American Psychological Association

Psychological Methods

1082-989X/19/$12.00

2019, Vol. 24, No. 3, 309–338
http://dx.doi.org/10.1037/met0000184

Quantifying Explained Variance in Multilevel Models: An Integrative Framework for Defining R-Squared Measures

Jason D. Rights and Sonya K. Sterba
Vanderbilt University

Abstract

Researchers often mention the utility and need for R-squared measures of explained variance for multilevel models (MLMs). Although this topic has been addressed by methodologists, the MLM R-squared literature suffers from several shortcomings: (a) analytic relationships among existing measures have not been established so measures equivalent in the population have been redeveloped 2 or 3 times; (b) a completely full partitioning of variance has not been used to create measures, leading to gaps in the availability of measures to address key substantive questions; (c) a unifying approach to interpreting and choosing among measures has not been provided, leading to researchers' difficulty with implementation; and (d) software has inconsistently and infrequently incorporated available measures. We address these issues with the following contributions. We develop an integrative framework of R-squared measures for MLMs with random intercepts and/or slopes based on a completely full decomposition of variance. We analytically relate 10 existing measures from different disciplines as special cases of 5 measures from our framework. We show how our framework fills gaps by supplying additional total and level-specific measures that answer new substantive research questions. To facilitate interpretation, we provide a novel and integrative graphical representation of all the measures in the framework; we use it to demonstrate limitations of current reporting practices for MLM R-squareds, as well as benefits of considering multiple measures from the framework in juxtaposition. We supply and empirically illustrate an R function, *r2MLM*, that computes all measures in our framework to help researchers in considering effect size and conveying practical significance.

Translational Abstract

R-squared measures are useful indications of effect size that are ubiquitously reported for single-level regression models. For multilevel models (MLMs), wherein observations are nested within clusters (e.g., students nested within schools), researchers likewise often mention the utility and necessity of R-squared measures; however, they find it difficult to relate, interpret, and choose among alternative existing measures, especially in the context of random slopes. Though methodologists have addressed this topic, the MLM R-squared literature suffers from several shortcomings: (a) the relationships among existing measures have not been established, leading to certain measures being redeveloped multiple times; (b) previous sets of measures have not considered all of the different ways that variance can be explained in MLMs, leading to gaps in the availability of measures to address key substantive questions; (c) a unifying approach to interpreting and choosing among measures has not been provided; and (d) existing software rarely incorporates available measures. In this article, we develop an integrative framework of R-squared measures for MLMs with random intercepts and/or slopes that addresses each of these shortcomings. We show that 10 existing measures are special cases of those from our framework, and show how our framework fills gaps by also supplying novel total and level-specific measures that answer important research questions. To facilitate interpretation, we introduce a unified graphical representation of all of the measures in our framework. We also demonstrate limitations of current R-squared reporting practices, and explain how considering our full framework overcomes these. We supply and illustrate new software that computes all of our measures to help researchers in considering effect size in MLM applications.

Keywords: R-squared, effect size, explained variance, random coefficient modeling, linear mixed effects modeling, hierarchical linear modeling

Supplemental materials: http://dx.doi.org/10.1037/met0000184.supp

This article was published Online First July 12, 2018.
Jason D. Rights and Sonya K. Sterba, Department of Psychology and Human Development, Vanderbilt University.
We thank Kristopher J. Preacher for helpful comments.
Correspondence concerning this article should be addressed to Jason D. Rights, Quantitative Methods Program, Department of Psychology and Human Development, Vanderbilt University, Peabody #552, 230 Appleton Place, Nashville, TN 37203. E-mail: jason.d.rights@vanderbilt.edu

Multilevel models (MLMs; also known as linear mixed effects models and hierarchical linear models) are commonly used to analyze nested data structures. Such data are particularly prevalent in social science, educational, and medical research wherein, for instance, students are nested within schools, patients are nested within clinicians, and repeated measures are nested within individuals. MLMs accommodate this nesting by allowing regression coefficients (intercept and/or slopes) to vary by cluster (e.g., school, clinician, individual).

For researchers applying MLMs, there is widely perceived to be a utility and need for R-squared (R^2) measures (Bickel, 2007; Edwards, Muller, Wolfinger, Qaqish, & Schabenberger, 2008; Jaeger et al., 2017; Johnson, 2014; Kramer, 2005; LaHuis, Hartman, Hakoyama, & Clark, 2014; Nakagawa & Schielzeth, 2013; Orelien & Edwards, 2008; Recchia, 2010; Roberts, Monaco, Stovall, & Foster, 2011; Wang & Schaalje, 2009; Xu, 2003; Zheng, 2000). For instance, LaHuis et al. (2014) emphasize that "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" (p. 446). In general, R^2's indicate the proportion of variance explained by the model and, as such, are considered measures of effect size[1] that (a) describe the correspondence between a model's predictions and the observed data (such that higher values of an R^2 measure mean predicted outcomes are more similar to the actual outcomes); (b) have an intuitive metric with well-defined endpoints (0 and 1); and (c) can be compared across studies with similar designs (e.g., Gelman & Hill, 2007; Kvålseth, 1985; Rights & Sterba, 2017; Xu, 2003). As summarized by Roberts et al. (2011) "With the further encouragement from editors to begin reporting effect sizes in all research, it is becoming more necessary for researchers using MLM to be able to explain their results in a way that is common with other statistical methods" (p. 229). Although several R^2 measures for MLMs have been separately developed (e.g., Bryk & Raudenbush, 1992; Hox, 2002; Johnson, 2014; Kreft & de Leeuw, 1998; Nakagawa & Schielzeth, 2013; Snijders & Bosker, 1994, 1999; Vonesh & Chinchilli, 1997; Xu, 2003), the MLM R^2 literature currently suffers from several shortcomings.

Shortcomings of the MLM R^2 Literature

Issue 1: Unknown Analytic Relationships Among Measures

Although lists of existing MLM R^2 measures proposed by different authors have been compiled (e.g., Jaeger et al., 2017; LaHuis et al., 2014; Orelien & Edwards, 2008; Roberts et al., 2011; Wang & Schaalje, 2009), general analytic relationships and equivalencies among these measures in the population have not been established. As we will see later, this has led to multiple instances wherein measures unknowingly equivalent in the population have been redeveloped two or three times.

Issue 2: Gaps in the Availability of Measures

Several authors (Johnson, 2014; Nakagawa & Schielzeth, 2013; Snijders & Bosker, 1999, 2012) had been thought to provide measures that "consider the full partitioning of variance for multilevel models" (LaHuis et al., 2014, p. 437). However, their approaches did not in fact use a full partitioning.[2] Their use of a limited partitioning led to fewer options and less flexibility in defining measures, which in turn has led to gaps in the availability of measures to address key substantive research questions.

Issue 3: Lack of a Unifying Approach to Interpreting Measures

R^2 measures for MLM are inherently more complicated to interpret than for single-level regression because of the potential for unexplained variance at each level of the hierarchy. Applied researchers have not been provided with a unifying approach to the interpretation of MLM R^2 measures in the context of random intercepts and slopes. As such, they have been found to be discouraged by the prospect of reconciling multiple definitions of MLM R^2's (Jaeger et al., 2017; Kreft & de Leeuw, 1998; Nakagawa & Schielzeth, 2013; Recchia, 2010), and have resorted to reporting typically only one R^2 measure, if any at all (LaHuis et al., 2014). The single reported measure has tended to be discipline-specific, likely because applied researchers have been unfamiliar with interpretation of measures developed outside of their discipline. This is because methodologists in the social sciences tend to recommend and use measures developed in their field (e.g., Bickel, 2007; Hox, 2010; McCoach & Black, 2008; Raudenbush & Bryk, 2002; Snijders & Bosker, 2012) whereas biostatisticians tend to recommend and use measures developed in their field (e.g., Edwards et al., 2008; Jaeger et al., 2017; Orelien & Edwards, 2008; Vonesh & Chinchilli, 1997; Xu, 2003).

Issue 4: Inconsistent and Incomplete Software Implementation of Measures

Researchers note that existing MLM R^2 measures are infrequently and inconsistently available in software (Bickel, 2007; Demidenko, Sargent, & Onega, 2012; Edwards et al., 2008; Jaeger et al., 2017; Kramer, 2005). This may be a byproduct of Issue 1 and 3, wherein the lack of understanding of how existing measures

[1] We employ Kelley and Preacher's (2012) definition of a measure of effect size as a "quantitative reflection of the magnitude of some phenomenon" that can help "inform a judgement about practical significance" (pp. 139–140).

[2] Specifically, previous measures constructed by partitioning model implied variance from a researcher's fitted model (Johnson, 2014; Nakagawa & Schielzeth, 2013; Snijders & Bosker, 1999, 2012) have not considered (a) partitioning outcome variance into *each* of total, within-cluster, and between-cluster variance; (b) creation of model-implied expressions for level-specific measures (which have the advantage of never being negative when existing level-specific measures such as those from Raudenbush & Bryk, 2002 can be); (c) partitioning explained total variance due to predictors at level-1 versus level-2 via fixed effects (which is facilitated by cluster-mean-centering); (d) explaining variance by predictors via random slope variation (in isolation or combination with another source); and (e) explaining variance by cluster-specific outcome means via random intercept variation (in isolation or combination with another source). These concepts will be defined later in the article, as will the motivation for (d) and (e) (see Vonesh & Chinchilli, 1997). Furthermore, note that Nakagawa and Schielzeth's (2013) partitioning does not allow random slopes, and Johnson's (2014) extension does not separately partition random intercept and slope variance. In contrast, features (a–e) are possible using the full partitioning used in our framework, and their utility will be explained later in the article.

relate and lack of consensus on which measure to use for what purpose both serve as obstacles for their incorporation into standard MLM software (Jaeger et al., 2017; Kramer, 2005). Compounding the inconvenience of software limitations, researchers state that MLM R^2 measures accommodating random slopes are complex and tedious to compute by hand (LaHuis et al., 2014; McCoach & Black, 2008; Nakagawa & Schielzeth, 2013; Snijders & Bosker, 2012). According to Edwards et al. (2008), MLM R^2 measures "that do exist have not become widely known. Even worse in terms of practical effect, most popular software does not provide easy access. Hence, the development, dissemination, and provision of easily accessible software . . . seems of the highest priority for linear mixed models" (p. 6150).

Contribution of Article

This article addresses the aforementioned four issues in the following manner. We begin by developing a general framework for defining and creating R^2 measures for MLMs with random intercepts and/or slopes. We comprehensively decompose model-implied variance from a researcher's fitted MLM into component parts and use these to construct a full set of 12 MLM R^2 measures. This approach and these measures are described in the section entitled Overview of an Integrative Framework of MLM R^2. Benefits of this general framework are: inclusion of existing measures, provision of substantively compelling new measures, convenience of having to fit only one model, and measures that cannot be negative.[3]

To address Issue 1 (Unknown Analytic Relationships Among Measures), we derive analytic equivalencies among 10 previously published MLM R^2 measures in the population, showing them to be special cases of five measures from our general framework. Proofs of these population equivalencies are provided in Appendixes B1–B10. In the section entitled Analytically Relating Pre-Existing MLM R^2's to the Current Framework these equivalencies are explained and a simulation illustration of their finite-sample correspondence is provided. The 10 previously published MLM R^2 measures that we analytically relate stem from across the social sciences and biostatistics fields, were developed independently using distinct mathematical approaches, and the set had not before been related analytically. Prior research had related existing measures by only listing them, discussing them conceptually, and/or comparing them empirically—while failing to acknowledge and/or demonstrate that many reflect the same population quantity (Jaeger et al., 2017; LaHuis et al., 2014; Orelien & Edwards, 2008; Roberts et al., 2011; Wang & Schaalje, 2009).

To address Issue 2 (Gaps in the Availability of Measures), our framework includes additional MLM R^2 measures obtained by reformulating individual components of existing measures in order to fulfill unmet substantive interpretational needs. Specifically, seven of the 12 measures in our framework are newly developed here in the sense that (a) they have not been used or proposed before in methodological or empirical work and (b) they answer novel and distinct substantive questions. Substantive motivation for these measures is provided in the section entitled Rationales for Newly-Developed Measures in the Framework and includes the need for measures representing a compromise between so-called "conditional" and "marginal" perspectives on explained variance (defined subsequently), the need for measures representing each source of explained variance individually, and the need for having "parallel" total versus level-specific measures (defined subsequently).

To address Issue 3 (Lack of a Unifying Approach to Interpreting Measures), we introduce a novel graphical representation of MLM R^2 measures depicting the component parts of each measure. This makes it straightforward to interpret a pattern of results across multiple MLM R^2 measures stemming from a single fitted model. This simple and integrated graphical approach to visualizing R^2 measures simultaneously increases accessibility and comprehension of not only measures from our framework but also other existing measures that correspond to those in our framework. Practical guidance for how to consult the framework's suite of R^2 measures in juxtaposition in order to obtain a comprehensive set of complementary information is provided in the section entitled Recommendations for Using MLM R^2 Framework in Practice.

To further address Issue 3, in the section entitled Limitations of the Common Practice of Reporting a Single MLM R^2 we describe and graphically illustrate how researchers can be potentially misled in four different ways about the interpretation of their results if they adhere to the current practice of reporting a single existing MLM R^2 measure in isolation. These illustrations also conversely demonstrate the utility of considering a suite of MLM R^2 measures in juxtaposition. The limitations of common practice that are illustrated in this section had not been previously noted in the methodological literature, yet they apply to empirical MLM applications that have reported R^2 to date. Hence, a key contribution of our article is to identify these limitations and show how to address them in concrete ways that are immediately relevant to applied practice.

Finally, to address Issue 4 (Inconsistent and Incomplete Software Implementation), we provide an R function, *r2MLM*, to aid researchers in computing all of the measures in our framework. This software computes all the measures from the output of a fitted MLM and automatically produces the graphical representation of the results. This software is described in the Software Implementation section. Subsequently, in the Empirical Examples section it is used to demonstrate our approach with three examples, each based on a prior analysis from a popular MLM textbook.

