Explained Variance Measures for Multilevel Models

**David M. LaHuis[1], Michael J. Hartman[1],
Shotaro Hakoyama[1], and Patrick C. Clark[1]**

Abstract
One challenge in using multilevel models is determining how to report the amount of explained variance. In multilevel models, explained variance can be reported for each level or for the total model. Existing measures have been based primarily on the reduction of variance components across models. However, these measures have not been reported consistently because they have some undesirable properties. The present study is one of the first to evaluate the accuracy of these measures using Monte Carlo simulations. In addition, a measure based on the full partitioning of variance in multilevel models was examined. With the exception of the Level 2 explained variance measure, all other measures performed well across our simulated conditions.

Keywords
multilevel, effect size, Monte Carlo

Multilevel modeling (MLM) has become an important tool for organizational scientists because it allows researchers to analyze data that reside at different levels. Organizations are inherently hierarchically structured. For example, employees work in teams, supervisors manage teams, and supervisors work in organizations that have their own unique culture. It is important to consider both individual-level variables and contextual variables when analyzing organizational relationships. Ignoring the influence of contextual variables discards important pieces of the puzzle of organizational behavior and may lead to erroneous conclusions about both micro- and macro-level relationships (Bliese & Hanges, 2004). In the past two decades, a number of articles (Bliese & Ployhart, 2002; Hofmann & Gavin, 1998; Kreft, de Leeuw, & Aiken, 1995; LaHuis & Ferguson, 2009; Maas & Hox, 2005), book chapters (Bliese, 2000; Long, 2012), and books (Hox, 2010; Klein & Kozlowski, 2000; Raudenbush & Bryk, 2002; Snijders & Bosker, 2012) have been dedicated to understanding both technical issues and the application of MLM in an organizational context. Researchers have utilized MLM to study topics such as organizational safety (Bacharach,

---

[1]Department of Psychology, Wright State University, Dayton, OH, USA

Corresponding Author:
David M. LaHuis, Department of Psychology, Wright State University, Dayton, OH 45435, USA.
Email: david.lahuis@wright.edu

Bamberger, & Doveh, 2008; Neal & Griffin, 2006), leadership and management (Atwater, Wang, Smither, & Fleenor, 2009; Chen, Kirkman, Kanfer, Allen, & Rosen, 2007), teams (Dierdorff, Rubin, & Morgeson, 2009), stress (Liu, Wang, Zhan, & Shi, 2009), and rater effects (Ng, Koh, Ang, Kennedy, & Chan, 2011). Furthermore, the use of MLM in organizational research is rising. Mathieu, Aguinis, Culpepper, and Chen (2012) conducted a thorough review of multilevel articles published in the *Journal of Applied Psychology* from 2002 to 2010. They found that there was an average of three published articles using MLM per year between 2000 and 2002. This number more than quadrupled to an average of 13 articles per year between 2008 and 2010. There were similar trends in other prominent journals, such as *Organizational Research Methods* (Aguinis, Pierce, Bosco, & Muslin, 2009). However, MLM is a relatively new technique and there are still some technical issues that require the attention of researchers. Due to the ubiquitous nesting structure in organizations and the rising popularity of MLM, it is important for researchers to address these issues.

One challenge facing researchers using multilevel models is how to estimate and report effect sizes. Reporting effect sizes is important because they are useful in theory building and testing (Aguinis et al., 2010) and communicating results to others from different professions (Thompson, 2002). The most common effect size in multilevel modeling is explained variance (R^2). However, the presence of multiple variance components in multilevel models complicates the estimation of these statistics. For example, in a basic model containing one Level 1 predictor, one Level 2 predictor, and a cross-level interaction, there are potentially three variance components (e.g., within-group variance, intercept variance, and slope variance) that a researcher can explain. Although a number of explained variance measures for multilevel models exist, there are some issues with these statistics. First, it is possible to find negative values for some of the measures. That is, adding predictors may actually decrease the amount of explained variance. In addition, the accuracy of these measures is unknown. The result of these issues is that researchers often do not report the amount of explained variance in multilevel studies. For example, we searched the 10 most prestigious journals identified by Zickar and Highhouse (2001) using the terms *multilevel* and *hierarchical linear modeling* and found 126 multilevel articles.[1] Of these, only 76 (60%) contained some measure of explained variance. This is unfortunate given the importance of effect sizes and the known issues of lower power for some of the relationships in these models (Mathieu et al., 2012).

In the present study, we evaluate the accuracy of the existing measures of explained variance for multilevel models across a number of conditions. To our knowledge, this is the first study to examine the performance of the various effect size indices. Furthermore, we identify some critical issues researchers face when using these measures and seek to provide guidance on these issues based on Monte Carlo simulations. Our hope is that the current study will facilitate the use of explained variance measures in multilevel research. Before discussing effect size measures and the issues that may arise when using them, we will briefly describe a basic multilevel model to introduce concepts for those new to MLM.

Multilevel Models

For simplicity, we focus on a basic two-level multilevel model that includes two Level 1 predictors (*X1*, *X2*), one Level 2 predictor (*Z*), and a cross-level interaction (*X1* × *Z*) predicting a Level 1 outcome (*Y*). It is often useful to conceptualize multilevel models as equations at different levels. The Level 1 equation regresses the outcome onto the Level 1 predictors:

$$Y_{ij} = \beta_{0j} + \beta_{1j}(X1_{ij}) + \beta_{2j}(X2_{ij}) + R_{ij}. \tag{1}$$

In this equation, β_{0j} is the intercept. β_{1j} and β_{2j} are the regression slopes predicting Y_{ij} using $X1_{ij}$ and $X2_{ij}$, respectively. The Level 1 error is R_{ij} and its variance (σ^2) represents within-group variance

not explained by the model. The subscripts i and j refer to the Level 1 and 2 units, respectively. The Level 2 equations model variance in β_{0j}, β_{1j}, and β_{2j}:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(Z_j) + U_{0j} \tag{2}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}(Z_j) + U_{1j} \tag{3}$$

$$\beta_{2j} = \gamma_{20} + U_{2j}, \tag{4}$$

where Z_j is a Level 2 predictor, the γs are Level 2 regression coefficients, and U_{0j}, U_{1j}, and U_{2j} represent Level 2 errors. The regression coefficients are referred to as fixed effects; the Level 1 and Level 2 error terms are random effects.

A succinct way of expressing more complicated models in a single equation is:

$$Y_{ij} = \gamma_{00} + \sum_{h=1}^{p} \gamma_{h0} X_{hij} + \sum_{k=1}^{q} \gamma_{0k} Z_{kj} + \sum_{k=1}^{q} \sum_{h=1}^{p} \gamma_{hk} Z_{kj} X_{hij} + U_{0j} + \sum_{h=1}^{p} U_{hj} X_{hij} + R_{ij}. \tag{5}$$

In Equation 5, there are p Level 1 predictors and q Level 2 predictors, and it is assumed that the Level 1 predictors X_{hij} do not correlate with the Level 2 predictors Z_{kj}. There are $p + 1$ random parameters because there is one random intercept term and every Level 1 predictor has a random slope. In practice, researchers may only wish to specify random slopes for a subset of Level 1 predictors. The first four terms contain the fixed part of the model and contribute to the prediction of the outcome. The last three terms contain the random part of the model. The variance of U_{0j} is τ_{00} and represents intercept variance not explained by Z_j. Similarly, the variances of U_{hj} are τ_{hh} and represent the unexplained variance in the Level 1 slopes. It is useful to predict unexplained Level 1 slope variance with a Level 2 predictor, which is referred to as a cross-level interaction. It is common to allow the intercept and slopes to covary. A positive covariance between a random intercept and random slope means that groups with higher intercept values have stronger within-group effects, whereas a negative covariance suggests groups with higher intercept values have weaker within-group effects.

