Aguinis, H., Beaty, J. C., Boik, R. J., & Pierce, C. A. (2005). Effect size and power in assessing moderating effects of categorical variables using multiple regression: A 30-year review. *Journal of Applied Psychology, 90*, 94-107.
- Aguinis, H., Gottfredson, R. K., & Culpepper, S. A. (2013). Best-practice recommendations for estimating cross-level interaction effects using multilevel modeling. *Journal of Management, 39*, 1490-1528.
- Aguinis, H., Pierce, C. A., Bosco, F. A., & Muslin, I. S. (2009). First decade of *Organizational Research Methods*: Trends in design, measurement, and data-analysis topics. *Organizational Research Methods, 12*, 69-112.
- Aguinis, H., & Stone-Romero, E. F. (1997). Methodological artifacts in moderated multiple regression and their effects on statistical power. *Journal of Applied Psychology, 82*, 192-206.
- Aguinis, H., Werner, S., Abbott, J. L., Angert, C., Park, J. H., & Kohlhausen, D. (2010). Customer-centric science: Reporting significant research results with rigor, relevance, and practical impact in mind. *Organizational Research Methods, 13*, 515-539.
- Bosco, F. A., Aguinis, H., Singh, K., Field, J. G., & Pierce, C. A. (in press). Correlational effect size benchmarks. *Journal of Applied Psychology*. doi: <http://dx.doi.org/10.1037/a0038047>
- Burstein, L., Linn, R. L., & Capell, F. J. (1978). Analyzing multilevel data in the presence of heterogeneous within-class regressions. *Journal of Educational Statistics, 3*, 347-383.
- Cao, J., & Ramsay, J. O. (2010). Linear mixed-effects modeling by parameter cascading. *Journal of the American Statistical Association, 105*, 365-374.
- Chen, G., Kirkman, B. L., Kanfer, R., Allen, D., & Rosen, B. (2007). A multilevel study of leadership, empowerment, and performance in teams. *Journal of Applied Psychology, 92*, 331-346.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Cortina, J. M., & Landis, R. S. (2009). When small effect sizes tell a big story, and when large effect sizes don't. In C. E. Lance & R. J. Vandenberg (Eds.), *Statistical and methodological myths and urban legends: Doctrine, verity, and fable in the organizational and social sciences* (pp. 287-308). New York, NY: Routledge.
- Crainiceanu, C., & Ruppert, D. (2004). Likelihood ratio tests in linear mixed models with one variance component. *Journal of the Royal Statistical Society-B, 66*, 165-185.
- 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*, 339-362.

- Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods, 12*, 121-138.
- Halbesleben, J. R. B., Wheeler, A. R., & Paustian-Underdahl, S. C. (2013). The impact of furloughs on emotional exhaustion, self-rated performance, and recovery experiences. *Journal of Applied Psychology, 98*, 492-503.
- Heck, R. H., Thomas, S. L., & Tabata, L. N. (2010). *Multilevel and longitudinal modeling with IBM SPSS*. New York, NY: Routledge.
- Hofmann, D. A., & Gavin, M. B. (1998). Centering decisions in hierarchical linear models: Theoretical and methodological implications for organizational science. *Journal of Management, 24*, 623-641.
- Hofmann, D. A., Griffin, M. A., & Gavin, M. B. (2000). The application of hierarchical linear modeling to management research. In K. J. Klein & S. W. J. Kozlowski (Eds.), *Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions* (pp. 467-511). Hoboken, NJ: Jossey-Bass.
- Hofmann, D. A., Morgeson, F. P., & Gerras, S. J. (2003). Climate as a moderator of the relationship between leader-member exchange and content specific citizenship: Safety climate as an exemplar. *Journal of Applied Psychology, 88*, 170-178.
- Hox, J. J. (2010). *Multilevel analysis: Techniques and applications* (2nd ed.). New York, NY: Routledge.
- Hu, J., & Liden, R. C. (2013). Relative leader-member exchange within team contexts: How and when social comparison impacts individual effectiveness. *Personnel Psychology, 66*, 127-172.
- Hülshöger, U. R., Alberts, H. J. E. M., Feinholdt, A., & Lang, J. W. B. (2013). Benefits of mindfulness at work: The role of mindfulness in emotion regulation, emotional exhaustion, and job satisfaction. *Journal of Applied Psychology, 98*, 310-325.
- Kenny, D. A., & Judd, C. M. (1996). A general procedure for the estimation of interdependence. *Psychological Bulletin, 119*, 138-148.
- Kenny, D. A., Korchmaros, J. D., & Bolger, N. (2003). Lower level mediation in multilevel models. *Psychological Methods, 8*, 115-128.
- Kim, E., Bhave, D. P., & Glomb, T. M. (2013). Emotion regulation in workgroups: The roles of demographic diversity and relational work context. *Personnel Psychology, 66*, 613-644.
- LaHuis, D. M., Hartman, M. J., Hakoyama, S., & Clark, P. C. (2014). Explained variance measures for multilevel models. *Organizational Research Methods, 17*, 433-451.
- Leavitt, K., Mitchell, R. R., & Peterson, J. (2010). Theory pruning: Strategies to reduce our dense theoretical landscape. *Organizational Research Methods, 13*, 644-667.
- LeBreton, J. M., & Senter, J. L. (2008). Answers to twenty questions about interrater reliability and interrater agreement. *Organizational Research Methods, 11*, 815-852.
- Liao, H., & Rupp, D. E. (2005). The impact of justice climate and justice orientation on work outcomes: A cross-level multifoci framework. *Journal of Applied Psychology, 90*, 242-256.
- Longford, N. (1993). *Random coefficient modeling*. Oxford, UK: Clarendon.
- Mathieu, J. E., Aguinis, H., Culpepper, S. A., & Chen, G. (2012). Understanding and estimating the power to detect cross-level interaction effects in multilevel modeling. *Journal of Applied Psychology, 97*, 951-966.
- Mathieu, J., Ahearne, M., & Taylor, S. R. (2007). A longitudinal cross-level model of leader and salesperson influences on sales force technology use and performance. *Journal of Applied Psychology, 92*, 528-537.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage Publications.
- Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review, 15*, 351-357.
- Searle, S. R., Casella, G., & McCulloch, C. E. (1992). *Variance components*. New York, NY: Wiley.
- Self, S. G., & Liang, K. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association, 82*, 605-610.
- Short, J. C., Ketchen, D. J., Bennett, N., & du Toit, M. (2006). An examination of firm, industry, and time effects on performance using random coefficients modeling. *Organizational Research Methods, 9*, 259-284.

- Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling* (2nd ed.). Thousand Oaks, CA: Sage.
- Stram, D., & Lee, J.W. (1994). Variance components testing in the longitudinal mixed effects model. *Biometrics*, *50*, 1171-1177.
- Takeuchi, R., Chen, G., & Lepak, D. P. (2009). Through the looking glass of a social system: Cross-level effects of high-performance work systems on employees' attitudes. *Personnel Psychology*, *62*, 1-29.
- Taylor, P. J., Li, W. D., Shi, K., & Borman, W. C. (2008). The transportability of job information across countries. *Personnel Psychology*, *61*, 69-111.
- Uy, M. A., Foo, M. D., & Aguinis, H. (2010). Using experience sampling methodology to advance entrepreneurship theory and research. *Organizational Research Methods*, *13*, 31-54.

Author Biographies

Herman Aguinis (<http://mypage.iu.edu/~haguinis>) is the John F. Mee Chair of Management and Founding Director of the Institute for Global Organizational Effectiveness in the Kelley School of Business, Indiana University. His multi-disciplinary, multi-method, and multi-level research addresses human capital acquisition, development, and deployment, and research methods and analysis. He has published five books and more than 120 articles in refereed journals. He is a Fellow of the Academy of Management, former editor-in-chief of *Organizational Research Methods*, and received the 2012 Academy of Management Research Methods Division Distinguished Career Award for lifetime contributions.

Steven Andrew Culpepper (<http://publish.illinois.edu/sculpepper/>) is an assistant professor in the Department of Statistics at the University of Illinois at Urbana-Champaign. He completed a doctorate in educational psychology from the University of Minnesota in 2006. His research focuses on statistical methods in the social sciences and includes the development of new methodologies, evaluation of existing procedures, and application of novel statistical techniques to substantive questions in demography, education, management, and psychology.