Scope of Article

Before continuing, it is important to clarify the scope of the present article in several respects. First, we will focus on two-level multilevel linear models with normal outcomes and homoscedastic residual variances at both level-1 and level-2, as this specification is most commonly employed in practice (Raudenbush & Bryk, 2002; Snijders & Bosker, 2012). Second, in order to avoid specifying level-1 predictors' effects that are conflated, "uninterpretable blend[s]" (Cronbach, 1976) of level-specific (i.e., within- and between-cluster) effects and to facilitate partitioning variance into level-specific components, here we will assume that level-1 predictors are cluster-mean-centered, as has been widely recom-

[3] Researchers have previously been concerned about the potential for some MLM R^2 measures to be negative (Hox, 2010; Jaeger et al., 2017; Kreft & de Leeuw, 1998; LaHuis et al., 2014; McCoach & Black, 2008; Nakagawa & Schielzeth, 2013; Recchia, 2010; Roberts et al., 2011; Wang et al., 2011).

mended (e.g., Hedeker & Gibbons, 2006; Kreft & de Leeuw, 1998; Preacher, Zyphur, & Zhang, 2010; Raudenbush & Bryk, 2002; Snijders & Bosker, 2012). Thus, if researchers are interested in a purely within-effect of a level-1 variable they could include a slope of the cluster-mean-centered level-1 variable. If they are also interested in a between-effect of that level-1 variable they should simultaneously include a slope of the cluster means of that level-1 variable. If a researcher wishes to fit a non-cluster-mean-centered MLM for a particular substantive reason (for instance, as described by Enders and Tofighi [2007], to examine level-2 effects while simply controlling for level-1 covariates), in the Discussion section we describe and provide modified formulae to compute a subset of our measures in the absence of cluster-mean-centering.

Third, we focus on R^2 measures for a single hypothesized model at a time (paralleling the focus of most methodological literature on MLM R^2 to date; Gelman & Pardoe, 2006), rather than on the use of R^2 in the context of model comparison. Fourth, the measures in our framework all afford the flexibility of accommodating random slopes, as random slopes are often used in practice. Consequently, we chose 10 previously published MLM R^2 measures to analytically relate to our framework that are themselves similarly general in accommodating random slopes. We did not focus on other measures that have particular kinds of restrictions (e.g., measures requiring fitting models with fixed slopes [e.g., Snijders & Bosker, 2012; p. 112, Equation 7.3]) as described further in the Discussion section.

As a final caveat, we emphasize that R^2 measures in our framework supplement rather than replace existing indicators of MLM model quality. Just as in single-level regression modeling, in MLM a high R^2 does not indicate that a model accurately reflects the data-generating process in the population; conversely, a low R^2 does not preclude a model from being informative for theory testing (see, e.g., Cohen, Cohen, West, & Aiken, 2003; King, 1986). MLM R^2 measures in our framework serve as intuitive measures of effect size, the use of which are increasingly recommended as focus shifts from overreliance on statistical significance (e.g., APA, 2009; Harlow, Mulaik, & Steiger, 1997; Kelley & Preacher, 2012; Panter & Sterba, 2011; Wilkinson & APA Task Force on Statistical Inference, 1999).

Multilevel Linear Model

To begin, we review the two-level MLM as background for the methodological developments in subsequent sections. Here, we are modeling some continuous outcome y_{ij} for observation i (level-1 unit) nested within cluster j (level-2 unit) with $i = 1 \ldots N_j$ and $j = 1 \ldots J$. We first present the level-1 regression model:

$$y_{ij} = \beta_{0j} + \sum_{p=1}^{P} \beta_{pj} v_{pij} + e_{ij}$$
$$e_{ij} \sim N(0, \sigma^2) \tag{1}$$

The intercept, β_{0j}, and the slopes, β_{pj}'s, of each of P level-1 predictors (v_{pij}'s) can both be cluster-specific, and the level-1 residual, e_{ij}, is normally distributed with variance σ^2. The level-2 regression equations that define the cluster-specific intercept and slopes are given as:

$$\beta_{0j} = \gamma_{00} + \sum_{q=1}^{Q} \gamma_{0q} z_{qj} + u_{0j}$$
$$\beta_{pj} = \gamma_{p0} + \sum_{q=1}^{Q} \gamma_{pq} z_{qj} + u_{pj}$$
$$\mathbf{u}_j \sim MVN(\mathbf{0}, \mathbf{T}) \tag{2}$$

Each cluster-specific intercept, β_{0j}, can be composed of a fixed component, γ_{00}, plus the sum of all Q level-2 predictors of the β_{0j} (i.e., z_{qj}'s) multiplied by their slopes (γ_{0q}'s), and the cluster-specific intercept deviation or residual, u_{0j}. The pth cluster-specific slope, β_{pj}, is similarly composed of a fixed component, γ_{p0}, plus the sum of all Q level-2 predictors of β_{pj} multiplied by their slopes (γ_{pq}'s), and the cluster-specific slope deviation or residual, u_{pj}. Note that any γ_{pq} denotes a cross-level interaction, that is, an interaction between z_{qj} and v_{pij}. Note also that any level-2 predictor not being used to model β_{0j} or β_{pj} would be given a coefficient of 0. The level-2 residuals in $\mathbf{u}_j$ (a $(P + 1) \times 1$ vector containing u_{0j} and P u_{pj}'s) are multivariate normally distributed with covariance matrix $\mathbf{T}$. To obtain a fixed intercept or fixed slope, the corresponding level-2 deviation in $\mathbf{u}_j$ would be set to 0 as would elements in $\mathbf{T}$ corresponding to its variance and covariance(s).

Representing this model in reduced form with the fixed components (the γ's) separated from the random components (the u's) makes it clear that an MLM can be conceptualized simply as a sum of fixed and random components (hence the common term "mixed effects model") like so:

$$y_{ij} = \left(\gamma_{00} + \sum_{q=1}^{Q} \gamma_{0q} z_{qj} + \sum_{p=1}^{P} \gamma_{p0} v_{pij} + \sum_{p=1}^{P} v_{pij} \sum_{q=1}^{Q} \gamma_{pq} z_{qj}\right) + \left(u_{0j} + \sum_{p=1}^{P} v_{pij} u_{pj}\right) + e_{ij} \tag{3}$$

This can be further simplified into a vector-based form, as in Equation (4):

$$y_{ij} = \mathbf{x}_{ij}'\boldsymbol{\gamma} + \mathbf{w}_{ij}'\mathbf{u}_j + e_{ij} \tag{4}$$

The $(1 + Q + P + QP) \times 1$ vector[4] $\boldsymbol{\gamma}$ contains all fixed components (i.e., all γ's in Equation [3]), with vector $\mathbf{x}_{ij}$ now consisting of 1 (for the intercept) and all predictors (i.e., level-1, level-2, and cross-level interactions). The $(1 + P) \times 1$ vector $\mathbf{w}_{ij}$ consists of 1 (for the intercept) and all level-1 predictors (level-1 predictors without random slopes have 0 elements in $\mathbf{u}_j$ so corresponding terms in $\mathbf{w}_{ij}'\mathbf{u}_j$ are 0).

For reasons mentioned above, in the current article we focus on the specification wherein all level-1 predictors are cluster-mean-centered, following common recommendations (e.g., Hedeker & Gibbons, 2006; Kreft & de Leeuw, 1998; Preacher et al., 2010; Raudenbush & Bryk, 2002; Snijders & Bosker, 2012). We can thus expand the $\mathbf{x}_{ij}'\boldsymbol{\gamma}$ term from Equation (4) to yield:

$$y_{ij} = \mathbf{x}_{ij}^{w\prime}\boldsymbol{\gamma}^{w} + \mathbf{x}_{j}^{b\prime}\boldsymbol{\gamma}^{b} + \mathbf{w}_{ij}'\mathbf{u}_j + e_{ij} \tag{5}$$

wherein $\mathbf{x}_{ij}^{w}$ denotes a vector of all cluster-mean-centered level-1 predictors and $\mathbf{x}_{j}^{b}$ denotes a vector of 1 (for the intercept) and all

[4] Dimensions of these vectors can be reduced when excluding coefficients that are set to 0—for instance for cross-level interactions that are not included in the model.

level-2 predictors (which could include cluster means of level-1 predictors). The vector $\boldsymbol{\gamma}^w$ denotes all level-1 fixed effects and $\boldsymbol{\gamma}^b$ denotes all level-2 fixed effects. This expansion in Equation (5) will facilitate description and presentation of the measures to follow. Note that cross-level interactions are included in $\mathbf{x}_{ij}^{w\prime}\boldsymbol{\gamma}^w$ when cluster-mean-centering; this is because such cross-level interaction terms (e.g., $v_{pij}z_{qj}$) have no variability across clusters, given that each cluster-specific mean of $v_{pij}z_{qj}$ will be equal to 0.[5]

Overview of an Integrative Framework of MLM R^2's

Generically, an R^2 can be defined in the population as

$$R^2 = 1 - \frac{\text{unexplained variance}}{\text{outcome variance}} \tag{6}$$

or, equivalently,

$$R^2 = \frac{\text{explained variance}}{\text{outcome variance}} \tag{7}$$

Estimating this proportion in a sample, then, involves estimating the outcome variance as well as either the unexplained or the explained variance.

One popular existing approach for estimating some MLM R^2 measures involves fitting two models: a *null* model (to estimate the outcome variance) and a *full* model (the researcher's model of substantive interest). This allows estimation of the population quantity:

$$R^2 = 1 - \frac{\text{unexplained variance from full model}}{\text{unexplained variance from null model}} \tag{8}$$

(Bryk & Raudenbush, 1992; Hox, 2002, 2010; Kreft & de Leeuw, 1998; Raudenbush & Bryk, 2002). Their null model consists of a fixed-intercept-only model or random-intercept-only model, depending on the measure. This two-model-fitting approach, however, has been criticized for the potential to yield negative estimates for certain MLM R^2 when implemented in a sample—which can occur, for instance, simply due to chance fluctuation in sample estimates from the two models (Hox, 2010; Jaeger et al., 2017; Kreft & de Leeuw, 1998; LaHuis et al., 2014; McCoach & Black, 2008; Nakagawa & Schielzeth, 2013; Recchia, 2010; Roberts et al., 2011; Wang, Xie, & Fisher, 2011). Hox (2010) has called this "unfortunate, to say the least" (p. 72).

Our approach, however, conveniently requires fitting only one model (the model of substantive interest from Equation [5]) and uses the reexpression in Equation (9) to estimate the population quantity:

$$R^2 = \frac{\text{explained variance from full model}}{\text{outcome variance from full model}} \tag{9}$$

Our approach uses model-implied variances from the researcher's single fitted model for both the denominator and numerator. Consequently, our measures will not be negative (provided a proper solution is obtained). Though others have previously also used model-implied variances from a researcher's single fitted model in the denominator and numerator of MLM total R^2's (Johnson, 2014; Nakagawa & Schielzeth, 2013; Snijders & Bosker, 1999, 2012) their partitioning of the outcome variance was more limited than ours (as described in Footnote 2 and in the next subsection), which led to them having fewer possibilities and less flexibility in defining measures (as shown later in Table 3). In the next subsection we extend their approach to allow a fuller partitioning of the outcome variance, which in turn, facilitates the construction of a complete suite of substantively interpretable R^2 measures.

Full Partitioning of Variance

The model-implied total outcome variance can be represented by Equation (10) (see Appendix A Section A1 for a detailed derivation of this equation):

$$\begin{aligned}&\textit{model-implied total outcome variance}\\ &\quad= \text{var}\left(\mathbf{x}_{ij}^{w\prime}\boldsymbol{\gamma}^w + \mathbf{x}_j^{b\prime}\boldsymbol{\gamma}^b + \mathbf{w}_{ij}^{\prime}\mathbf{u}_j + e_{ij}\right)\\ &\quad= \boldsymbol{\gamma}^{w\prime}\boldsymbol{\Phi}^w\boldsymbol{\gamma}^w + \boldsymbol{\gamma}^{b\prime}\boldsymbol{\Phi}^b\boldsymbol{\gamma}^b + tr(\mathbf{T}\boldsymbol{\Sigma}) + \tau_{00} + \sigma^2\end{aligned} \tag{10}$$

Here, $\boldsymbol{\Phi}^w$ and $\boldsymbol{\Phi}^b$ denote the covariance matrix of $\mathbf{x}_{ij}^w$ and the covariance matrix of $\mathbf{x}_{ij}^b$, respectively.[6] $\boldsymbol{\Sigma}$ denotes the covariance matrix of all elements of $\mathbf{w}_{ij}$. The penultimate term, τ_{00}, denotes the random intercept variance. This model-implied variance expression is useful in that it clarifies which specific components comprise the total outcome variance. Specifically, in Equation (10), each term represents variance attributable to one of five specific sources:

$$\boldsymbol{\gamma}^{w\prime}\boldsymbol{\Phi}^w\boldsymbol{\gamma}^w = \text{variance attributable to } \textit{level-1 predictors via fixed slopes} \text{ (shorthand: variance attributable to "}f_1\text{")} \tag{11}$$

$$\boldsymbol{\gamma}^{b\prime}\boldsymbol{\Phi}^b\boldsymbol{\gamma}^b = \text{variance attributable to } \textit{level-2 predictors via fixed slopes} \text{ (shorthand: variance attributable to "}f_2\text{")} \tag{12}$$

$$tr(\mathbf{T}\boldsymbol{\Sigma}) = \text{variance attributable to } \textit{level-1 predictors via random slope variation/covariation} \text{ (shorthand: variance attributable to "}v\text{")} \tag{13}$$

$$\tau_{00} = \text{variance attributable to } \textit{cluster - specific outcome means via random intercept variation} \text{ (shorthand: variance attributable to "}m\text{")} \tag{14}$$

$$\sigma^2 = \text{variance attributable to } \textit{level-1 residuals} \tag{15}$$

Three components in Equation (10) reflect purely within-cluster variance–namely, Equations (11), (13), and (15). Thus,

$$\begin{aligned}&\textit{model-implied within-cluster outcome variance}\\ &\quad= \boldsymbol{\gamma}^{w\prime}\boldsymbol{\Phi}^w\boldsymbol{\gamma}^w + tr(\mathbf{T}\boldsymbol{\Sigma}) + \sigma^2\end{aligned} \tag{16}$$

Hence, of the sources defined above, there can only be variance attributable to f_1 (Equation [11]) and/or v (Equation [13]) within a cluster beyond that attributable to level-1 residuals.

[5] With N_j denoting the number of level-1 units in cluster j, the mean for each of the J clusters is given as $(\sum_{i=1}^{N_j} v_{pij}z_{qj})/N_j = (z_{qj}/N_j)(\sum_{i=1}^{N_j} v_{pij}) = 0$.

[6] As noted by Snijders and Bosker (2012) in a similar context, it is atypical to think of exogenous predictors in terms of a population distribution; nonetheless, the predictors themselves contribute to the outcome variance, and thus the model-implied outcome variance must incorporate aspects of the distributions of the predictors.

Two components from Equation (10) reflect purely between-cluster variance—namely, Equations (12) and (14). Thus:

$$\text{model-implied between-cluster outcome variance}$$
$$= \boldsymbol{\gamma}^{b\prime}\, \boldsymbol{\Phi}^{b} \boldsymbol{\gamma}^{b} + \tau_{00} \quad (17)$$

Hence, of the sources defined above, there can only be variance attributable to f_2 (Equation [12]) and/or m (Equation [14]) between clusters.