As previously mentioned, the presence of multiple variance components complicates the calculation of explained variance for multilevel models. In traditional ordinary least squares (OLS) regression, explained variance can be thought of as a ratio of explained variance to total variance. However, in multilevel models, the total variance is partitioned into Level 1 and Level 2 variance components.

Explained Variance Measures for Multilevel Models

For the most part, organizational researchers have used multilevel explained variance measures based on the reduction in variance components when adding predictors. These measures compare variance components from different models. A disadvantage of these measures is that they can result in negative explained variance estimates because of the way fixed effects and variance components are estimated. A few researchers (Hofmann, Morgeson, & Gerras, 2003; Vandenberghe et al., 2007; Wallace, Edwards, Arnold, Frazier, & Finch, 2009) reported explained variance from an OLS regression analysis, which is based on a variance partitioning approach. Although the OLS measure ignores the multilevel structure of the data, this will provide an unbiased estimate of explained variance because the fixed-effects estimates are typically estimated without bias. More recently, Nakagawa and Schielzeth (2013) proposed an explained variance measure based on variance partitioning that retains the multilevel structure of the model. In the following section, we describe the indices based on each approach in more detail.

Reduction in Variance Components-Based Measures

One of the first explained variance measures for multilevel models is based on the reduction of unexplained variance when predictors are added (Bryk & Raudenbush, 1992). These approximate R^2 values can be calculated at each level by first estimating the null model to determine the amount of variance at each level. Next, predictors are added to the model, which allows the researcher to compare the residual variance components to those from the null model. Level 1 explained variance can be calculated using the formula:

$$R_1^2(approx.) = \frac{(\sigma_{null}^2 - \sigma_{full}^2)}{\sigma_{null}^2}. \tag{6}$$

Similarly, Level 2 explained variance is computed using

$$R_2^2(approx.) = \frac{(\tau_{00/null} - \tau_{00/full})}{\tau_{00/null}}. \tag{7}$$

A major advantage of these measures is that they provide intuitive simple measures of explained variance based on the logic of partitioning unexplained variance in two-level models. However, they can be problematic because they can result in negative values. That is, adding predictors may actually increase the amount of unexplained variance. This can occur because the between-group variance is a function of both Level 1 and Level 2 variances. Specifically, Snijders and Bosker (1994) show that

$$var(\overline{Y}_j) = \tau_{00} + \frac{\sigma^2}{n_j}, \tag{8}$$

where n_j is the group size. Thus, if a Level 1 predictor is added and σ^2 decreases, but the between-group variability is not affected, then τ_{00} must increase to balance out the decrease in σ^2. That is, adding a Level 1 predictor would increase the unexplained Level 2 variance.

Snijders and Bosker (1994) offered alternative measures to address this issue of negative explained variance estimates. They originally described measures for Level 1 and Level 2 R^2s; however, according to our review only the Level 1 version has been used because it reflects the total amount of explained variance. The Level 2 measure did not have a straightforward interpretation. Thus, we focused on the Level 1 version. They described explained variance as the proportional reduction in mean squared prediction error. To address the issue of the interplay between τ_{00} and σ^2, they used both to calculate explained variance. One measure provides an estimate of a total R^2 by using the σ^2s and τ_{00}s from null and full models using the formula

$$R^2(S\&B) = 1 - \frac{(\sigma_{full}^2 + \tau_{00full})}{(\sigma_{null}^2 + \tau_{00null})}. \tag{9}$$

$R^2(S\&B)$ considers the interplay of Level 1 and Level 2 variance components, and as such, population estimates of these cannot be negative (although sample estimates can be negative).

Equations 6, 7, and 9 are appropriate for random intercept models. When random slopes are present, the formulas for partitioning unexplained variance are more complicated as described below. Snijders and Bosker (1994) provided a formula for calculating $R^2(S\&B)$ with random slopes. However, they note that adding random slopes does not aid in predicting the outcome so incorporating those formulas will not result in important differences in explained variance. Thus, they suggest that researchers can estimate a random-intercept model to calculate explained variance and ignore random slopes (Snijders & Bosker, 2012). In the present study, we evaluate this suggestion empirically.

Variance Partitioning–Based Measures

Several authors (Hofmann et al., 2003; Vandenberghe et al., 2007; Wallace et al., 2009) have used explained variances estimates from an OLS regression equation where Level 2 variables are disaggregated down to Level 1 units. $R^2(OLS)$ can be expressed in many ways. In the present context, it is useful to consider $R^2(OLS)$ in the terms of how variance in the outcome is partitioned. In OLS, the regression variance of Y is partitioned into the variance of predicted scores $(\hat{Y}_i)$ and the variance of residuals. Predicted scores are simply the intercept added to the sum of the products of the predictor scores and their respective unstandardized regression coefficient. Thus, $R^2(OLS)$ can be calculated using the following equation:

$$R^2(OLS) = \frac{\text{var}(\hat{Y}_i)}{\text{var}(\hat{Y}_i) + \sigma^2}. \tag{10}$$

Equation 10 is the ratio of explained variance to total variance. Although OLS will result in biased standard errors for regression coefficients in multilevel contexts (Bliese & Hanges, 2004), it will usually produce unbiased estimates for the regression coefficients themselves. This will result in $R^2(OLS)$ being unbiased.

Nakagawa and Schielzeth (2013) proposed an R^2 measure that considers the full partitioning of variance for multilevel models as described by Snijders and Bosker (2012). To simplify the presentation, we refer to the variance of predicted scores as $\text{var}(\hat{Y}_{ij})$. Predicted scores are computed by summing the product of the fixed effects and their respected predictors as with OLS. Snijders and Bosker (2012) show that without random slopes, variance in Y can be expressed using

$$\text{var}(Y_{ij}) = \text{var}(\hat{Y}_{ij}) + \tau_{00} + \sigma^2. \tag{11}$$

The last two terms are the unexplained variances at Level 2 and Level 1 as defined previously. From this equation, an explained variance measure based on multilevel variance partitioning $R^2(MVP)$ can be calculated. The equation is

$$R^2(MVP) = \frac{\text{var}(\hat{Y}_{ij})}{\text{var}(\hat{Y}_{ij}) + \tau_{00} + \sigma^2}. \tag{12}$$

Put simply, $R^2(MVP)$ is the variance of the predicted scores divided by the variance of the predicted scores plus the Level 1 and Level 2 variance components. This is equivalent to the R^2 measure proposed by Nakagawa and Schielzeth (2013).