Previous partitionings used in creating MLM R^2's did not decompose outcome variance into *each* of total, within-cluster, and between-cluster variance (Johnson, 2014; Nakagawa & Schielzeth, 2013; Snijders & Bosker, 2012) which precluded them from creating model-implied expressions for level-specific measures. Previous partitionings also did not distinguish variance attributable to level-1 versus level-2 predictors via fixed effects (i.e., "f_1" vs. "f_2"; Johnson, 2014; Nakagawa & Schielzeth, 2013; Snijders & Bosker, 2012)[7] nor distinguish between random intercept and random slope variation (i.e., "v" vs. "m"; Johnson, 2014), or simply excluded random slope variation (i.e., "v"; Nakagawa & Schielzeth, 2013). Note that the reason our variance partitioning expands upon previous partitionings used for MLM R^2's does not simply amount to the fact that we assumed cluster-mean-centering, unlike previous partitionings. As one example, total variance attributable to "v" and "m" can be partitioned whether or not cluster-mean-centering is assumed (though their interpretation would differ when not cluster-mean-centering; see Discussion section). Furthermore, the use of cluster-mean-centering itself does not automatically split total variance attributable to predictors via fixed effects (f) into within-cluster (f_1) and between-cluster (f_2) components, but rather the formulae we provide allows for this. Consequently, previous partitioning used in creating total R^2 measures by Johnson (2014), Nakagawa and Schielzeth (2013), and Snijders and Bosker (2012) would still be incomplete even if cluster-mean-centering were assumed; for instance, their proportion of total variance explained by predictors via fixed effects would then implicitly combine that explained via within-cluster (f_1) and between-cluster (f_2) effects.[8] Taken together, omissions from previous partitionings serve to restrict possibilities for constructing measures as will be seen later in Table 3. Our fuller partitioning will yield more informative results and a more complete set of options for defining MLM R^2 measures.

Using Equations (10–17), further defining an R^2 measure in the context of the researcher's fitted MLM thus involves two considerations: (a) what *outcome variance* is of interest (total, within-cluster, or between-cluster), which determines the denominator; and (b) which sources contribute to *explained variance*, which determines the numerator. *Total* R^2 measures incorporate *all* outcome variance (i.e., both within-cluster and between-cluster) in the denominator of Equation (9) and they quantify variance explained in an omnibus sense. They help one to ascertain how much outcome variance—as a whole—can be explained or understood by a given model. In contrast, *within-cluster* R^2 measures incorporate only within-cluster variance in the denominator of Equation (9) and they help researchers understand the degree to which within-cluster variance can be explained by a given model. For instance, with students nested within classrooms, a researcher may want to know the degree to which a given model can explain why students—within the same classroom—differ on the outcome of interest. Conversely, *between-cluster* R^2's can be useful for researchers interested specifically in the degree to which between-cluster variance can be explained by a given model; they incorporate only between-cluster variance in the denominator of Equation (9). For instance, with students nested within classrooms, a researcher may be interested in the degree to which a model can explain outcome differences between classrooms. Details follow on constructing total, within, and between measures and choosing the sources of explained variance for each.

Overview of Total R^2 Measures in the Framework

Having fully partitioned the model-implied outcome variance, we can transform each of the five components in Equations (11–15) into the proportion metric of R^2 by dividing by the total model-implied outcome variance. This transformation of the first four components—dividing each Equation (11) to (15) by Equation (10)—is shown in the first four rows of Table 1. Table 1 also provides a definition and symbol for each of these four proportions in the population ($R_t^{2(f_1)}$, $R_t^{2(f_2)}$, $R_t^{2(m)}$, $R_t^{2(v)}$). Each symbol's *superscript* denotes the *source* contributing to explained variance: "f_1" = level-1 predictors via fixed slopes, "f_2" = level-2 predictors via fixed slopes; "v" = predictors via slope variation/covariation; "m" = cluster-specific outcome means via intercept variation.[9] To conserve space, we hereafter rely on this shorthand "f_1", "f_2", "f", "v", and "m" to refer to these sources of explained variance (wherein "f" = the combination of f_1 and f_2). Each symbol's *subscript* denotes the outcome variance (here, "t" = total). For instance, the proportion of total outcome variance attributable to level-1 predictors via fixed slopes can be obtained by dividing $\boldsymbol{\gamma}^{w\prime}\boldsymbol{\Phi}^{w}\boldsymbol{\gamma}^{w}$ by Equation (10), which is denoted $R_t^{2(f_1)}$ in Table 1. Thus, the total variance can be decomposed into five separate components with corresponding proportions that sum to 1. It is pedagogically useful to visualize this decomposition graphically; hence, we introduce a graphical depiction, given in the Figure 1 bar chart. The leftmost column in the Figure 1 bar chart shows a

[7] Although Snijders and Bosker (2012, p. 117) said that a partitioning of variance attributable to f_1 and f_2 separately could be done, they did not provide specific formula to utilize this in their R^2 computation, and thus researchers computing their measure while cluster-mean-centering would be unable to separately consider the two components.

[8] Moreover, this splitting of f is not fully dependent on cluster-mean-centering, as described in the Discussion section. Of course, when choosing not to cluster-mean-center, researchers would need to ensure that the often more-restrictive assumptions of their fitted MLM are upheld or risk parameter bias.

[9] Given that random effect variance in MLM is generally termed residual variance at level-2, it may seem unintuitive that such variance could potentially count as "explained" variance in some measures. However, several existing MLM R^2 measures stemming from biostatistics—termed "conditional" measures—already do consider such random effect variation to be explained variance (e.g., Vonesh & Chinchilli, 1997), for reasons discussed in our later section entitled *There is substantive need for measures representing a compromise between so-called "conditional" and "marginal" perspectives*. Additionally, similar to the suggestion by Snijders and Bosker (1994), rather than just thinking of such measures in terms of *explained* variance, researchers may prefer the more neutral conceptualization of *modeled* variance.

Table 1
Definitions of Multilevel Model (MLM) R^2 Measures in Integrative Framework

Measure	Definition (Interpretation)
Total MLM R^2 measures	
$R_t^{2(f_1)} = \dfrac{\boldsymbol{\gamma}^{w\prime}\boldsymbol{\Phi}^w\boldsymbol{\gamma}^w}{\boldsymbol{\gamma}^{w\prime}\boldsymbol{\Phi}^w\boldsymbol{\gamma}^w + \boldsymbol{\gamma}^{b\prime}\boldsymbol{\Phi}^b\boldsymbol{\gamma}^b + tr(\mathbf{T\Sigma}) + \tau_{00} + \sigma^2}$	Proportion of total outcome variance explained by *level-1 predictors via fixed slopes*
$R_t^{2(f_2)} = \dfrac{\boldsymbol{\gamma}^{b\prime}\boldsymbol{\Phi}^b\boldsymbol{\gamma}^b}{\boldsymbol{\gamma}^{w\prime}\boldsymbol{\Phi}^w\boldsymbol{\gamma}^w + \boldsymbol{\gamma}^{b\prime}\boldsymbol{\Phi}^b\boldsymbol{\gamma}^b + tr(\mathbf{T\Sigma}) + \tau_{00} + \sigma^2}$	Proportion of total outcome variance explained by *level-2 predictors via fixed slopes*
$R_t^{2(f)} = \dfrac{\boldsymbol{\gamma}^{w\prime}\boldsymbol{\Phi}^w\boldsymbol{\gamma}^w + \boldsymbol{\gamma}^{b\prime}\boldsymbol{\Phi}^b\boldsymbol{\gamma}^b}{\boldsymbol{\gamma}^{w\prime}\boldsymbol{\Phi}^w\boldsymbol{\gamma}^w + \boldsymbol{\gamma}^{b\prime}\boldsymbol{\Phi}^b\boldsymbol{\gamma}^b + tr(\mathbf{T\Sigma}) + \tau_{00} + \sigma^2}$	Proportion of total outcome variance explained by *all predictors via fixed slopes*
$R_t^{2(v)} = \dfrac{tr(\mathbf{T\Sigma})}{\boldsymbol{\gamma}^{w\prime}\boldsymbol{\Phi}^w\boldsymbol{\gamma}^w + \boldsymbol{\gamma}^{b\prime}\boldsymbol{\Phi}^b\boldsymbol{\gamma}^b + tr(\mathbf{T\Sigma}) + \tau_{00} + \sigma^2}$	Proportion of total outcome variance explained by *level-1 predictors via random slope variation/covariation*
$R_t^{2(m)} = \dfrac{\tau_{00}}{\boldsymbol{\gamma}^{w\prime}\boldsymbol{\Phi}^w\boldsymbol{\gamma}^w + \boldsymbol{\gamma}^{b\prime}\boldsymbol{\Phi}^b\boldsymbol{\gamma}^b + tr(\mathbf{T\Sigma}) + \tau_{00} + \sigma^2}$	Proportion of total outcome variance explained by *cluster-specific outcome means via random intercept variation*
$R_t^{2(fv)} = \dfrac{\boldsymbol{\gamma}^{w\prime}\boldsymbol{\Phi}^w\boldsymbol{\gamma}^w + \boldsymbol{\gamma}^{b\prime}\boldsymbol{\Phi}^b\boldsymbol{\gamma}^b + tr(\mathbf{T\Sigma})}{\boldsymbol{\gamma}^{w\prime}\boldsymbol{\Phi}^w\boldsymbol{\gamma}^w + \boldsymbol{\gamma}^{b\prime}\boldsymbol{\Phi}^b\boldsymbol{\gamma}^b + tr(\mathbf{T\Sigma}) + \tau_{00} + \sigma^2}$	Proportion of total outcome variance explained by *predictors via fixed slopes and random slope variation/covariation*
$R_t^{2(fvm)} = \dfrac{\boldsymbol{\gamma}^{w\prime}\boldsymbol{\Phi}^w\boldsymbol{\gamma}^w + \boldsymbol{\gamma}^{b\prime}\boldsymbol{\Phi}^b\boldsymbol{\gamma}^b + tr(\mathbf{T\Sigma}) + \tau_{00}}{\boldsymbol{\gamma}^{w\prime}\boldsymbol{\Phi}^w\boldsymbol{\gamma}^w + \boldsymbol{\gamma}^{b\prime}\boldsymbol{\Phi}^b\boldsymbol{\gamma}^b + tr(\mathbf{T\Sigma}) + \tau_{00} + \sigma^2}$	Proportion of total outcome variance explained by *predictors via fixed slopes and random slope variation/covariation* and by *cluster-specific outcome means via random intercept variation*
Within-cluster MLM R^2 measures	
$R_w^{2(f_1)} = \dfrac{\boldsymbol{\gamma}^{w\prime}\boldsymbol{\Phi}^w\boldsymbol{\gamma}^w}{\boldsymbol{\gamma}^{w\prime}\boldsymbol{\Phi}^w\boldsymbol{\gamma}^w + tr(\mathbf{T\Sigma}) + \sigma^2}$	Proportion of within-cluster outcome variance explained by *level-1 predictors via fixed slopes*
$R_w^{2(v)} = \dfrac{tr(\mathbf{T\Sigma})}{\boldsymbol{\gamma}^{w\prime}\boldsymbol{\Phi}^w\boldsymbol{\gamma}^w + tr(\mathbf{T\Sigma}) + \sigma^2}$	Proportion of within-cluster outcome variance explained by *level-1 predictors via random slope variation/covariation*
$R_w^{2(f_1v)} = \dfrac{\boldsymbol{\gamma}^{w\prime}\boldsymbol{\Phi}^w\boldsymbol{\gamma}^w + tr(\mathbf{T\Sigma})}{\boldsymbol{\gamma}^{w\prime}\boldsymbol{\Phi}^w\boldsymbol{\gamma}^w + tr(\mathbf{T\Sigma}) + \sigma^2}$	Proportion of within-cluster outcome variance explained by *level-1 predictors via fixed slopes and random slope variation/covariation*
Between-cluster MLM R^2 measures	
$R_b^{2(f_2)} = \dfrac{\boldsymbol{\gamma}^{b\prime}\boldsymbol{\Phi}^b\boldsymbol{\gamma}^b}{\boldsymbol{\gamma}^{b\prime}\boldsymbol{\Phi}^b\boldsymbol{\gamma}^b + \tau_{00}}$	Proportion of between-cluster outcome variance explained by *level-2 predictors via fixed slopes*
$R_b^{2(m)} = \dfrac{\tau_{00}}{\boldsymbol{\gamma}^{b\prime}\boldsymbol{\Phi}^b\boldsymbol{\gamma}^b + \tau_{00}}$	Proportion of between-cluster outcome variance explained by *cluster-specific outcome means via random intercept variation*

Note. These measures were developed and defined under the assumption that the fitted MLM used cluster-mean-centering.

breakdown of the total outcome variance for a hypothetical example into the five distinct proportions. From this leftmost bar chart, we could say, for instance, that "20% of the total outcome variance is attributable to level-1 predictors via fixed slopes," as depicted by $R_t^{2(f_1)}$.

The proportions $R_t^{2(f_1)}$, $R_t^{2(f_2)}$, $R_t^{2(m)}$, $R_t^{2(v)}$ can be used individually *or in combination* to define total R^2 measures. The suite of total R^2 measures focused on in our framework includes these measures tapping individual sources of explained variance ($R_t^{2(f_1)}$, $R_t^{2(f_2)}$, $R_t^{2(m)}$, $R_t^{2(v)}$) as well as the following three measures that combine multiple sources of explained variance, as defined in Table 1: $R_t^{2(f)}$ (combining sources f_1 and f_2), $R_t^{2(fv)}$ (combining sources f and v), and $R_t^{2(fvm)}$ (combining sources f, v, and m). Although other combined-source measures could be created from within the framework, we focus on this subset because (a) it includes previously published measures (as shown in the upcoming Analytic Relations section); and (b) it fills compelling substantive needs (as shown in the upcoming Rationales section).

Overview of Within-Cluster R^2 Measures in the Framework

Restricting focus to within-cluster outcome variance in Equation (16), we can again transform each component into a proportion. Figure 1 (middle bar chart) graphically illustrates the decomposition of scaled within-cluster variance into three separate propor-

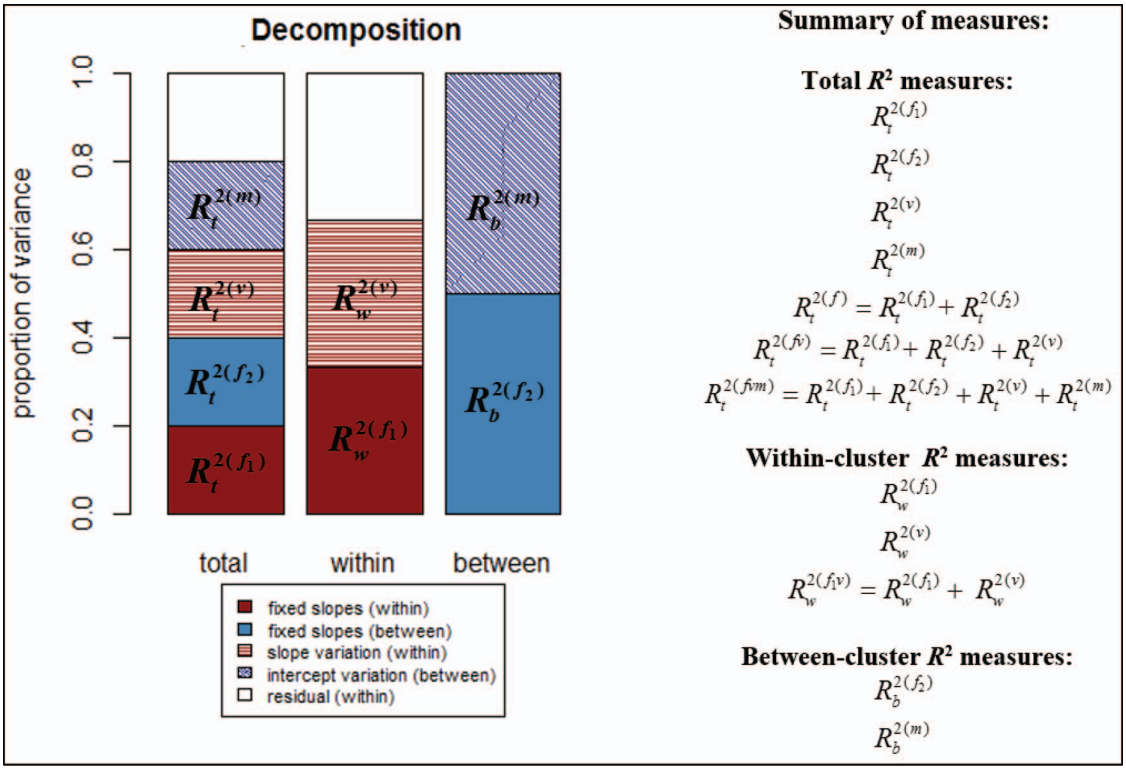

Figure 1. Decomposition of scaled outcome variance into proportions to construct R^2 measures in the framework: Bar chart graphic for a hypothetical example. In the bar chart, the symbol for each measure (from Table 1) is superimposed on its corresponding proportion of variance. The first column of the bar chart decomposes scaled total variance into proportions; the second column decomposes scaled within-cluster variance into proportions; and the third column decomposes scaled between-cluster variance into proportions. A given shade (e.g., horizontal stripe) refers to a source of explained variance, and when this same shade appears in multiple columns (e.g., total column *and* within column) it means that same source is counted as explained variance (in the numerator of the R^2) in measures with different outcome variances (in the denominator of the R^2). The white space in the first and second columns refers to the level-1 residual variance divided by either the total or within variance, respectively. To the right of the figure, measures in our framework are listed, some of which combine proportions from the figure. More detailed definitions of each measure were given in Table 1. See the online article for the color version of this figure.

tions for a hypothetical example. Thus, the three within-cluster components in the middle bar chart of Figure 1 display the proportion of variance relative to all within-cluster variance (e.g., 33% of within-cluster variance is attributable to level-1 predictors via fixed slopes). The transformation of the first two components into proportions—dividing Equations (11) and (13) by (16)—is shown in Table 1, alongside definitions and symbols for these proportions in the population ($R_w^{2(f_1)}$, $R_w^{2(v)}$). For instance, the proportion of within-cluster outcome variance attributable to level-1 predictors via fixed slopes, denoted $R_w^{2(f_1)}$ in Table 1, is obtained by dividing $\boldsymbol{\gamma}^{w\prime}\boldsymbol{\Phi}^{w}\boldsymbol{\gamma}^{w}$ by Equation (16).[10] The "w" *subscript* on these proportions indicate the outcome variance is within-cluster and the *superscripts* indicate that, of the potential sources of explained variance defined earlier, only f_1 and/or v can explain variance within a cluster beyond that attributable to level-1 residuals. Hence, in the remainder of this paper we use $R_w^{2(f_1)}$, $R_w^{2(v)}$ and their combination $R_w^{2(f_1v)}$ (also defined in Table 1) as the within-cluster R^2 measures in our framework.

Overview of Between-Cluster R^2 Measures in the Framework

Now restricting focus to between-cluster outcome variance in Equation (17), we again transform each component into a proportion by dividing Equations (12) and (14) by (17). The decomposition of scaled between-cluster variance into two proportions is illustrated in Figure 1 (rightmost bar chart) for a hypothetical example. Thus, the two between-cluster components in the rightmost bar chart of Figure 1 display the proportion of variance relative to all between-cluster variance (e.g., 50% of between-cluster variance is attributable to

[10] These level-specific measures are conceptually similar to a partial R^2 familiar from single-level multiple regression analyses. Whereas a partial R^2 in multiple regression defines the outcome variance (i.e., the R^2 denominator) as the variance that is not accounted for by a set of predictors, the *within-cluster* R^2 in our framework defines the outcome variance as the variance that is not accounted for by *between-cluster* sources, and vice versa.

level-2 predictors via fixed slopes). Table 1 provides definitions and symbols for each proportion measure in the population (denoted $R_b^{2(f_2)}$, $R_b^{2(m)}$). For instance, the proportion of between-cluster outcome variance attributable to level-2 predictors via fixed slopes, denoted $R_b^{2(f_2)}$ in Table 1, is obtained by dividing $\boldsymbol{\gamma}^{b\prime}\boldsymbol{\Phi}^b\boldsymbol{\gamma}^b$ by Equation (17). The "b" *subscript* on these proportions indicates that the outcome variance is between-cluster and the *superscripts* indicate that, of the potential sources of explained variance defined earlier, only f_2 and m can explain variance between clusters. In contrast to the total MLM R^2's and within-cluster MLM R^2's, wherein level-1 residuals were considered to *always* contribute to unexplained variance, in between-cluster MLM R^2's, there is no source that need always contribute to unexplained variance. Consequently, in contrast to the total MLM R^2's, there is no need to create a between-cluster MLM R^2 measure treating both sources f_2 and m together, as this would necessarily equal 1. Hence the suite of two between-cluster measures in our framework considers sources f_2 and m individually (i.e., $R_b^{2(f_2)}$, $R_b^{2(m)}$).

Summary

Table 2 provides a summary of the MLM R^2 measures in our framework. Table 2 has three columns denoting choices of outcome variance (total, within-cluster, between-cluster) and rows denoting choices of which sources contribute to explained variance. Each MLM R^2 measure is defined by what is in the subscript (choice of outcome variance: t, w, b) and in the superscript (sources contributing to explained variance: f_1, f_2, f, v, m, and combinations thereof).

Recommendations for Using MLM R^2 Framework in Practice

In reporting MLM R^2 measures, the sheer number of options may initially feel overwhelming, but these can be organized and streamlined. As a straightforward approach, researchers can report the measures that contain only a single source of explained variance at a time in the numerator, that is, the total measures $R_t^{2(f_1)}$, $R_t^{2(f_2)}$, $R_t^{2(v)}$, and $R_t^{2(m)}$ and their level-specific counterparts $R_w^{2(f_1)}$, $R_b^{2(f_2)}$, $R_w^{2(v)}$, and $R_b^{2(m)}$. All of these can simultaneously be visualized and reported in a bar chart such as that in Figure 1 (software provided for doing so is discussed later). Each of these can be reported as a quantitative effect size to supplement qualitative interpretation. For example, a statement such as "there is heterogeneity in the effects of the predictors" can be made more informative with statements such as "specifically, 15% of the total outcome variance is attributable to the predictors via slope variation," corresponding to an $R_t^{2(v)}$ estimate of .15.

If a summary measure that combines sources of explained variance is also desired, researchers can substantively justify what outcome variance is of interest, which combination of sources should contribute to explained variance, and then use the appropriate R^2 measure from Table 1. However, there is no single, one-size-fits-all measure that will address every research question. Hence, we do not recommend reporting just one combined-source measure in isolation (for reasons illustrated graphically in a later section, entitled Limitations of the Common Practice of Only Reporting a Single MLM R^2). Rather, we encourage researchers to interpret a given measure in juxtaposition to other measures within the context of the full decomposition (relatedly, see Rights & Sterba, 2017).

Analytically Relating Pre-Existing MLM R^2's to the Current Framework

Of the 12 MLM R^2 measures in the integrative framework, from Table 2, there are seven which have not previously been proposed ($R_t^{2(f_1)}$, $R_t^{2(f_2)}$, $R_t^{2(fv)}$, $R_t^{2(m)}$, $R_w^{2(f_1)}$, $R_w^{2(v)}$, $R_b^{2(m)}$) and five which correspond to the same population quantities as previous authors' measures ($R_t^{2(fvm)}$, $R_t^{2(f)}$, $R_t^{2(v)}$, $R_w^{2(f_1 v)}$, $R_b^{2(f_2)}$). Appendix B sections B1 through B10 provide analytic derivations showing how the latter five measures from the framework represent the same population quantities as published measures from other authors, despite their differences in computation (Aguinis & Culpepper, 2015; Bryk & Raudenbush, 1992; Hox, 2002, 2010; Johnson, 2014; Kreft & de Leeuw, 1998; Raudenbush & Bryk, 2002; Snijders & Bosker, 1999, 2012; Vonesh & Chinchilli, 1997; Xu, 2003). Appendix B shows these equivalencies with or without the assumption of cluster-mean-centering, for total measures, and with the assumption of cluster-mean-centering, for level-specific measures. Table 3 overviews the correspondence between R^2 MLM measures from our integrative framework and those developed by previous authors. Each *column* of Table 3 refers to one of the 12 measures in our framework. Each *row* of Table 3 refers to a different author who has previously developed a MLM R^2 measure. Cells of Table 3 indicate the page number and symbol for each previously published measure that corresponds to the same population quantity as one of our measures (with supporting proofs in each case provided in Appendix B1–B10).

It can be seen from Table 3 that certain measures were developed previously by, not one, but multiple sets of authors. As previously mentioned in Issue 1 (Unknown Analytic Relationships Among Measures) it has not before been appreciated that these multiple measures are estimating the same population quantity. Specifically, three previous sets of authors independently developed measures corresponding to $R_t^{2(fvm)}$ in the population: Vonesh and Chinchilli (1997) (proof of correspondence given in Appendix B Section B1), Xu (2003) (proof given in Appendix B Section B2), and Johnson (2014) (proof given in Appendix B Section B3). Also three previous sets of authors independently developed measures corresponding to $R_t^{2(f)}$ in the population: Snijders and Bosker (1999, 2012) (proof given in Appendix B Section B4), Vonesh and Chinchilli (1997) (proof given in Appendix B Section B5), and Johnson (2014) (proof given in Appendix B Section B6). Furthermore, two previous authors independently developed measures corresponding to $R_w^{2(f_1 v)}$ in the population: Raudenbush and Bryk (2002, see also 1992 edition) (proof given in Appendix B Section B7) and Vonesh and Chinchilli (1997) (proof given in Appendix B Section B8).[11] One previous set of authors developed a measure corresponding to the same population quantity as $R_b^{2(f_2)}$: Raudenbush and Bryk (2002, see also 1992 edition) (proof given in Appendix B Section B9). Notably, the two latter measures were also widely disseminated by Hox (2002, 2010) and Kreft and de Leeuw (1998). Lastly, one previous set of authors developed a

[11] Xu (2003) also includes a measure similar to $R_w^{2(f_1 v)}$, but formulae are provided under the restrictive assumption that intercepts and slopes are uncorrelated; hence, we exclude this measure from Table 3.

Table 2
MLM R^2 Measures in the Integrative Framework, Distinguished by Outcome Variance of Interest (Denominator) and Sources Contributing to Explained Variance (Numerator)

*Sources contributing to explained variance**	*Outcome variance of interest †*		
	total	**within-cluster**	**between-cluster**
L1 predictors via fixed slopes ("f_1")	$R_t^{2(f_1)}$	$R_w^{2(f_1)}$	N/A (no between variability is explained by f_1 so this measure would always be 0)
L2 predictors via fixed slopes ("f_2")	$R_t^{2(f_2)}$	N/A (no within variability is explained by f_2 so this measure would always be 0)	$R_b^{2(f_2)}$
all predictors via fixed slopes ("f")	$R_t^{2(f)}$	N/A (no within variability is explained by f_2 so this measure would always be equal to $R_w^{2(f_1)}$)	N/A (no between variability is explained by f_1 so this measure would always be equal to $R_b^{2(f_2)}$)
L1 predictors via random slope (co)variation ("v")	$R_t^{2(v)}$	$R_w^{2(v)}$	N/A (no between variability is explained by v so this measure would always be 0)
cluster-specific outcome means via random intercept variation ("m")	$R_t^{2(m)}$	N/A (no within variability is explained by m so this measure would always be 0)	$R_b^{2(m)}$
predictors via fixed slopes and random slope (co)variation ("fv")	$R_t^{2(fv)}$	$R_w^{2(fv)}$	N/A (no between variability is explained by v so this measure would always be equal to $R_b^{2(f_2)}$)
predictors via fixed slopes and random slope (co)variation and cluster-specific outcome means via random intercept variation ("fvm")	$R_t^{2(fvm)}$	N/A (no within variability is explained by m so this measure would always be equal to $R_w^{2(fv)}$)	N/A (v and m already account for all between variability so this measure would always be 1)

† Subscripts denote the outcome variance: t = total; w = within-cluster; b = between-cluster; Ll = level-1; L2 = level-2. * Superscripts denote source(s) contributing to explained variance for a given measure: f = predictors via fixed slopes; f_1 = level-1 predictors via fixed slopes; f_2 = level-2 predictors via fixed slopes; v = predictors via slope variation/covariation; m = cluster-specific outcome means via intercept variation. Shaded cells correspond to combinations that are not applicable, as described in the text of the cells.

measure corresponding to $R_t^{2(v)}$ in the population: Aguinis and Culpepper (2015) (proof given in Appendix B Section B10).[12]

Simulation Illustration

It is important to underscore that, though all rows within a given column of Table 3 correspond with the same population R^2 measure (as derived in Appendix B), that measure would be computed differently by each author in a given sample. More specifically, within a given column of Table 3, in service of estimating the same population R^2 measure, authors may use different combinations of estimates (e.g., some use estimates from one fitted [full] model, others use estimates from two [null and full] models; some use estimates of level-1 residual variance but not estimates of random effect variances and vice versa; some require outputting cluster-specific empirical Bayes predicted scores and others do not). In some cases, certain of our measures are not only equivalent to pre-existing measures in the population, but also in the sample (for Johnson, 2014 extension of Nakagawa & Schielzeth, 2013 and Snijders & Bosker, 2012). In other cases, denoted by "*" in Appendix B, equivalencies between our measures and pre-existing measures hold in the population but not necessarily in a given sample (for Aguinis & Culpepper, 2015; Raudenbush & Bryk, 2002; Vonesh & Chinchilli, 1997; and Xu, 2003). In all cases, different authors' sample estimates of the same population R^2 measure would show greater correspondence with each other, and with the population value, given a larger number of clusters and/or cluster size. Even at moderate N_j and J, however, across repeated samples, the average value of a given measure computed using each author's approach should be similar. We illustrate the latter point using a simulation, wherein we generated 500 samples having a known population value for each of the 12 MLM R^2 in our framework. The generating multilevel model had a random intercept, random slopes of three cluster-mean-centered level-1 predic-

[12] Although Aguinis and Culpepper (2015) did not interpret their measure as an R^2, it is such if considering v as a source of explained variance. They termed their measure ICC beta and viewed it as a complementary measure to the intraclass correlation coefficient (ICC). Whereas the conventional ICC captures the degree of mean outcome variation across clusters, their ICC beta (i.e., $R_t^{2(v)}$) captures "the degree of variability of a lower-level relationship across higher-order units" (Aguinis & Culpepper, 2015, p. 168).