Nakagawa and Schielzeth (2013) did not consider models with random slopes. Consistent with Snijders and Bosker's (2012) recommendation, they suggested simply basing the explained variance on a model without random slopes because they generally do not affect the amount of explained variance. However, incorporating random slopes in the partitioning variance in Y_{ij} is straightforward. We again simplify the presentation somewhat by using $\text{var}(\hat{Y}_{ij})$ to refer to the variance of predicted values with the understanding that predicted scores are based on the direct effects of the Level 1 and 2 predictors and any cross-level interactions. Snijders and Bosker (2012, p. 116) show that in this case the variance in Y_{ij} can then be partitioned as:

$$\text{var}(Y_{ij}) = \text{var}(\hat{Y}_{ij}) + \tau_{00} + 2\mu'_{X(q)}T_{10} + \mu'_{X(q)}T_{11}\mu_{X(q)} + trace(T_{11}\Sigma_{X(p)}) + \sigma^2, \tag{13}$$

where $\mu_{X(q)}$ is a vector of predictor means, T_{10} is a vector of intercept-slope covariances, and T_{11} is the covariance matrix of random slopes. The term $trace(T_{11}\Sigma_{x(p)})$ represents the contribution of the random slopes to the variance in Y_{ij}, and it is the sum of the diagonals of the resulting matrix when the slope covariance matrix T_{11} is postmultiplied by the covariance matrix of predictors with

random slopes. Based on Equation 13, an explained variance measure for models with random slopes can be calculated using:

$$R^2(MVP) = \frac{\text{var}(\hat{Y}_{ij})}{\text{var}(\hat{Y}_{ij}) + \tau_{00} + 2\mu'_{X(q)}T_{10} + \mu'_{X(q)}T_{11}\mu_{X(q)} + trace(T_{11}\Sigma_{X(p)}) + \sigma^2}. \qquad (14)$$

Researchers may find Equation 14 somewhat daunting. However, the statistical program R provides the necessary components. As with $R^2(S\&B)$, researchers could reestimate a model without random slopes and use Equation 12 to estimate explained variance. The result should be similar to estimates produced by Equation 14 because the contribution of random slopes to the variance of Y_{ij} is typically small. In our Monte Carlo simulations, we included both approaches for comparison.

Potential Issues in Reporting Explained Variance in Multilevel Models

We see three major issues that researchers may encounter when reporting explained variance in multilevel models. The first issue concerns what measure to report. This will likely depend on if researchers are interested in explaining level-specific or total model variance. $R_1^2(approx.)$ and $R_2^2(approx.)$ are the only measures that describe level-specific explained variance, while $R^2(S\&B)$, $R^2(OLS)$, and $R^2(MVP)$ describe total explained variance. The latter three can estimate the amount of total variance Level 1 or Level 2 predictors explain using a hierarchical approach. The choice will also depend on the accuracy of these measures, which to our knowledge, has not been assessed empirically. Based on our literature review of the studies that reported explained variance, the majority (55%) reported either $R_1^2(approx.)$ or $R_2^2(approx.)$. The next most common was $R^2(S\&B)$ (20%) followed by $R^2(OLS)$ (3%). $R^2(MVP)$ has not yet been applied in organizational research. It would be interesting to determine if the measures reported most often are in fact the most accurate. To address this, we conducted a series of Monte Carlo simulations to investigate the accuracy these measures.

A second issue is how to deal with random slopes. Researchers may either ignore slope variance or incorporate it in the equations. Snijders and Bosker (2012) and Nakagawa and Schielzeth (2013) advocate ignoring slope variance and to base explained variance calculation on the random intercept model. However, to our knowledge this suggestion has not been tested empirically under a variety of conditions. We compared these two approaches using Monte Carlo simulations to assess if one of these approaches should be preferred.

A final issue is how to compute and interpret the change in explained variance between two nested models. In OLS, this is calculated by taking the difference between the R^2s produced by the two models. An alternative approach for the explained variance measures based on the reduction in variance components is to compute explained variance using the variance component estimates. It is important to note that the two approaches provide estimates of different things. For example, if Model 1 and Model 2 were nested models, we could compute a change in $R_2^2(approx.)$ by first computing it for each model and then taking the difference. The second option would be to compute $R_2^2(approx.)$ using the variance components for the two models. The difference between the two approaches can be explained by describing the first approach as a single equation:

$$\Delta R_2^2 = \frac{\tau_{00null} - \tau_{00M2}}{\tau_{00null}} - \frac{\tau_{00null} - \tau_{00M1}}{\tau_{00null}} = \frac{\tau_{00M1} - \tau_{00M2}}{\tau_{00null}}. \qquad (15)$$

Thus, the first approach compares the change in variance components from Models 1 and 2 relative to the variance component from the null model. The latter approach compares the change in variance components against the Model 1 variance and is simply the ratio of the difference between

Model 1 and Model 2 variance to the variance component from Model 1. Thus, the first approach expresses the amount of incremental explained variance relative to the total variance. This is consistent with the traditional interpretation of change in R^2. The second approach provides the amount of explained variance relative to the Model 1 residual variance. This applies to explained variance measures based on the reduction in variance components ($R_1^2(approx.)$, $R_2^2(approx.)$, and $R^2(S\&B)$). There is still some question as to how accurate the various measures would be in measuring changes in explained variance. In our simulations, we demonstrate how to compute and interpret change in R^2 for the various measures. Furthermore, we assess the accuracy of change in explained variance across several multilevel models.

The Present Study

The purpose of the current study is to shed light on the three issues we believe researchers are likely to encounter when calculating and reporting explained variance measures in multilevel models. The accuracy of these measures plays an integral role in resolving each issue. It is not clear how accurate any of the explained variance measures are or how they are affected by various characteristics of multilevel studies. To address this, we evaluated the performance of the explained variance measures using three criteria: unbiasedness, consistency, and efficiency. An estimator is unbiased when the expected value of the estimator equals the population value. In our case, an explained variance measure would be unbiased if the mean across samples equaled the population R^2. We adopted Hoogland and Boomsma's (1998) criterion of .05 or less for acceptable parameter estimate bias. A consistent estimator is one whose estimates converge toward the population value as sample size increases. In the present study, an explained variance measure is consistent if it exhibits less bias as sample size increases. In the multilevel context, sample size refers to the number of groups, group size, or both. We evaluate consistency against all three criteria. Efficiency refers to the precision of an estimator. In the present study, we evaluated efficiency by computing the standard deviation of point estimates for each condition. These standard deviations can essentially be interpreted as standard errors because our generated data represent a population. Finally, effect size measures should be scaled properly given the question of interest (Kelley & Preacher, 2012). All of the measures in the present study seek to quantify the percentage of variance explained by the predictors. Thus, they should have a minimum of zero and a maximum of one. Negative estimates greatly complicate the interpretation of explained variance and of change in explained variance. We tracked the percentage of times an R^2 index returned a negative value.

Method

Design

To evaluate the various explained variance measures, we conducted two sets of Monte Carlo simulations that examined different generating multilevel models. We chose these models because they contain the possible effects one might encounter in multilevel models (cross-level effects, cross-level interactions, etc.). In our first set of simulations, we used a random-intercept model with two Level 1 predictors and one Level 2 predictor. In the second set of simulations, we added a random slope for a Level 1 predictor and a cross-level interaction to the previous model.

For both sets of simulations, we manipulated the number of groups, average group size, the intraclass correlation (ICC) in the outcome, and the amount of explained variance. We based these values on our literature review of multilevel articles in the organizational literature. For each article, we recorded the number of groups, group size, parameter estimates for fixed and random effects, and whether or not explained variance was reported. We derived the number for groups, average group sizes, and R^2 values based on the 25th, 50th, and 75th percentiles of those reported in the studies.

Table 1. Data Generation for the Random-Intercept Model Simulations.

	Level 1 Covariance Matrix		
	1	2	3
1. Y	1		
2. X1	$\sqrt{\frac{.91(R^2)}{1.4}}$	1	
3. X2	$\sqrt{\frac{.91(R^2)}{1.4}}$	.3	1

	Level 2 Covariance Matrix			
	1	2	3	4
1. Y	τ_{00}			
2. Z	$\sqrt{R^2(\tau_{00})}$	1		

Note: τ_{00} equaled approximately .176 for the ICCY = .15 condition and .429 for the ICCY = .30 condition.

Based on these results, we defined the number of groups to be 40, 70, or 160. We used unbalanced group sizes with average group sizes of 4, 8, or 21. Sizes for each group were sampled randomly from a vector ranging from 2 to 27 until the desired number of groups and average group sizes were reached. The R^2 values for each level were .1, .3, or .5 and reflected a broad range of explained variance values. We set the ICC for Y to either .15 or .3. In the random-intercept model simulations, we set the correlation between Level 1 predictors to .3. To simplify the data generation process for the random-slope models, we set the correlation between Level 1 predictors to 0.

Data Generation

For the random-intercept models, we generated data for basic two-level multilevel models using the software program R. Essentially, we specified Level 1 and Level 2 population correlation matrices resulting from the specified model and sampled randomly from the multivariate normal distribution using the mvrnorm command in R. The R code used for the data generation is available from the first author.

For the random-intercept model, the equations for the underlying model were

$$Y_{ij} = \beta_{0j} + \beta_{1j}(X1_{ij}) + \beta_{2j}(X2_{ij}) + R_{ij} \tag{16}$$

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(Z_j) + U_{0j} \tag{17}$$

$$\beta_{1j} = \gamma_{10} \tag{18}$$

$$\beta_{2j} = \gamma_{20}. \tag{19}$$

Table 1 presents the equations used to generate the correlation matrices. We generated the correlations between the outcome and predictors to correspond to the desired R^2 values at Level 1 based on the equations. That is, these values changed systematically to produce the desired population Level 1 R^2. A similar procedure was used for the Level 2 variable. We used the mvrnorm command and specified the population correlation matrix at Level 2. This included correlations between the between-group variance of the outcome and the Level 2 predictor. Again, the actual values of the correlation matrix varied to produce the desired between-group variances in the outcome and the

predictors, as well as the desired Level 2 R^2. We then combined the Level 1 and Level 2 data sets to create the full multilevel data set.

For the random-slope model, we used a different procedure to generate the data because specifying population Level 1 and Level 2 correlations is considerably more complicated with random-slope models. Instead, we generated the data directly from the equation:

$$Y_{ij} = \sqrt{\frac{R^2}{3}}(X1_{ij}) + \sqrt{\frac{R^2}{3}}(X2_{ij}) + \sqrt{\frac{R^2}{3}}(X1_{ij})(Z_j) + U_{0j} + U_{1j}(X1_{ij}) + R_{ij}. \tag{20}$$

Values for $X1$, $X2$, and Z were drawn randomly from standard normal distributions. Values for U_{0j} were drawn from a random normal distribution with a mean of 0 and a standard deviation equal to $\sqrt{\tau_{00}}$. Values for U_{1j} were sampled from a random normal distribution with a mean of zero and a standard deviation of .1. The standard deviation was based on Mathieu et al. (2012). As shown in Equation 20, we set the effect of Z on the intercept at zero; as such, no Level 2 variance was explained. We specified the two Level 1 predictors and the cross-level interaction as explaining one third of the Level 1 R^2. That is, the coefficients for these equaled the square root of one third of the desired R^2.

Analyses

We analyzed the data using restricted maximum likelihood estimation as implemented in the LME4 package in R. A null model was first estimated to obtain baseline σ^2 and τ_{00} variance components. Next, we estimated the reduced and full models to obtain their respective variance components. We then calculated the various R^2 measures using the appropriate equations. We assessed the accuracy of the effect size measures by calculating mean bias and standard deviation of the point estimates. Bias was the mean difference between the estimated R_S^2 and , where R_S^2 is the effect size for sample s and R_P^2 is the appropriate population value for that measure. For incremental variance scenarios, we calculated ΔR_S^2 and ΔR_P^2. For the level-specific measures, the population value was the amount of variance explained at the level and for the total explained variance measures the population value was the amount of total variance explained. This provided an indication of whether or not the measures under- or overestimated the amount of explained variance. We determined consistency by simply examining whether bias decreased as sample size increased. We calculated the standard deviation of point estimates within each condition across 1,000 samples. This provided an estimate of the efficiency of the R^2 measures.

Results

Accuracy of R^2 Measures

Random-Intercept Models. Table 2 presents the overall results for the random-intercept and random-slope models. For the random-intercept model, all measures displayed acceptable levels of bias with values ranging from –0.011 to 0.006 with $R_2^2(approx.)$ performing the worst. $R^2(S\&B)$ (bias = –0.001) and $R_1^2(approx.)$(bias = –0.001) were the most accurate, followed by $R^2(MVP)$ (bias = 0.003). In terms of efficiency, $R_2^2(approx.)$ had the highest standard deviations of 0.157, indicating that this measure was the least precise estimator. The remaining measures produced very similar results with standard deviations ranging from 0.032 to 0.036. $R_1^2(approx.)$ and $R^2(MVP)$ were the most precise, with values of 0.032 and 0.034, respectively. We tracked the percentage of times each measure resulted in negative values. It is important to note that for all of the measures in random-intercept models the population R^2 was at least .10. We found negative effect size values

Table 2. Overall Results.

	$R_1^2(approx.)$	$R_2^2(approx.)$	$R^2(S\&B)$	$R^2(OLS)$	$R^2(MVP)$
Random intercept					
Bias	−0.001	−0.011	−0.001	0.006	0.003
SD	0.032	0.157	0.036	0.036	0.034
Random slope with cross-level interaction					
Bias	0.008 (−0.006)	−0.042 (−0.014)	−0.004 (−0.003)	0.008	0.004 (0.004)
SD	0.042 (0.041)	0.149 (0.149)	0.039 (0.034)	0.037	0.035 (0.035)
Random slope without cross-level interaction					
Bias	0.104 (−0.002)	−0.042 (−0.009)	−0.009 (−0.001)	0.010(−)	0.009 (0.007)
SD	0.045 (0.039)	0.159 (0.139)	0.039 (0.035)	0.034(−)	0.031 (0.032)

Note: Estimates averaged across 54,000 samples for random-intercept conditions and across 54,000 samples for random-slope conditions. Values in parentheses are based on estimates using only the random-intercept model. There is only one value for $R^2(OLS)$ because it is based on a disaggregated version of the multilevel model that does not consider Level 2 variance.

for $R_1^2(approx.)(<1\%)$, $R_2^2(approx.)(8\%)$, and $R^2(S\&B)(<1\%)$. We did not find negative values for $R^2(OLS)$ or $R^2(MVP)$.