Table 3
Population Relationships Among Previous Authors' MLM R^2 Measures and Those in Our Integrative Framework (Supporting Derivations in Appendix B)

		Total measures							Within-cluster measures			Between-cluster measures	
	Rights & Sterba:	$R_t^{2(f_1)}$	$R_t^{2(f_2)}$	$R_t^{2(v)}$	$R_t^{2(m)}$	$R_t^{2(f)}$	$R_t^{2(fv)}$	$R_t^{2(fvm)}$	$R_w^{2(f_1)}$	$R_w^{2(v)}$	$R_w^{2(f_1v)}$	$R_b^{2(f_2)}$	$R_b^{2(m)}$
Author(s)	Vonesh & Chinchilli (1997)					p. 422 Eqn. 8.3.7 if $\hat{y}_{ij} = \mathbf{x}'_{ij}\boldsymbol{\gamma}$		p. 422 Eqn 8.3.7 if $\hat{y}_{ij} = \mathbf{x}'_{ij}\boldsymbol{\gamma} + \mathbf{w}_j\mathbf{u}_j$			p. 422 Eqn 8.3.7 if $\hat{y}_{ij} = \mathbf{x}'_{ij}\boldsymbol{\gamma} + \mathbf{w}_j\mathbf{u}_j$ and replacing $\bar{y}$ with $\hat{y}_{ij}^{null}$ from Eqn. 8.3.9 *		
	Snijders & Bosker (2012) (see also 1999)					p. 117 divide first term in Eqn 7.9 by entirety of Eqn 7.9							
	Xu (2003)							termed Ω_0^2 p. 3530 Eqn 6					
	Aguinis & Culpepper (2015)			termed ρ_β p. 8 Eqn. 20									
	Johnson (2014) (extension of Nakagawa & Schielzeth [2013])					p. 945 Eqn 1 (replace $\sum_{l=1}^{u}\sigma_l^2$ with Eqn 10; drop σ_d^2)		p. 945 Eqn 2 (replace $\sum_{l=1}^{u}\sigma_l^2$ with Eqn 10; drop σ_d^2)					
	Raudenbush & Bryk (2002) (see also Bryk & Raudenbush 1992, Hox, 2002, 2010; Kreft & de Leeuw, 1998)										termed "proportion reduction in variance or 'variance explained' at L1" p. 79 Eqn 4.20*	termed "proportion reduction in variance or 'variance explained' at L2" p. 74 Eqn 4.12*	

Note. We show analytically in Appendix B that all measures in the same column are estimates of the same population quantity, with or without the assumption of cluster-mean-centering for total measures (Appendices B1–B6) and with the assumption of cluster-mean-centering for level-specific measures (Appendices B7–B10). Page numbers in the table correspond with authors' original sources. The table has blank cells because each previous author provided only 1, 2, or 3 out of the 12 measures in our framework. MLM = multilevel model.
* Computation involves fitting two models (full and null); in particular, Raudenbush-Bryk's and Vonesh-Chinchilli's analog to $R_w^{2(f_1v)}$ and the former's analog to $R_b^{2(f_2)}$ define the null model as random-intercept-only (see Appendix B).

tors, and fixed slopes of two level-2 predictors.[13] For each sample, $J = 200$ and $N_j = 50$. This sample size would be generally considered sufficient for multilevel modeling (e.g., Maas & Hox, 2005). The generating model was fit to all 500 samples. All 12 MLM R^2 measures were computed using each available previously published approach from Table 3, as well as our own approach. Computing sample estimates of measures in our framework involves simply replacing the parameters in each equation in Table 1 with their corresponding sample estimates (i.e., estimated fixed effects and estimated random effect [co]variances as well as the sample-estimated predictor covariance matrices). Table 4 presents the population value of each of the 12 measures (top row), along with the average estimates obtained using our approach (second row) and each previous author's approach (subsequent rows).

The results in Table 4 indicate that estimates from our measures and from other authors' measures correspond closely to each other and to the population values, on average. Two exceptions, however, were that Vonesh and Chinchilli's (1997) method of computing $R_w^{2(f_1v)}$ and Raudenbush-Bryk/Hox/Kreft-de-Leeuw's method of computing $R_b^{2(f_2)}$ did not perform as well. The former may largely be due to the fact that these Vonesh and Chinchilli's (1997) measures require outputting cluster-specific empirical Bayes estimates for intercept and slope random effects (i.e., $\hat{\mathbf{u}}_j$) to compute the predicted outcomes for each observation, which may introduce an additional source of error (Snijders & Bosker, 2012) that can lead to bias in their measures at lower sample sizes. It is less clear analytically why the Raudenbush and Bryk (2002) between-clusters measure performed less well, but it should be noted that it has been shown to perform poorly in another simulation (LaHuis et al., 2014).

[13] Predictors were multivariate normally distributed. Generating parameters were: $\boldsymbol{\Phi}^w = \begin{pmatrix} 2 & .3 & .75 \\ .3 & 1.5 & .2 \\ .75 & .2 & 1 \end{pmatrix}$, $\boldsymbol{\Phi}^b = \begin{pmatrix} 2 & .5 \\ .5 & 1.5 \end{pmatrix}$, $\boldsymbol{\gamma}^w = \begin{bmatrix} 1 \\ -2 \\ 3 \end{bmatrix}$, $\boldsymbol{\gamma}^b = \begin{bmatrix} 1 \\ -.5 \\ 2 \end{bmatrix}$, $\mathbf{T} = \begin{pmatrix} 10 & .25 & .5 & .6 \\ .25 & 1 & .36 & .4 \\ .5 & .36 & 1.5 & .6 \\ .6 & .4 & .6 & 2 \end{pmatrix}$, $\sigma^2 = 17$.

Table 4
Simulation Results: Finite Sample Correspondence Among Measures Listed in Table 3

	Measure											
	$R_t^{2(f_1)}$	$R_t^{2(f_2)}$	$R_t^{2(f)}$	$R_t^{2(v)}$	$R_t^{2(m)}$	$R_t^{2(fv)}$	$R_t^{2(fvm)}$	$R_w^{2(f_1v)}$	$R_w^{2(f_1)}$	$R_w^{2(v)}$	$R_b^{2(f_2)}$	$R_b^{2(m)}$
Population value	**.31**	**.10**	**.41**	**.13**	**.17**	**.53**	**.71**	**.60**	**.42**	**.17**	**.36**	**.65**
Author(s)												
Rights and Sterba framework	.31	.10	.40	.13	.17	.53	.70	.59	.42	.17	.36	.64
Snijders and Bosker (2012)			.40									
Raudenbush and Bryk (2002)								.60			.33*	
Xu (2003)							.70					
Vonesh and Chinchilli (1997)			.41				.72	.76*				
Aguinis and Culpepper (2015)				.13								
Johnson (2014) (extension of Nakagawa and Schielzeth, 2013)			.40				.70					

Note. Each table cell provides the average estimate across 500 samples. Note that the table has blank cells because each previous author provided only 1, 2, or 3 out of the 12 measures in our framework.
* Discrepancies of these Vonesh and Chinchilli (1997) and Raudenbush and Bryk (2002) estimates from population values are discussed in the article text. Bolded values are the population-generating values whereas plain-text values are across-sample average estimates of those values.

Our contribution in the current article was to show the analytic equivalencies of existing measures with several of our measures in the population (derived in Appendix B), as well as to show an example of their finite sample correspondence and performance (in this section). It is outside our scope to examine and contrast finite sampling properties of these estimators under a variety of generating conditions; this could be a direction for future research.

Rationales for Newly Developed Measures in the Framework

Deriving a suite of R^2's based on a more complete decomposition of variance than in prior literature allowed us to identify and fill gaps where additional measures can be used to fill sensible interpretational needs (addressing Issue 2). As was summarized in Table 3, seven of the measures in our framework were newly added here. Next we discuss substantive rationales motivating these measures ($R_t^{2(fv)}$, $R_t^{2(m)}$, $R_t^{2(f_1)}$, $R_t^{2(f_2)}$, $R_w^{2(f_1)}$, $R_w^{2(v)}$, $R_b^{2(m)}$).

There is Substantive Need for Measures Representing a Compromise Between So-Called "Conditional" and "Marginal" Perspectives: Rationale for $R_t^{2(fv)}$

Compared with single-level contexts, in MLM the choice of numerator for R^2 is complicated by the presence of multiple variance components. A researcher must consider: Should variance attributable to predictors and cluster means via random effects be treated as explained variance or unexplained variance? Two alternative perspectives have previously been offered in the literature for total R^2 measures, corresponding with the use of the terms "marginal" versus "conditional" measures (e.g., Edwards et al., 2008; Orelien & Edwards, 2008; Vonesh & Chinchilli, 1997; Wang & Schaalje, 2009; Xu, 2003). The first alternative is to count all variance attributable to predictors and cluster means via random effects (v and m) as unexplained, and count variance attributable only to predictors via fixed effects (f) as explained—that is, computing $R_t^{2(f)}$. This corresponds to what some authors have termed a "marginal" total R^2 (meaning that predicted scores are marginalized across random effects and thus are based on only the fixed portion from Equation [5], $\mathbf{x}_{ij}^{w\prime}\boldsymbol{\gamma}^w + \mathbf{x}_j^{b\prime}\boldsymbol{\gamma}^b$). The second alternative is to consider *all* variance attributable to v and m as explained, along with that attributable to f—that is, computing $R_t^{2(fvm)}$. This measure corresponds to what some authors have termed a "conditional" total R^2 (meaning that predicted scores are conditioned on random effects and thus are based on *both* the fixed and random portion from Equation [5], $\mathbf{x}_{ij}^{w\prime}\boldsymbol{\gamma}^w + \mathbf{x}_j^{b\prime}\boldsymbol{\gamma}^b + \mathbf{w}_{ij}'\mathbf{u}_j$). The marginal perspective is currently the more dominant view in psychology and may feel intuitive given that random effect variation is commonly termed "residual variance" at level-2 (see Footnote 9). The conditional perspective may feel foreign outside of the biostatistics field, where it primarily originated, due to the disciplinary divide in the dissemination of measures (see Issue 3). As for why a researcher may wish to use the conditional perspective to include variance attributable to v and m in the numerator, one argument for $R_t^{2(fvm)}$ given by Vonesh and Chinchilli (1997) is that "Typically in longitudinal studies, there tends to be greater variability between subjects rather than within subjects. Consequently, the [$R_t^{2(f)}$] may be somewhat undervalued since [it doesn't] account for the presence of subject-specific random effects. A moderately low value for [$R_t^{2(f)}$] may mislead the user into thinking the selected fixed effects fit the data poorly. Therefore, it is important that we also assess the fit of both the fixed and random effects based on the conditional mean response" (p. 423). As another example, suppose a researcher is studying math achievement for students (nested within classrooms) and is particularly interested in the effect of hours spent studying (e.g., Rights & Sterba, 2016). This effect might reasonably be expected to differ across classrooms due to any number of factors that may be unmeasured; for instance, classrooms may vary in how effectively material is taught and the degree to which independent study is expected of students. Similarly, classrooms likely have different baseline levels of math achievement, reflected by a random intercept. A researcher explicitly interested in both this slope and intercept heterogeneity would likely want to include such variance in an R^2 measure and thus report $R_t^{2(fvm)}$.

Though the marginal ($R_t^{2(f)}$) versus conditional ($R_t^{2(fvm)}$) distinction has always been framed as an "all-or-nothing" consideration (i.e., counting as explained *all* or *none* of the variation attributable to predictors/cluster means via random effects), a new measure we presented in Table 1 serves as a compromise by considering variance explained by predictors via random *slopes* explained but by cluster means via random *intercepts* unexplained—$R_t^{2(fv)}$. Consider, for instance, the example described above of predicting math scores from hours spent studying. Suppose a researcher were indeed interested in the heterogeneity in this effect across classrooms. This does not, however, imply there to be interest in the *intercept* heterogeneity in math scores across classrooms. A researcher interested in slope heterogeneity but not intercept heterogeneity may wish to report $R_t^{2(fv)}$, which conveniently can be interpreted as the proportion of variance explained by the predictors (because predictors explain the outcome via both f and v). Such a measure is also consistent with recent recommendations to develop measures that evaluate random effects in conjunction with fixed effects (Demidenko et al., 2012; Edwards et al., 2008; Jaeger et al., 2017; Kramer, 2005).

There is Substantive Need for Measures Representing Each Source of Explained Variance Individually: Rationale for $R_t^{2(m)}$, $R_t^{2(f_1)}$, $R_t^{2(f_2)}$

Researchers previously have been interested in making use of measures that represent a given source of explained variance in isolation. For instance, a pre-existing measure, $R_t^{2(v)}$, isolates the impact of predictors via random slope variation, or v (Aguinis & Culpepper, 2015). Aguinis and Culpepper (2015) have recommended that $R_t^{2(v)}$ be used to compare slope heterogeneity across studies and also used to assess the degree of clustering, in conjunction with the ICC, when the researcher is determining the need for a MLM.

Newly developed measures in our framework extend this principle more completely in providing measures that isolate other sources of explained variance. Specifically, $R_t^{2(m)}$ isolates the proportion of total variance explained by m, $R_t^{2(f_1)}$ isolates the proportion of total variance explained by f_1, and $R_t^{2(f_2)}$ isolates the proportion of total variance explained by f_2. Though we do not necessarily anticipate that researchers would be interested in reporting only one of these isolated-source measures, each can be useful to contrast with the others ($R_t^{2(v)}$ vs. $R_t^{2(m)}$ vs. $R_t^{2(f_1)}$ vs. $R_t^{2(f_2)}$) to get a broader understanding of the full decomposition. Also, because researchers already are widely using published versions of measures combining multiple sources of explained variance (e.g., $R_t^{2(fvm)}$ that includes f_1, f_2, v, and m as sources), it stands to reason that these researchers would also be interested in examining the proportion of variance explained solely by each, one at a time.