Table 3 presents the average bias and standard deviation values as a function of the conditions we manipulated for the random-intercept models. All measures were consistent in that they exhibited less bias as group size increased. The amount of between-group variance in the outcome had little influence on bias. There was more bias as population R^2 increased for all of the measures except $R^2(OLS)$ and $R^2(MVP)$, which became more accurate as the size of the effect increased.

The overall results for the random-slope models are also displayed in Table 2. $R_1^2(approx.)$(bias = 0.008), $R^2(S\&B)$(bias = −0.004), $R^2(OLS)$ (bias = 0.008), and $R^2(MVP)$ (bias = 0.004) produced acceptable levels of bias. In contrast, the Level 2 explained variance measure did not perform as well. $R_2^2(approx.)$ resulted in an acceptable level of average bias (−0.042). However, the negative bias suggests that on average $R_2^2(approx.)$ produced negative estimates, as the Level 2 population R^2 was 0. The efficiency of the indices was very similar to those for the random-intercept model. $R_2^2(approx.)$ had the highest standard deviation values of 0.149. $R^2(MVP)$ ($SD = 0.035$) and $R^2(OLS)$ ($SD = 0.037$) had the lowest values. $R^2(S\&B)$ and $R_1^2(approx.)$ produced slightly higher SD values of 0.038 and 0.041, respectively. We found negative effect size values for $R_1^2(approx.)(<1\%)$, $R_2^2(approx.)(51\%)$, and $R^2(S\&B)(<1\%)$. The frequency of negative values for the Level 2 measures can partially be explained by the fact that the population value for Level 2 R^2 was 0. We did not find negative values for $R^2(OLS)$ or $R^2(MVP)$.

Table 4 presents the bias and standard deviation values as a function of the manipulated conditions for random-slope models. In general, the manipulated conditions had little impact on bias and standard deviation values. For the most part, the measures were consistent and produced less bias as the sample size increased.

Random-Slope Models. We were also interested in how researchers should deal with random slope variance. One solution is to incorporate it into the relevant formulas. Another solution is to base explained variance off a random-intercept model where the slopes are fixed. To compare these two approaches, we estimated two sets of models. One is the random-slope model with two Level 1 predictors and a cross-level interaction reported previously. The other random-slope model omitted the cross-level interaction thereby resulting in more unexplained slope variance. We used two different approaches to calculate explained variance for each model. The first approach uses the parameters produced by the random-slope model to calculate the various explained variance estimates. The second approach uses the parameter estimates derived from the random-intercept version of the two

Table 3. Simulation Results for the Random-Intercept Model Simulations.

Bias

	$R_1^2(approx.)$	$R_2^2(approx.)$	$R^2(S\&B)$	$R^2(OLS)$	$R^2(MVP)$
NGS					
40	−0.003	−0.017	−0.002	0.010	0.004
70	−0.001	−0.009	−0.001	0.006	0.003
160	−0.000	−0.007	−0.001	0.002	0.001
NJ					
4	−0.004	−0.006	−0.001	0.009	0.005
8	−0.001	−0.017	−0.001	0.006	0.002
21	−0.000	−0.009	−0.001	0.003	0.001
R^2					
.10	−0.001	−0.002	−0.000	0.009	0.007
.30	−0.002	−0.013	−0.001	0.006	0.002
.50	−0.002	−0.018	−0.002	0.003	−0.001
ICC Y					
.15	−0.002	−0.009	−0.001	0.006	0.003
.30	−0.001	−0.013	−0.002	0.006	0.003

SD

	$R_1^2(approx.)$	$R_2^2(approx.)$	$R^2(S\&B)$	$R^2(OLS)$	$R^2(MVP)$
NGS					
40	0.048	0.220	0.046	0.045	0.043
70	0.031	0.148	0.035	0.036	0.033
160	0.018	0.103	0.026	0.026	0.025
NJ					
4	0.043	0.217	0.048	0.047	0.045
8	0.032	0.156	0.036	0.036	0.034
21	0.021	0.099	0.024	0.024	0.023
R^2					
.10	0.032	0.155	0.036	0.036	0.033
.30	0.030	0.148	0.033	0.033	0.032
.50	0.035	0.168	0.038	0.038	0.037
ICC Y					
.15	0.030	0.159	0.032	0.032	0.030
.30	0.034	0.155	0.039	0.039	0.037

Note: NGS = number of groups; NJ = cases per group; Estimates averaged across 18,000 samples for NGS, NJ, and R^2 conditions and across 27,000 samples for ICC Y conditions.

models. As shown in Table 2, both approaches performed well in terms of bias and efficiency for most of the measures when the cross-level interaction was included. However, when the cross-level interaction was not included $R_1^2(approx.)$ did not perform well when based on the random-slope model but did perform well when based on a random-intercept model. $R_2^2(approx.)$ had large standard deviations for both approaches and both models. This is not surprising because these measures do not incorporate the effects of slope variance.

Changes in R^2. Table 5 presents the results for the changes in explained variance analyses. We first assessed the change in explained variance when a Level 1 predictor is added. We accomplished this by comparing the full random-intercept model previously described to a reduced model where one of

Table 4. Simulation Results for the Random-Slopes Model Simulations.

Bias

	$R_1^2(approx.)$	$R_2^2(approx.)$	$R^2(S\&B)$	$R^2(OLS)$	$R^2(MVP)$
NGS					
40	−0.008	−0.018	−0.005	0.015	0.007
70	−0.006	−0.015	−0.004	0.008	0.003
160	−0.004	−0.009	−0.003	0.002	0.000
NJ					
4	−0.009	−0.018	−0.006	0.012	0.005
8	−0.005	−0.019	−0.003	0.009	0.004
21	−0.004	−0.006	−0.003	0.004	0.002
R^2					
.1	−0.002	−0.004	−0.002	0.011	0.008
.3	−0.006	−0.016	−0.004	0.009	0.004
.5	−0.010	−0.022	−0.006	0.005	−0.001
ICC Y					
.15	−0.006	−0.015	−0.005	0.006	0.002
.30	−0.005	−0.013	−0.003	0.011	0.005

SD

	$R_1^2(approx.)$	$R_2^2(approx.)$	$R^2(S\&B)$	$R^2(OLS)$	$R^2(MVP)$
NGS					
40	0.043	0.217	0.051	0.050	0.047
70	0.032	0.156	0.038	0.037	0.035
160	0.021	0.099	0.025	0.025	0.023
NJ					
4	0.055	0.223	0.049	0.048	0.045
8	0.040	0.147	0.038	0.037	0.035
21	0.027	0.071	0.027	0.027	0.026
R^2					
.1	0.038	0.121	0.034	0.033	0.031
.3	0.038	0.155	0.036	0.035	0.033
.5	0.046	0.172	0.044	0.044	0.041
ICC Y					
.15	0.035	0.133	0.032	0.031	0.029
.30	0.047	0.166	0.045	0.044	0.041

Note: NGS = number of groups; NJ = cases per group; estimates averaged across 36,000 samples for NGS, NJ, and R^2 conditions and across 54,000 samples for ICC Y and ICC X1 and X2 conditions.

the Level 1 predictors is omitted. Thus, the explained Level 1 variance in the reduced model is half that of the full model. For example, for an R^2 of .3 and an ICC of .15, the difference in Level 1 explained variance between the full and reduced models was .15. This would be the population value for change in $R_1^2(approx.)$. We did not assess changes in $R_2^2(approx.)$ because we would not expect that to change and as such would not be of interest to most researchers. For the total explained variance measures ($R^2(S\&B)$, $R^2(OLS)$, or $R^2(MVP)$) the population value would be approximately .13 because of the unequal Level 1 and Level 2 variances. As shown in Table 5, all of the measures performed well in assessing changes in R^2 when a Level 1 predictor was added.

We also assessed changes in R^2 due to adding a Level 2 predictor for random-intercept models. This was accomplished by comparing the full model to a reduced model where the Level 2 predictor

Table 5. Results for Changes in R^2.

	$R_1^2(approx.)$	$R_2^2(approx.)$	$R^2(S\&B)$	$R^2(OLS)$	$R^2(MVP)$
Adding a Level 1 predictor					
Bias	0.000	—	0.000	0.001	0.000
SD	0.023	—	0.019	0.019	0.018
Adding a Level 2 predictor					
Bias	–	0.004	0.000	0.003	0.001
SD	–	0.136	0.029	0.030	0.029
Adding a cross-level interaction					
Bias	−0.004	–	0.005	−0.002	−0.005
SD	0.030	–	0.029	0.024	0.025

Note: Estimates averaged across 54,000 samples for random-intercept conditions and across 54,000 samples for random-slope conditions.

Z was omitted. This resulted in an expected change in $R_2^2(approx.)$ equal to the amount of explained variance at Level 2. For example, for an R^2 of .3 and an ICC of .15, the difference in Level 2 explained variance between the full and reduced models was .3. We did not assess changes in $R_1^2(approx.)$ because we would not expect that to change and as such, would not be of interest to most researchers. For the total explained variance measures ($R^2(S\&B)$, $R^2(OLS)$, or $R^2(MVP)$) the population value would be approximately .05 because of the unequal Level 1 and Level 2 variances. As shown in Table 5, all three total explained variance measures performed well. $R_2^2(approx.)$ had acceptable levels of bias but was not very efficient, producing an average standard deviation of .136 across conditions.

Finally, we assessed changes in R^2 when a cross-level interaction was added to an equation containing two Level 1 predictors and one Level 2 predictor. Thus, we compared the full random-slope model with a random-slope model without the cross-level interaction. The population values for changes in R^2 were identical to those when a Level 1 predictor was added. As shown in Table 5, all of the measures demonstrated acceptable levels of bias and efficiency. All measures except $R^2(OLS)$ returned negative estimates, but the occurrence was rare (<1% of the time).

Supplemental Analyses. In the previous simulations, we found only small differences between the explained variance measures. One reason for this may have been the clean delineation between Level 1 and Level 2 effects. In practice, this may not be the case. For example, it is possible for a variable to have different Level 1 and Level 2 effects (Snijders & Bosker, 2012). When this is the case, OLS and MLM may produce bias estimates if that variable is uncentered or grand-mean centered. In contrast, group-mean centering the variable and adding the group mean as a Level 2 predictor will produce fairly accurate estimates. However, it is not clear how each of the explained variance measures will perform when this occurs.

We investigated this by conducting some smaller simulations to assess if unequal Level 1 and Level 2 effects produced different results. Given the results of the aforementioned analyses, we generated data with 70 groups, an average of 8 cases per group, and a .30 ICC for the outcome. In one simulation, we specified Level 1 and Level 2 coefficients to be opposite in sign using the same levels of effect size as described previously. For example, in one condition, at Level 1, the variable had a positive effect that explained 10% of the Level 1 variance, and at Level 2, the variable had a negative effect that explained 50% of the Level 2 variance. We fully crossed the signs and effect size levels to produce 18 conditions. In the other simulation, we specified different positive effects at Level 1 and Level 2. For example, one condition had a positive Level 1 effect that explained 10% of the Level 1 variance and a positive Level 2 effect that explained 50% of the Level 2 variance. For both simulations, we generated 1,000 data sets

Table 6. Unequal Effect Sizes at Level 1 and Level 2.

	$R_1^2(approx.)$	$R_2^2(approx.)$	$R^2(S\&B)$	$R^2(OLS)$	$R^2(MVP)$
Opposite sign different effect					
Bias	−0.001 (−0.003)	−0.050 (−0.736)	−0.013 (−0.216)	0.001 (−0.150)	−0.005 (−0.120)
SD	0.031 (0.031)	0.125 (0.160)	0.041 (0.041)	0.042 (0.029)	0.040 (0.024)
Same sign different effect					
Bias	−0.002 (−0.003)	−0.054 (−0.105)	−0.014 (−0.030)	0.001 (−0.017)	−0.007 (−0.027)
SD	0.031 (0.031)	0.128 (0.138)	0.042 (0.042)	0.043 (0.042)	0.041 (0.034)

Note: Estimates averaged across 18,000 samples for opposite sign different effect results and 6,000 samples for same sign different effect results. Values outside parenthesis are group-mean centered; values in parenthesis are grand-mean centered.

per condition. For each sample, we analyzed the data using grand-mean centering as well as group-mean centering with the group means added in as a Level 2 predictor.

Table 6 presents the results. As expected, there were clear differences between the centering approaches with the group-mean centering approach producing less biased estimates. The bias for grand-mean centering was greater as the discrepancy between the effects sizes grew. However, we again found little differences across explained variance measures suggesting that centering choices and unequal effect sizes did not have different effects on the explained variance measures.

Discussion

A major challenge for researchers conducting multilevel analyses is to determine how to report explained variance. As such, the amount of explained variance is often not reported for multilevel analyses in organizational studies. Explained variance measures provide a useful summary of the magnitude of effects and may be particularly useful in multilevel studies where unstandardized coefficients are reported often. In the present study, we compared several existing multilevel explained variance measures and investigated several issues likely to occur when using these measures. A summary of the comparisons is presented in Table 7. Our results allow us to draw several conclusions about the measures and provide specific recommendations for their use.

Overall, most of the measures ($R_1^2(approx.)$, $R^2(S\&B)$, $R^2(OLS)$, and $R^2(MVP)$) exhibited acceptable levels of bias, consistency, and efficiency across all conditions and models. In addition, the difference in mean bias values for these measures was small. $R_1^2(approx.)$ and $R^2(S\&B)$ were the least biased in the random-intercept models and $R^2(S\&B)$ and $R^2(MVP)$ were the least biased in the random-slope models. In terms of efficiency, the $R_1^2(approx.)$ and $R^2(MVP)$ had the lowest standard deviation values in the random-intercept model. $R^2(MVP)$ and $R^2(OLS)$ had the lowest standard deviations in the random-slope model. In general, $R_2^2(approx.)$ was not an efficient estimator.