There is a Substantive Need for Having "Parallel" Total Versus Level-Specific Measures: Rationale for $R_w^{2(f_1)}$, $R_w^{2(v)}$, $R_b^{2(m)}$

For a given source of explained variance, it can be informative to consider how much it explains relative to the total variance as well as relative to level-specific (i.e., within or between) variance. It could be the case, for instance, that a given source explains a large proportion of level-specific variance, but explains little of the total variance, or vice versa (as seen later in the section titled Limitations of common practice of reporting level-specific measures $R_w^{2(f_1v)}$ and $R_b^{2(f_2)}$ without total measures). Thus, considering a level-specific measure in isolation does not inform one of the importance of a given source with respect to the total variance, whereas considering a total measure in isolation does not inform one of the importance of a given source with respect to level-specific variance. To allow researchers to consider both types of importance (total vs. level-specific) simultaneously, our framework provides pairs of measures that we term "parallel" in that they consider the same source of explained variance (numerator), but with different denominators. For instance, we introduce $R_w^{2(f_1)}$ to assess the amount of variance explained by level-1 predictors via fixed slopes relative to the within-cluster variance; one can compare this value with the total measure $R_t^{2(f_1)}$. Similarly, we provide $R_w^{2(v)}$ as a parallel to $R_t^{2(v)}$, and $R_b^{2(m)}$ as a parallel to $R_t^{2(m)}$. In the empirical examples, we further illustrate limitations of considering either total or level-specific measures in isolation and illustrate the utility of considering parallel measures.

Limitations of the Common Practice of Reporting Only a Single MLM R^2

As mentioned earlier, currently when researchers using MLM report an R^2, they tend to report exclusively a total measure *or* exclusively level-specific measure(s) (LaHuis et al., 2014). Currently, common choices for reporting a total measure are $R_t^{2(fvm)}$ or $R_t^{2(f)}$ analogs and common choices for reporting level-specific measures are $R_w^{2(f_1v)}$ and/or $R_b^{2(f_2)}$ analogs. Earlier we recommended that one need not choose a single R^2 to report because the measures in our framework provide complementary information and thus it is more informative to consider the suite of measures together. This can be accomplished by inspecting the decomposition of scaled variance into proportions graphic (e.g., Figure 1). In this section, we present simulated demonstrations, in Figures 2–5, that highlight limitations of the typical practice of reporting just one or a pair of these common measures. Furthermore, we describe how consulting the integrative framework can yield more informative substantive interpretations. Note that each of our simulated demonstrations in this section contains a single predictor simply for ease of graphical depiction of results in line plots; the points we make also apply to the context of multiple predictors.

Limitations of the Common Practice of Reporting Estimate of $R_t^{2(fvm)}$ Without Decomposing Total Variance

We first demonstrate limitations of the common practice of reporting exclusively a $R_t^{2(fvm)}$ (e.g., Xu's/Johnson's/Vonesh-Chinchilli's measure) without decomposing total variance to form our suite of total R^2 measures (as in, e.g., Payne, Lee, & Federmeier, 2015; Stolz, 2015). Figure 2 Panel A depicts four conditions. In each condition, the level-1 residual variance is the same and there is a single level-1 predictor. Further, each condition has an identical $R_t^{2(fvm)}$ of .80. Nonetheless, each condition corresponds to a unique substantive interpretation regarding the proportion of variance explained; namely, each condition involves different sources that contribute to explained variance. This is illustrated for each condition using both a line plot (wherein each thin [black] line is a cluster-specific regression line and the thick [red] line is the average regression line) and in a corresponding bar chart

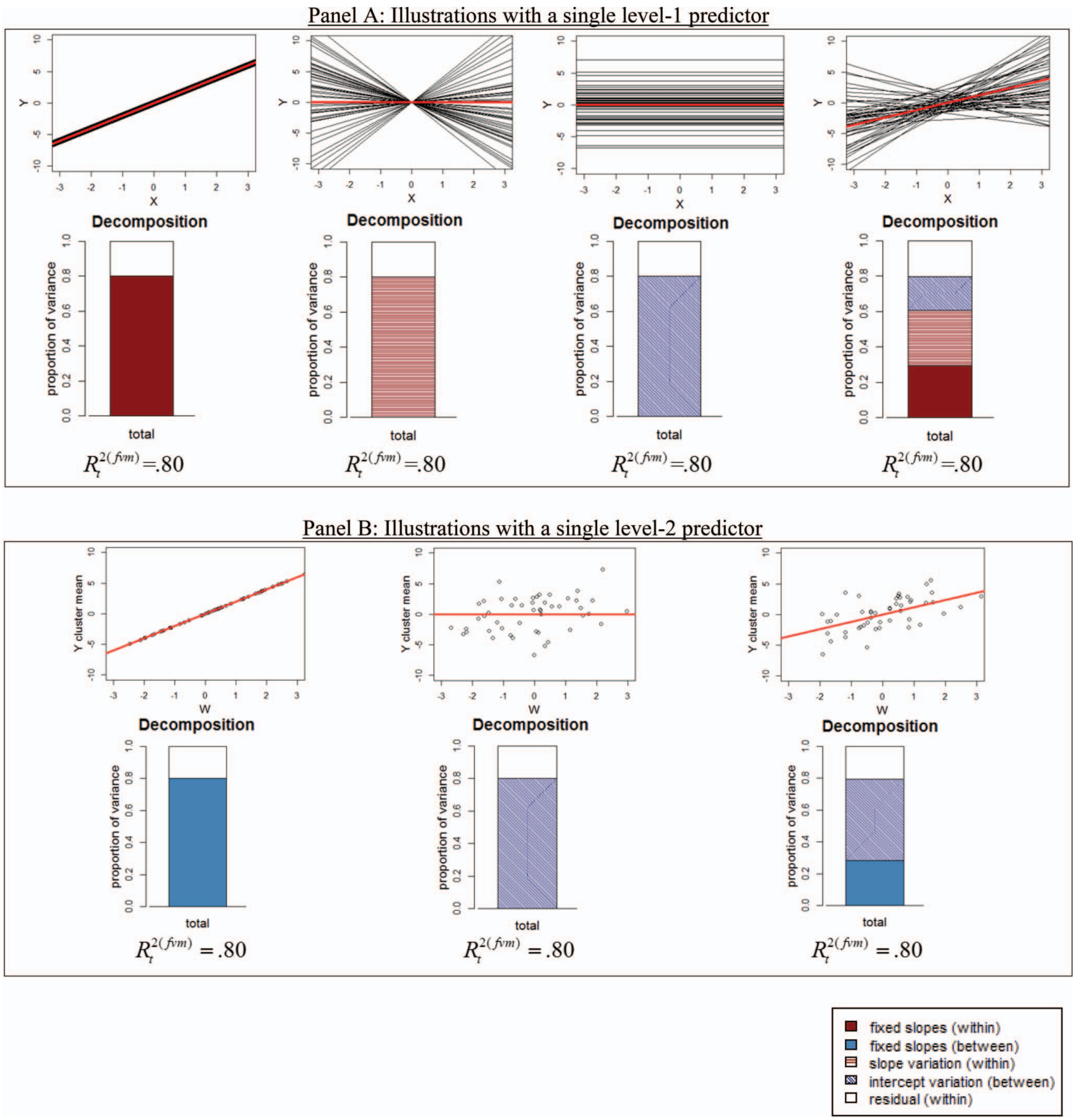

Figure 2. Limitations of the common practice of reporting a $R_t^{2(fvm)}$ measure analog (e.g., Vonesh-Chinchilli/Xu/Johnson's measure) without decomposing scaled total variance into proportions: Substantively different patterns yield the same $R_t^{2(fvm)}$. $R_t^{2(fvm)}$ was defined in Figure 1 and Table 1. Thin (black) lines = cluster-specific regression lines; thick (red) line = marginal (mean) regression line in panel A and regression line in panel B; dots = cluster-specific values. See the online article for the color version of this figure.

(showing the decomposition of scaled variance into proportions, previously defined in Figure 1). For instance, panel A, column 1 reflects a situation wherein $R_t^{2(fvm)}$ = .80 such that variance is explained solely by f_1, and no variance is explained via random effects (i.e., v or m). In contrast, panel A, columns 2 and 3 depict situations wherein again $R_t^{2(fvm)}$ = .80, but now variance is explained exclusively via random effects (v or m, respectively) and the fixed slope is actually equal to 0. Clearly these situations in Figure 2, panel A correspond with vastly different interpretations regarding the influence of the predictor on the outcome and the extent of across-cluster heterogeneity. In column 1, cluster membership yields no predictive information, whereas in column 3, the predictor yields no predictive information. In column 2, the predictor yields *cluster-specific* predictive information but no *marginal* information. These are extreme examples; a more realistic and nuanced situation is presented in panel A, column 4, wherein each of the aforementioned sources explains some variance in the outcome, combining to yield $R_t^{2(fvm)} = .80$.

Figure 2, panel B illustrates this same concept with a single *level-2* predictor. Note that for each of the line plots in panel B, the *y*-axis is now the cluster-specific outcome mean, each point is cluster specific, and the line represents the regression line for the level-2 predictor. Thus, greater vertical spread of these points about the regression line indicates greater intercept variance. Panel B, column 1 reflects a situation wherein variance is explained exclusively by f_2, panel B, column 2 reflects a situation wherein variance is explained exclusively by m, and panel B, column 3 reflects a situation with a mix of these sources. Despite the vastly different corresponding substantive interpretations, all three of these situations yield the same $R_t^{2(fvm)}$ of .80.

The key point of the demonstration in Figure 2 is that, without also considering our suite of total R^2 measures (readily visualized from the bar chart decomposition of scaled variance into proportions defined in Figure 1), it is not clear *in what way* variance is being explained by the model when only reporting $R_t^{2(fvm)}$.

Limitations of the Common Practice of Reporting an Estimate of $R_t^{2(f)}$ Without Decomposing Total Variance

Though $R_t^{2(f)}$ is simpler than $R_t^{2(fvm)}$ in the sense that the only sources of explained variance are f_1 and f_2, the common practice of reporting a $R_t^{2(f)}$ (e.g., Snijders-Bosker's/Vonesh-Chinchilli's/Johnson's measure) exclusively (as in, e.g., Engert, Plessow, Miller, Kirschbaum, & Singer, 2014; Lusby, Goodman, Yeung, Bell, & Stowe, 2016) can nonetheless be misleading. The limitations of this practice are illustrated in Figure 3. In Figure 3, panel A, again holding level-1 residual variance constant, we present four different situations with a level-1 predictor that yield the same $R_t^{2(f)}$, despite each corresponding to a unique substantive pattern of explained variance. In panel A, column 1, the total outcome variance is attributable to only two sources: f_1 and level-1 residuals. As can be seen in the panel A, column 1 line plot, a relatively modest slope yields an $R_t^{2(f)}$ of .40. In panel A, columns 2–4, however, the total outcome variance is now attributable to the sources mentioned for column 1 as well as at least one other—v (column 2), m (column 3), or the combination thereof (column 4). Notice that, despite the slope of the predictor being greater in columns 2–4, the $R_t^{2(f)}$ of .40 is the same as in column 1. One may have expected intuitively that panel A, columns 2–4—having a larger slope of x_{ij} than panel A, column 1 but having the same level-1 residual variance—would also have had a higher $R_t^{2(f)}$. However, this is not the case here because there is greater total variance in panel A, columns 2–4. Similarly, in panel B with a level-2 predictor, the first column consists of only variance attributable to level-1 residuals and f_2, whereas column 2 also consists of variance attributable to m. Despite the slope being much larger in column 2, they both have the same $R_t^{2(f)}$ of .40.

The key point in this Figure 3 illustration is that $R_t^{2(f)}$ reflects the proportion of variance explained by f *relative* to total variance—the latter of which is composed of several distinct components. Large slopes of predictors and a small level-1 residual variance does not mean $R_t^{2(f)}$ will similarly be large; if there is substantial variance attributable to other sources, $R_t^{2(f)}$ may still be quite small. This can be elucidated by interpreting $R_t^{2(f)}$ in the context of the other total R^2 measures in our framework by decomposing total scaled variance into proportions using the bar chart graphic (defined in Figure 1).

Limitations of the Common Practice of Reporting an Estimate of $R_w^{2(f_1v)}$ Without Decomposing Within Variance

We next consider the common practice of reporting a $R_w^{2(f_1v)}$ (e.g., Raudenbush-Bryk/Hox/Kreft-de-Leeuw/Vonesh-Chinchilli's measure) as an index of within-cluster variance explained, without decomposing within-cluster scaled variance into proportions to form our suite of within-cluster measures (as in, e.g., Sasidharan, Santhanam, Brass, & Sambamurthy, 2012; Wells & Krieckhaus, 2006). Figure 4 presents three generating conditions yielding the same $R_w^{2(f_1v)} = .50$ despite different substantive interpretations regarding the proportion of within-cluster variance explained. In Figure 4, within-cluster variance is explained exclusively by f_1 in column 1, by both f_1 and v in column 2, and exclusively by v in column 3. Thus, similar to the Figure 2 demonstration, when reporting $R_w^{2(f_1v)}$ in isolation it is not clear *in what way* variance is being explained. This can be assessed only by examining all three within-cluster measures in juxtaposition, which is straightforward to do using the bar chart graphic.