The conditions that we manipulated did not appear to have large effects on the majority of the explained variance measures. We manipulated the number of groups, group size, the amount of explained variance at each level, the amount of between-group variance in the outcome, and centering choices. None of these appeared to have large effects on the bias or standard deviations of the measures for either the random-intercept or random-slope models. Finally, most measures resulted in some negatives estimates even when the population effect size was nonzero. This occurred only for the measures that relied on comparing variance components across models. These negative values can likely be attributed to estimation error when sample sizes are smaller.

The presence of random slopes in multilevel models can create problems for estimating explained variance because the partitioning of variance is considerably more tedious. As such, some (Nakagawa & Schielzeth, 2013; Snijders & Bosker, 2012) have suggested basing explained variance on results from a random-intercept model because random slopes do not generally affect the prediction

Table 7. Comparison of Multilevel Explained Variance Measures.

Measure	Interpretation	Advantages	Disadvantages	General Comments
$R_1^2(approx.)$	Explained variance at Level 1	- Easy to calculate - Intuitive	- Possible negative values - Requires multiple models	- Can be used when Level 1 explained variance is the focus - Must be calculated using random intercept models
$R_2^2(approx.)$	Explained variance at Level 2	- Easy to calculate - Intuitive	- Possible negative values - Requires multiple models	- Poor bias and efficiency results in simulations - Not recommended
$R^2(S\&B)$	Total explained variance	- Easy to calculate - Intuitive	- Possible negative values - Requires multiple models	- Can be used when total explained variance is the focus - Can be calculated with random intercept model or random slope models
$R^2(OLS)$	Total explained variance	- Easy to calculate	- Based on inappropriate statistical model - Need to estimate multilevel model to obtain accurate standard errors	- Can be used when total explained variance is the focus - Must be accompanied by a multilevel analysis to obtain correct standard errors
$R^2(MVP)$	Total explained variance	- Requires single model - Will not result in negative estimates	- Tedious to calculate - Unwieldy formula	- Can be used when total explained variance is the focus - Can be calculated with random-intercept model or random-slope models

of the outcome. We compared this approach with incorporating the random-slope variance into the measures when applicable. Our results suggest little differences between the two approaches when there was little slope variance but that $R_1^2(approx.)$ and $R_2^2(approx.)$ did not perform well when there was more slope variance.

The results of the changes in explained variance analyses were generally consistent with those of the overall analyses. One exception was that negative values were found for $R^2(MVP)$. This primarily occurred because $R^2(MVP)$ was overestimated in the model without the cross-level interaction and is consistent with the notion that negative explained variance estimates from multilevel models are a result of model misspecification.

Recommendations

The results of the present study allow us to make some specific recommendations for reporting explained variance. Table 6 provides a summary of our recommendations. In general, $R_1^2(approx.)$, $R^2(S\&B)$, $R^2(OLS)$, and $R^2(MVP)$ appear to be viable options for reporting explained variance in multilevel models. In addition, there were very small differences in terms of bias and efficiency. As such, the choice of measure can be a matter of preference and nature of the research question. Calculating $R_1^2(approx.)$ and $R^2(S\&B)$ involve an intuitive comparison between the variance components of a null and full model. However, when there is missing data in the Level 1 predictors, it is possible that sample sizes for the models will differ. In contrast, $R^2(MVP)$ provides an

internally consistent partitioning of variance from a single multilevel model. However, $R^2(MVP)$ also requires additional calculations to compute explained variance. To aid in calculating all of the indices, we created an R function that will compute each measure.[2]

Researchers wishing to examine level-specific explained variance can use $R_1^2(approx.)$ to report Level 1 explained variance. However, there does not appear to be a viable option for Level 2 explained variance. $R_2^2(approx.)$ was not an efficient estimator, and we do not recommend its use. As an alternative, we suggest using the change in total variance associated with adding Level 2 predictors, which would allow researchers to estimate the amount of total variance that these predictors explain.

Random slopes can complicate the partitioning of variance in multilevel models. Our results support Snijder and Bosker's (1994) suggestion of using random-intercept models to calculate explained variance because random slopes do not generally predict the outcome. In fact, our results suggest that this approach is necessary for estimating $R_1^2(approx.)$ when there are random slopes. However, it is important to note that random-slope models still need to be estimated to obtain accurate standard errors.

We suggest that researchers use changes in explained variance for expressing the magnitude of cross-level interaction effects. $R_1^2(approx.)$, $R^2(S\&B)$, $R^2(OLS)$, and $R^2(MVP)$ performed well in estimating changes in explained variance, and $R^2(OLS)$ was the only measure that did not produce negative estimates. In addition, when calculating changes in explained variance based on variance component methods, researchers should compare R^2s from full and reduced models rather than compare variance components from those models.

Finally, we caution researchers to consider whether it is possible for variables to have different within- and between-group effects when calculating explained variance for multilevel models. When this is the case, OLS and MLM approaches may produce biased estimates unless group-mean centering is used and the group means are added as a Level 2 predictor. We recommend that researchers always compare results between grand-mean centered and group-mean centered analyses. In addition, although $R^2(OLS)$ can be used to assess explained variance in multilevel models, the OLS-based standard errors will be biased. Thus, researchers need to use MLM to calculate accurate significance levels and confidence intervals.

Limitations and Future Research

One limitation of the present study is that we examined a limited amount of multilevel models. We assessed the performance of explained variance measures across multilevel models that were selected because they represent some of the common scenarios in multilevel research and because they contain most of types of effects that will interest organizational researchers. We believe they generalize to models with more predictors and/or random effects because they are easily added to the models. However, it is possible that this is not the case.

We did not manipulate between-group variance in the Level 1 predictors in our simulations. As such, our results generalize to group-mean centered models. In practice, the decision about which centering method to use ultimately depends on the research question. Our focus on group-mean centering is consistent with Mathieu et al. (2012) and the recommended approach of testing cross-level interactions (Hofmann & Gavin, 1998). In addition, when there is considerable between-group variance in Level 1 predictors, we recommend using group-mean centering and using the group means as Level 2 predictors. This allows researchers to follow Snijders and Bosker's (2012) recommendation to consider the possibility of different within- and between-group effects.

Finally, we agree with Kelley and Preacher (2012) that effect sizes are estimates and should be reported with confidence intervals. Unfortunately, computing standard errors for multilevel R^2 measures in a single sample is not straightforward because they are partially a function of variance

components that have standard errors with undesirable properties (Snijders & Bosker, 2012). The most logical approach would be to employ bootstrapping; however, there is some debate about whether researchers should use case or residual bootstrapping (Goldstein, 2011). Thus, the question of how to estimate standard errors for multilevel explained variance measures still needs to be addressed.

Conclusion

In sum, our results suggest that researchers have several viable options for assessing the amount of explained variance in multilevel models. For the most part, the choice of measure can be a matter of preference although several issues need to be considered in choosing a measure. Specifically, researchers need to consider the presence of random slopes and the possibility of different between- and within-group effects for a variable as these will impact how each measure can be used.