Limitations of the Common Practice of Reporting Estimates of Level-Specific Measures $R_w^{2(f_1v)}$ and $R_b^{2(f_2)}$ Without Total Measures

Lastly, we consider the common practice of reporting only level-specific measures $R_w^{2(f_1v)}$ and/or $R_b^{2(f_2)}$ (e.g., Raudenbush-Bryk/Hox/Kreft-de-Leeuw's measures) without simultaneously considering their relation to the *total* outcome variance (as in, e.g., Holland & Neimeyer, 2011; McCrae et al., 2008). Figure 5, panel A reflects a situation wherein $R_w^{2(f_1v)}$ is substantially smaller than $R_b^{2(f_2)}$ (as seen in the comparison of the middle and right bars). One might be tempted to conclude that this result implies that the level-2 predictors are "more important" than the level-1 predictors, such that less *total variance* is explained within-cluster than between-cluster. However, this is not true. What is true, in this case, is that the *proportion of within-cluster variance* that is explained is less than the *proportion of between-cluster variance* that is explained. In fact, in this illustration the opposite pattern holds for total variance—that is, much more is explained by within-cluster sources (f_1 and v) than between-cluster sources (f_2). This can be seen in Figure 5, panel A by comparing $R_w^{2(f_1v)}$ and $R_b^{2(f_2)}$ to their parallel *total* R^2 counterparts (namely, $R_t^{2(f_1v)}$—the proportion of total variance explained by f_1 and v—and $R_t^{2(f_2)}$—the

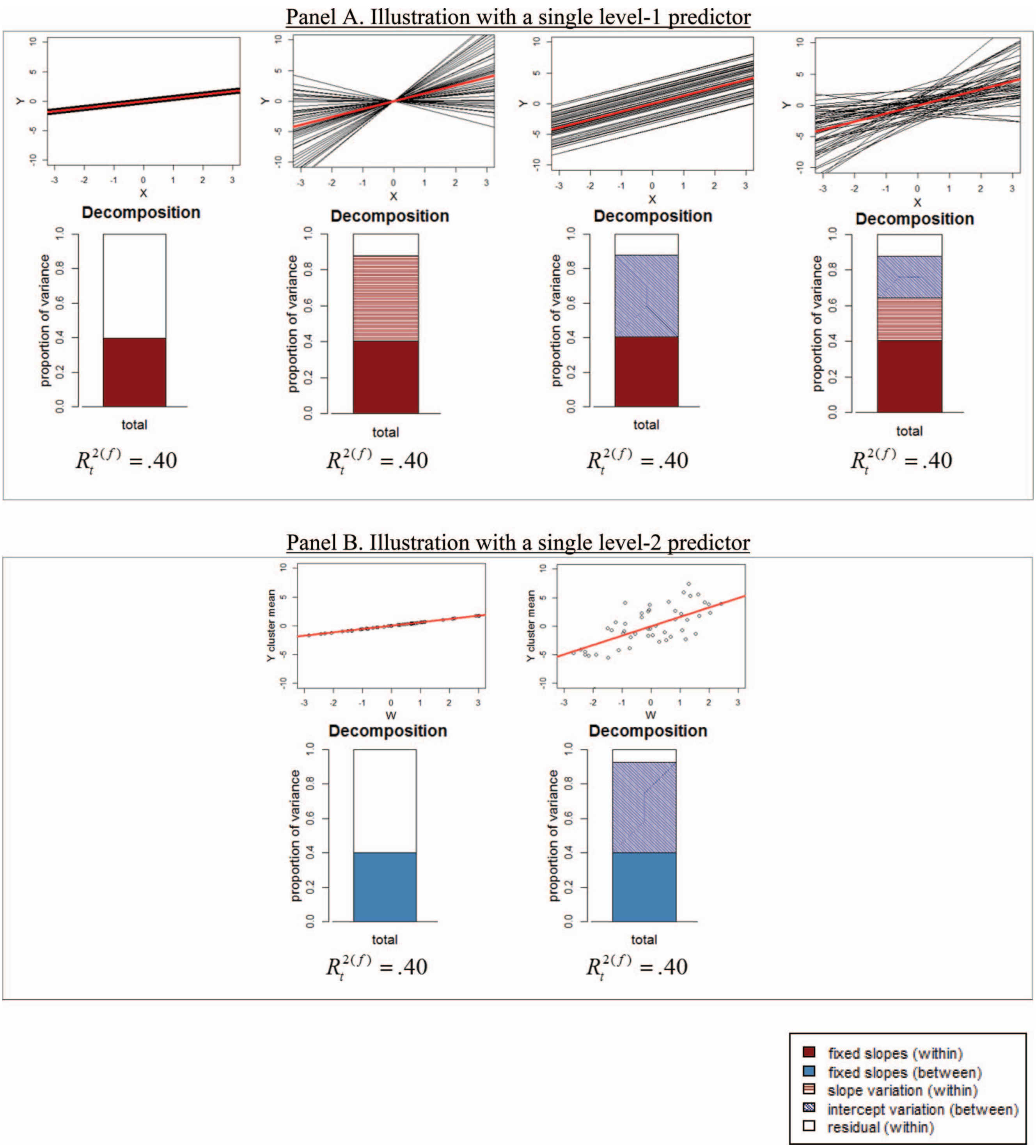

Figure 3. Limitations of the common practice of reporting a $R_t^{2(f)}$ measure analog (e.g., Snijders-Bosker/Vonesh-Chinchilli/Johnson's measure) without decomposing scaled total variance into proportions: Substantively different patterns yield the same $R_t^{2(f)}$. $R_t^{2(f)}$ was defined in Figure 1 and Table 1. Thin (black) lines = cluster-specific regression lines; thick (red) line = marginal (mean) regression line in panel A and regression line in panel B, dots = cluster-specific values. See the online article for the color version of this figure.

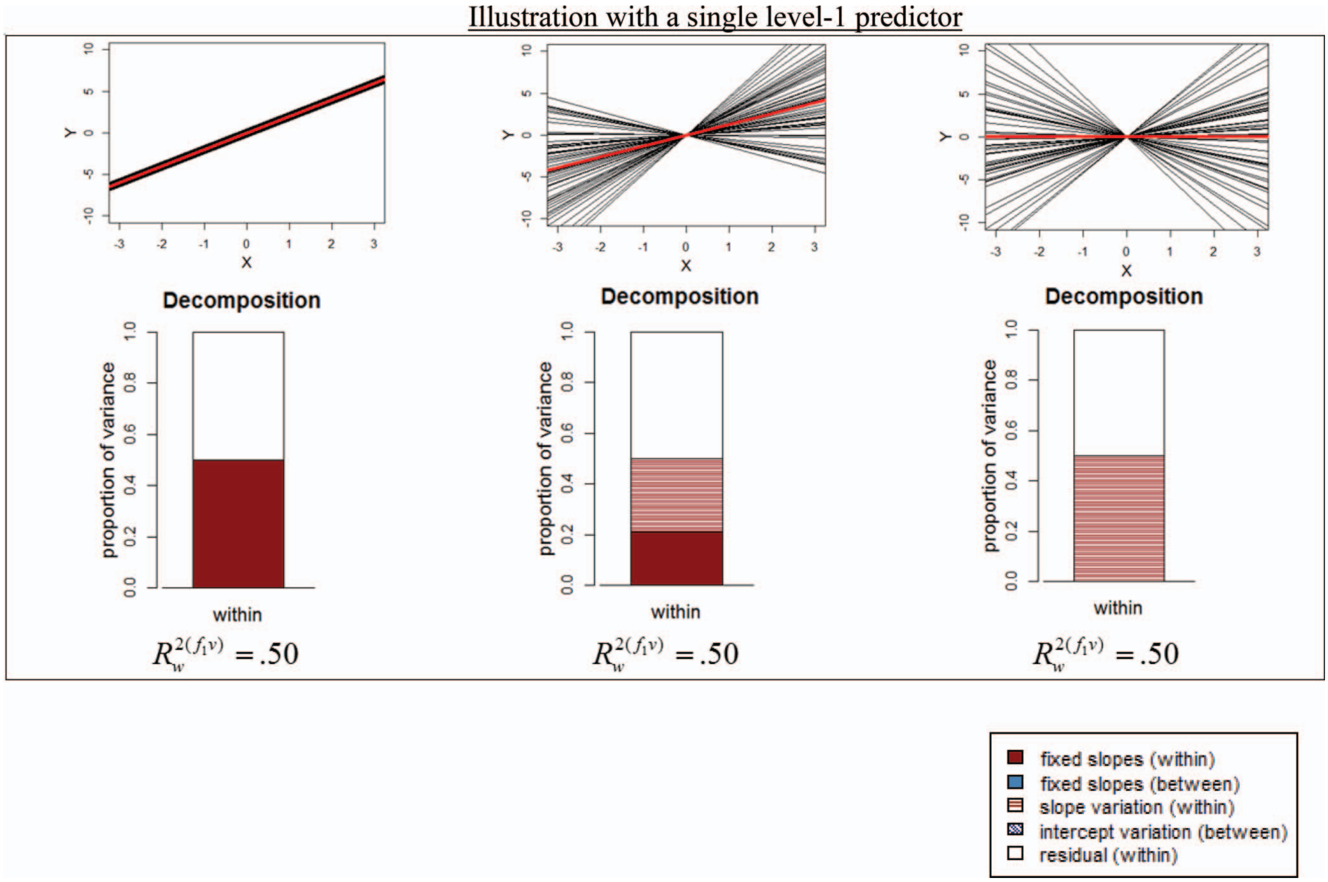

Figure 4. Limitations of the common practice of reporting a $R_w^{2(f_1v)}$ measure analog (e.g., Raudenbush-Bryk/Hox/Kreft-de Leeuw/Vonesh-Chinchilli's measure) without decomposing scaled within-variance into proportions: Substantively different patterns yield the same $R_w^{2(f_1v)}$. $R_w^{2(f_1v)}$ was defined in Figure 1 and Table 1. Thin (black) lines = cluster-specific regression lines, thick (red) line = marginal (mean) regression line. See the online article for the color version of this figure.

proportion of total variance explained by f_2). Note that this pattern can also be reversed; Figure 5, panel B reflects a situation wherein there is very little within-cluster variance and $R_w^{2(f_1v)} > R_b^{2(f_2)}$ but $R_t^{2(f_1v)} < R_t^{2(f_2)}$. The key point to this illustration in Figure 5 is that restricting focus to measures representing purely within- and/or purely between-cluster variance explained risks misunderstanding as it tells a researcher little about the proportion of the *total* variance explained.

Software Implementation

As mentioned previously (Issue 4), a current impediment to the use of R^2 measures for MLM in practice is the lack of available software (Bickel, 2007; Demidenko et al., 2012; Edwards et al., 2008; Jaeger et al., 2017; Kramer, 2005; LaHuis et al., 2014; McCoach & Black, 2008; Nakagawa & Schielzeth, 2013; Snijders & Bosker, 2012). Demidenko et al. (2012) emphasized that such measures should be "computed as a standard output [in] software every time a mixed model or a meta-analysis model is estimated" (p. 967).

To aid researchers in computing measures in our framework, we developed an R function, *r2MLM*. With this function, a user inputs all MLM estimates and raw data. The function then outputs the following: all 12 R^2 measures listed in Table 1, all decompositions illustrated in Figure 1, and a bar chart to visualize them (such as that in Figure 1). In online supplemental Appendix A, R code is provided both for the function itself and for example input.

Empirical Examples

In this section, we reanalyze three empirical examples from popular MLM textbooks (Hox, 2010; Kreft & de Leeuw, 1998; Snijders & Bosker, 1999). In these reanalyses, we highlight useful insights and information gained by computing the suite of R^2 measures in our framework with *r2MLM* software and visualizing the decomposition of scaled variance into proportions.[14] The original analyses of Examples 1 and 2 did not present R^2 measures.

[14] Note that, in the textbooks, some of the below examples included level-1 predictors that were not cluster-mean-centered. To facilitate a decomposition of scaled variance into proportions within and between cluster, we cluster-mean-centered level-1 variables while adding the cluster mean as a level-2 variable. Also, for substantive reasons, we included a random slope for parental education in Empirical Example 1.

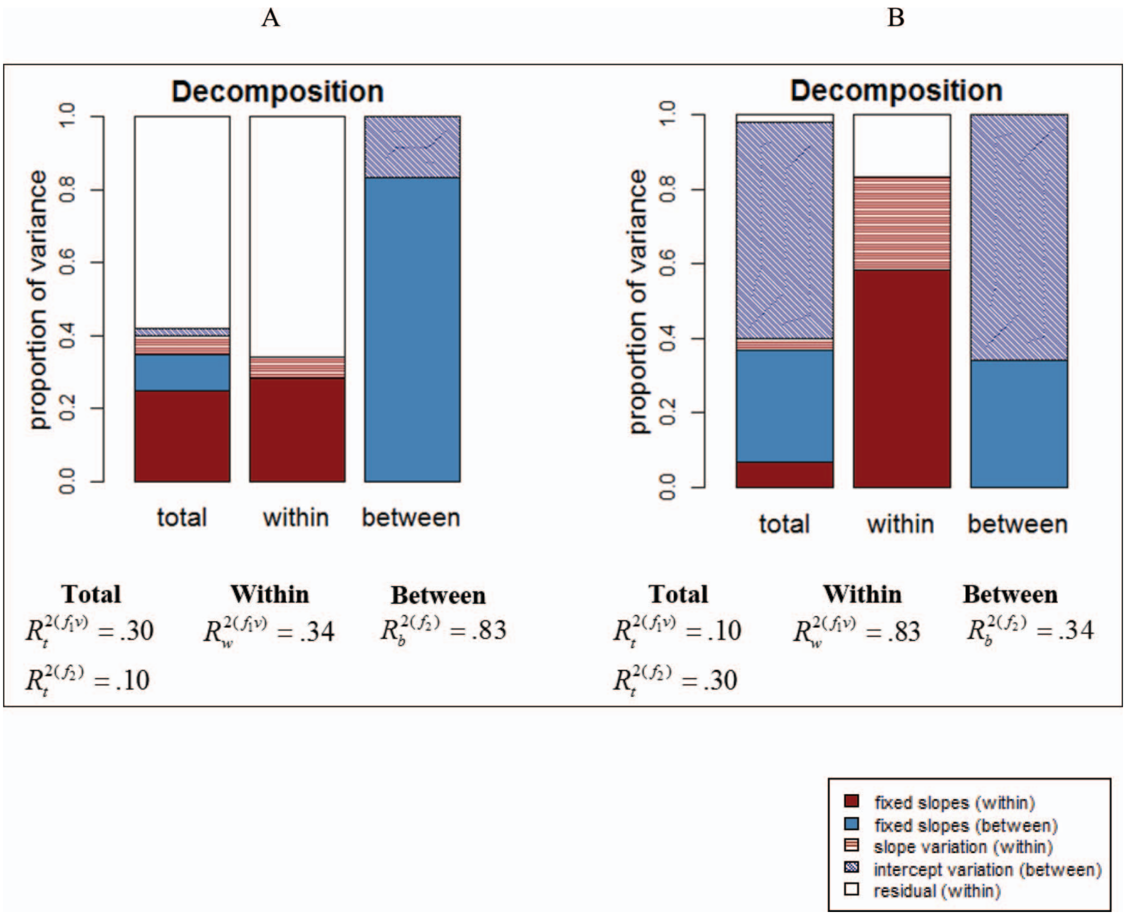

Figure 5. Limitations of the common practice of reporting level-specific $R_w^{2(f_1v)}$ and/or $R_b^{2(f_2)}$ measure analogs (e.g., Raudenbush-Bryk measures) without considering total measures. $R_t^{2(f_1v)}$ = proportion of total variance explained by f_1 and v; $R_t^{2(f_2)}$ = proportion of total variance explained by f_2; $R_w^{2(f_1v)}$ and $R_b^{2(f_2)}$ were defined in Figure 1 and Table 1. See the online article for the color version of this figure.