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.

Notes

1. A complete list of references from the literature search is available from http://psych-scholar.wright.edu/lahuis/software/multilevel-effect-size-literature-review.
2. A function that will produce multiple *R* squared measures for multilevel models based on lmer models from lme4 package in R can be downloaded from http://psych-scholar.wright.edu/lahuis/software/multilevel-effect-size-function.

References

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., Werner, S., Abbott, J. L., Angert, C., Park, J. H., & Kohlhausen, D. (2010). Customer-centric science: Reporting research results with rigor, relevance, and practical impact in mind. *Organizational Research Methods*, *13*, 515-539. doi:10.1177/1094428109333339

Atwater, L., Wang, M., Smither, J. W., & Fleenor, J. W. (2009). Are cultural characteristics associated with the relationship between self and others' ratings of leadership? *Journal of Applied Psychology*, *94*(4), 876-886. doi:10.1037/a0014561

Bacharach, S. B., Bamberger, P. A., & Doveh, E. (2008). Firefighters, critical incidents, and drinking to cope: The adequacy of unit-level performance resources as a source of vulnerability and protection. *Journal of Applied Psychology*, *93*(1), 155-169. doi:10.1037/0021-9010.93.1.155

Bliese, P. D. (2000). Within-group agreement, non-independence, and reliability: Implications for data aggregation and analyses. In K. J. Klein & S. W. J. Kozlowski (Eds.), *Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions* (pp. 349-381). San Francisco, CA: Jossey-Bass.

Bliese, P. D., & Hanges, P. J. (2004). Being both too liberal and too conservative: The perils of treating grouped data as though they were independent. *Organizational Research Methods*, *7*, 400-417. doi:10.1177/1094428104268542

Bliese, P. D., & Ployhart, R. E. (2002). Growth modeling using random coefficient models: Model-building, testing, and illustrations. *Organizational Research Methods*, *5*, 101-116. doi:10.1177/109442802237116

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models for social and behavioural research: Applications and data analysis methods*. Newbury Park, CA: Sage Publications.

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*(2), 331-346. doi:10.1037/0021-9010.92.2.331

Dierdorff, E. C., Rubin, R. S., & Morgeson, F. P. (2009). The milieu of managerial work: An integrative framework linking work context to role requirements. *Journal of Applied Psychology*, *94*(4), 972-988. doi:10.1037/a0015456

Goldstein, H. (2011). Bootstrapping in multilevel models. In J. Hox & K. Roberts (Eds.), *Handbook of advanced multilevel analysis* (pp. 163-171). New York, NY: Taylor Francis Group.

Hofmann, D. A., & Gavin, M. B. (1998). Centering decisions in hierarchical linear models: Implications for research in organizations. *Journal of Management*, *5*, 623-641. doi: 10.1016/S0149-2063(99)80077-4

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*(1), 170-178. doi:10.1037/0021-9010.88.1.170

Hox, J. (2010). *Multilevel analysis: Techniques and applications* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Hoogland, J. J., & Boomsma, A. (1998). Robustness studies in covariance structural modeling: An overview and meta-analysis. *Sociological Methods & Research*, *26*, 329-367. doi:10.1177/0049124198026003003

Kelley, K., & Preacher, K. (2012). On effect size. *Psychological Methods*, *17*, 137-152. doi:10.1037/a0028086

Klein, K. J., & Kozlowski, S. W. J. (2000). *Multilevel theory, research, and methods in organizations*. San Francisco, CA: Jossey-Bass.

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

LaHuis, D. M., & Ferguson, M. W. (2009). The accuracy of statistical tests of variance components in multilevel random coefficient modeling. *Organizational Research Methods*, *12*(3), 418-435. doi:10.1177/1094428107308984

Liu, S. Q., Wang, M., Zhan, Y. J., & Shi, J. Q. (2009). Daily work stress and alcohol use: Testing the cross-level moderation effects of neuroticism and job involvement. *Personnel Psychology*, *62*(3), 575-597.

Long, J. D. (2012). *Longitudinal data analysis for the behavioral sciences using R*. Thousand Oaks, CA: Sage Publications.

Maas, C. J. M., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, *1*, 85-91.

Mathieu, J. E., Aguinis, H., Culpepper, S. A., & Chen, G. (2012). Power analyses and accurate inferences about cross-level interaction tests in random coefficients modeling. *Journal of Applied Psychology*, *97*, 951-966. doi:10.137/a0029358

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

Neal, A., & Griffin, M. A. (2006). A study of the lagged relationships among safety climate, safety motivation, safety behavior, and accidents at the individual and group levels. *Journal of Applied Psychology*, *91*(4), 946-953. doi:10.1037/0021-9010.91.4.946

Ng, K., Koh, C., Ang, S., Kennedy, J. C., & Chan, K. (2011). Rating leniency and halo in multisource feedback ratings: Testing cultural assumptions of power distance and individualism-collectivism. *Journal of Applied Psychology*, *96*(5), 1033-1044. doi:10.1037/a0023368

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

Snijders, T. A. B., & Bosker, R. J. (1994). Modeled variance in two-level models. *Sociological Methods & Research*, *22*, 342-363. doi:10.1177/0049124194022003004

Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling* (2nd ed.). Thousand Oaks, CA: Sage Publications.

Thompson, B. (2002). "Statistical," "practical," and "clinical": How many kinds of significance do counselors need to consider? *Journal of Counseling and Development*, *80*, 64-71. doi: 10.1002/j.1556-6678.2002.tb00167.x

Vandenberghe, C., Bentein, K., Michon, R., Chebat, J., Tremblay, M., & Fils, J. (2007). An examination of the role of perceived support and employee commitment in employee-customer encounters. *Journal of Applied Psychology*, *92*(4), 1177-1187. doi:10.1037/0021-9010.92.4.1177

Wallace, J. C., Edwards, B. D., Arnold, T., Frazier, L., & Finch, D. M. (2009). Work stressors, role-based performance, and the moderating role of organizational support. *Journal of Applied Psychology*, *94*, 254-262. doi:10.1037/a0013090

Zickar, M. J., & Highhouse, S. (2001). Measuring prestige of journals in industrial-organizational psychology. *The Industrial-Organizational Psychologist*, *38*, 29-36.

Author Biographies

David M. LaHuis received his doctorate in industrial and organizational psychology from the University of Connecticut in 2002 and is currently an associate professor of psychology at Wright State University in Dayton, Ohio. His research interests include item response theory, and multilevel modeling.

Michael J. Hartman received his master's degree in industrial and organizational psychology from Angelo State University in 2011 and is currently working on his PhD at Wright State University in Dayton, Ohio. His research interests include test development and measurement.

Patrick C. Clark received his doctorate in industrial and organizational psychology from Wright State University in 2013 and is currently a research and development scientist at Infor PeopleAnswers in Dallas, Texas. His research interests include item response theory and multilevel modeling.

Shotaro Hakoyama received his master's degree in industrial and organizational psychology from Wright State University in 2014. He is currently pursuing Ph.D. at Wright State University. His research interests include measurement and R programming.