The original analysis of Example 3 presented only a subset of the R^2's provided by our framework.

We fit all models using SAS Proc Mixed with REML estimation. Here, to conserve space, we focus on results pertaining to R^2's and decompositions. For each example, the complete set of parameter estimate and standard error results can be found in online supplemental Appendix B. We summarize t-test results for fixed effects here; test results for variance components are in online supplemental Appendix B.

Empirical Example 1: Kreft & de Leeuw (1998), Predicting Math Scores

In this first example, we reanalyze Kreft and de Leeuw's (1998) multilevel example with 519 students nested within 23 schools wherein math scores are predicted from time spent on homework and parental education. Specifically, our predictors include school-mean-centered time spent on homework, school-mean-centered parental education, school-mean time spent on homework, and school-mean parental education. Across-school heterogeneity was modeled with a random intercept and random slopes of both level-1 predictors. Results indicated significant ($p < .05$) fixed effects for cluster-mean-centered time spent on homework and parental education and cluster-mean parental education.

Figure 6 displays bar charts of the total, within-cluster, and between-cluster decompositions of scaled variance into proportions. The first thing to note is that $R_t^{2(fvm)}$ is fairly high (.56) whereas $R_t^{2(f)}$ is fairly low (.20). Considering both of these simultaneously clarifies that a large proportion of variance is attributable to random effect variation. In particular, there is a sizable amount of slope heterogeneity, reflected by the $R_t^{2(v)}$ of .24. If this slope heterogeneity in addition to the marginal effects is deemed of interest, one may wish to focus on the new measure $R_t^{2(fv)}$ which is .44, meaning that 44% of the total variance in math scores is explained by the predictors via f and v.

Considering the within-cluster measures, previous common practice was to report just a $R_w^{2(f_1v)}$ (here, .43) analog, but this risks leading researchers to erroneously assume that marginal effects of parental education and time spent on homework explain a large amount of within-cluster variance, which is not the case because the new measure $R_w^{2(f_1)}$ is quite low (.12).

Considering the between-cluster measures, note that $R_b^{2(f_2)}$ is fairly large (.47). Looking at the decomposition of total scaled variance into proportions, however, clarifies that little of the total variance is actually explained by level-2 predictors (11%) because the amount of overall between-cluster variance is relatively small (23% of the total). Additionally, juxtaposing the total measures versus level-specific measures clarifies that although $R_b^{2(f_2)}$ is over twice as large as $R_w^{2(f_1 v)}$ this doesn't imply that f_2 explains more of the *total* variance than f_1 (in fact, their proportions of total variance explained are about equal).

Empirical Example 2: Snijders and Bosker (1999), Predicting Language Scores

Next, we reanalyze Snijders and Bosker's (1999) multilevel example with 2,287 students nested within 131 schools wherein they predict language scores from verbal IQ, socioeconomic status (SES), and within-school and between-school interactions thereof. Specifically, our predictors include school-mean-centered verbal IQ, school-mean-centered SES, school-mean verbal IQ, school-mean verbal IQ × school-mean SES, school-mean SES, and school-mean-centered verbal IQ × school-mean-centered SES. We included a random intercept and a random slope of school-mean-centered verbal IQ. Results indicated significant ($p < .05$) fixed effects of all predictors except the latter two.

Figure 7 displays the decomposition of scaled variance into proportions for Empirical Example 2. In contrast to Empirical Example 1, here there is not a large difference between $R_t^{2(fvm)}$ (.52) and $R_t^{2(f)}$ (.41), indicating that outcome variance is not driven as strongly by random effect variation and is more attributable to f. This fact may have been overlooked if a researcher had just focused on statistical significance of random slope variance, because it is statistically significant. In fact, if we inspect the effect size $R_t^{2(v)}$, we see it is actually quite small (.01) and perhaps substantively unimportant. Thus, in this example, we have evidence that the marginal effects of the predictors are of primary importance, whereas there is little heterogeneity in effects. When considering the within-cluster measures, $R_w^{2(f_1 v)}$ (.39) and $R_w^{2(f_1)}$ (.38) are very similar, unlike for Empirical Example 1. When considering the between-cluster measures, note that $R_b^{2(f_2)}$ is again larger than any of the within-cluster R^2 measures, despite the fact that f_2 explains much less of the total variance than does f_1. This is similar to the situation depicted in the Figure 5 demonstration.

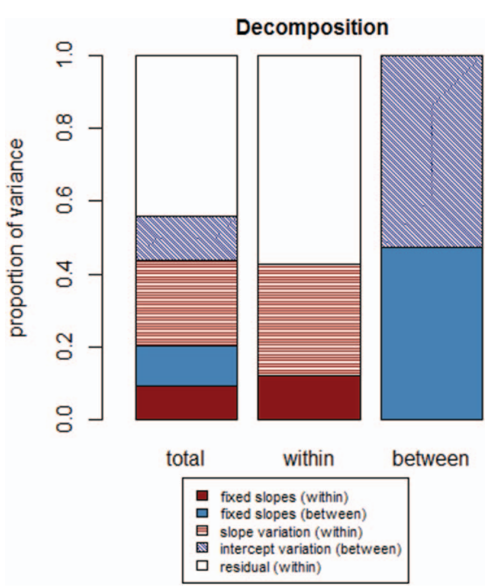

Figure 6. Empirical Example 1: Results from model predicting math scores (Kreft & de Leeuw, 1998). See the online article for the color version of this figure.

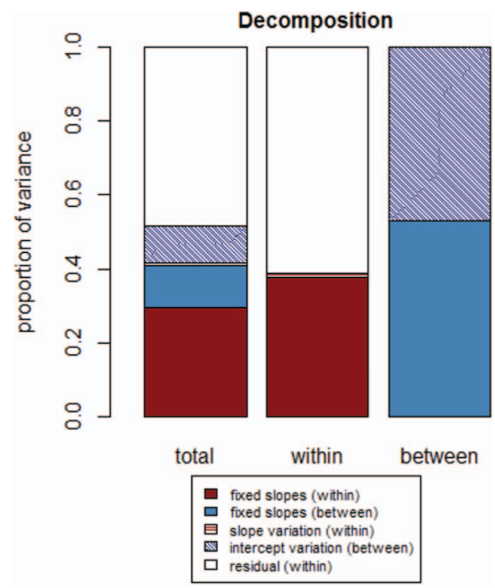

Figure 7. Empirical Example 2: Results from model predicting language scores (Snijders & Bosker, 2012). See the online article for the color version of this figure.

Empirical Example 3: Hox (2010), Predicting Popularity

Lastly, we reanalyze Hox's (2010) example with 2,000 students nested within 100 classrooms wherein he predicts popularity from extraversion, sex, teacher experience, and the cross-level interaction of extraversion and teacher experience.[15] Specifically, our predictors include class-mean-centered extraversion, class-mean-centered sex, teacher experience, and class-mean-centered extraversion × teacher experience. A random intercept and slope of extraversion were modeled. All mean slopes were significant.

The decomposition of scaled variance into proportions for Empirical Example 3 is in Figure 8. Here $R_t^{2(fvm)}$ is fairly high (.71) whereas $R_t^{2(f)}$ is just over half as large (.41). The $R_t^{2(m)}$ of .29 clarifies that much of this total variance is attributable to cluster means via random intercept variation, whereas the $R_t^{2(v)}$ of .01 indicates that little variance is explained by predictors via slope heterogeneity, similar to Empirical Example 2 but different from Empirical Example 1. When examining the level-specific measures, note that the proportion of within-cluster variance explained

[15] Note that Hox (2010) reports that this dataset was artificially constructed for pedagogical purposes in the textbook.

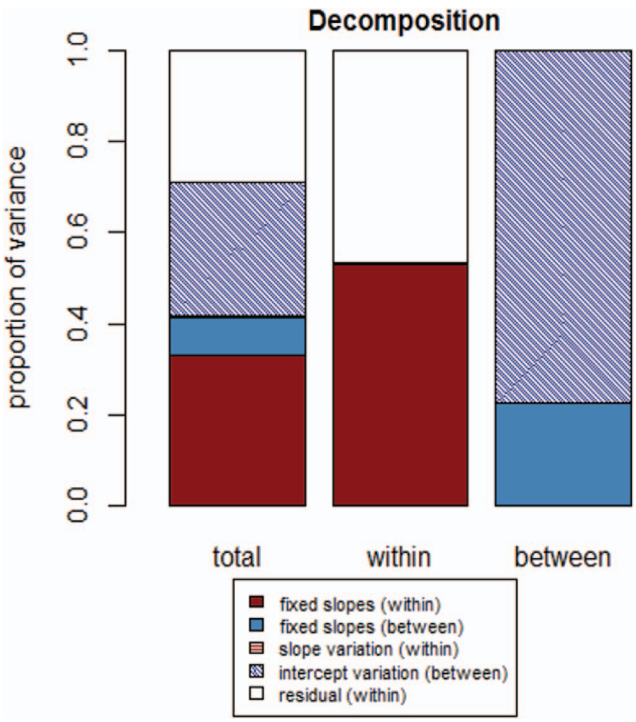

Figure 8. Empirical Example 3: Results from model predicting popularity (Hox, 2010). See the online article for the color version of this figure.

by f_1 and v is greater than the proportion of between-cluster variance explained by f_2 (i.e., $R_w^{2(f_1 v)} > R_b^{2(f_2)}$). In contrast to the previous two empirical examples, this same pattern holds for the parallel *total* measures in the Figure 8 bar charts. Thus, we do not have a situation like that depicted in the Figure 5 pedagogical demonstration.

For all three examples, had we considered only point estimates and statistical significance of slopes and random effect (co)variances, it would have been difficult to determine what potential sources of variance are most important. Having a graphical decomposition and set of R^2 measures gives a detailed suite of effect sizes that indicate the degree to which total versus within-cluster versus between-cluster variance in math/language/popularity can be attributed to different sources.

Discussion

Motivated by researchers' continued interest in R^2 measures for MLM (Bickel, 2007; Edwards et al., 2008; Jaeger et al., 2017; Johnson, 2014; Kramer, 2005; LaHuis et al., 2014; Nakagawa & Schielzeth, 2013; Orelien & Edwards, 2008; Recchia, 2010; Roberts et al., 2011; Wang & Schaalje, 2009; Xu, 2003; Zheng, 2000), several such measures have been developed. Nonetheless, the current methodological literature on R^2 measures for MLM had suffered from several issues: Issue 1) analytic relationships and equivalencies among popular existing measures in the population had not been established; Issue 2) there were gaps in the availability of measures needed to answer substantive questions because a completely full partitioning of variance had not been used to create measures; Issue 3) a unifying approach to interpreting and choosing among measures had not been supplied; and Issue 4) measures were infrequently and inconsistently available in software. In this article, we have addressed each of these concerns in the following manner. We developed an integrative framework of R^2 measures for models with random intercepts and/or slopes that is based on a complete decomposition of scaled variance into proportions (see Tables 1–2). We addressed Issue 1 by analytically showing how 10 existing R^2 measures (from the social sciences as well as biostatistics) correspond to the same population quantities as five measures in our framework (see Table 3 and Appendix B). We addressed Issue 2 by reformulating contents of our framework to create additional measures that answer novel substantively relevant research questions. We addressed Issue 3 by developing an integrated graphical representation of the suite of MLM R^2 measures to straightforwardly communicate their relationships (see Figure 1) and by providing four demonstrations of the advantages of an integrated interpretation of a suite of MLM R^2 measures, rather than the more common approach of interpreting a single measure in isolation (see Figures 2–5). Finally, we addressed Issue 4 by creating an R function, *r2MLM*, that computes all measures and decompositions in our framework and we illustrated this function using three empirical examples drawn from popular MLM textbooks (Figures 6–8). Next we discuss several extension topics.

Extensions and Future Directions

Following widespread recommendations, here we decided to ensure unconflated slopes of level-1 predictors by focusing on the MLM specification with cluster-mean-centered level-1 predictors. In this context, each predictor has (and explains) either purely level-1 or level-2 variance. If researchers wish to fit MLMs with non-cluster-mean-centered level-1 variables, then level-specific R^2 measures would likely no longer be of interest but total R^2 measures would likely still be of interest.[16] It is important to remember that fitted MLMs have stricter assumptions when not cluster-mean-centering and so, as usual, researchers fitting such MLMs need to ensure their model assumptions are plausible. Our R function *r2MLM* can also be used to compute total R^2 measures in the absence of cluster-mean-centering; these supplementary measures are given in Table 5 and they use a model-implied total outcome variance derivation provided in Appendix A Section A2. For the Table 5 measures, f retains the same interpretation as in

[16] If researchers unconflate level-1 predictors' effects using a "contextual effect" approach (see Raudenbush & Bryk, 2002 or Snijders & Bosker, 2012) rather than cluster-mean-centering, besides the measures in Table 5 they may be also interested in separately estimating $R_t^{2(f_1)}$ and $R_t^{2(f_2)}$. Although measures shown in Table 5 do not split f into within- and between-cluster components, the within-cluster component f_1 can be computed by separately (i.e., post-model-fitting) cluster-mean-centering each level-1 predictor and taking their (co)variances to compute $\mathbf{\Phi}^w$ (see Equation 10) and then inserting the fixed component of each level-1 predictor's slope into $\boldsymbol{\gamma}^w$. The between-cluster component of f (i.e., f_2) can be computed by taking the (co)variances of cluster-means of each level-1 predictor to compute $\mathbf{\Phi}^b$ and then adding to the fixed component of the slope of each cluster-mean (i.e., the contextual effect) the fixed component of the slope of each level-1 predictor, and placing the resultant between-effect into $\boldsymbol{\gamma}^b$. If such researchers additionally wish to compute level-specific measures, we suggest that they use a cluster-mean-centered MLM together with the Table 1 formulas.

Table 5
Supplement to Framework: Total R^2 Formulas for Multilevel Models (MLMs) Not Employing Cluster-Mean-Centering

Measure	Definition (Interpretation)
$R_t^{2(f)} = \dfrac{\boldsymbol{\gamma}'\boldsymbol{\Phi}\boldsymbol{\gamma}}{\boldsymbol{\gamma}'\boldsymbol{\Phi}\boldsymbol{\gamma} + tr(\mathbf{T\Sigma}) + \mathbf{m}'\mathbf{Tm} + \sigma^2}$	Proportion of total outcome variance explained by *predictors via fixed slopes*
$R_t^{2(v)} = \dfrac{tr(\mathbf{T\Sigma})}{\boldsymbol{\gamma}'\boldsymbol{\Phi}\boldsymbol{\gamma} + tr(\mathbf{T\Sigma}) + \mathbf{m}'\mathbf{Tm} + \sigma^2}$	Proportion of total outcome variance explained by *predictors via random slope variation/covariation*
$R_t^{2(m)} = \dfrac{\mathbf{m}'\mathbf{Tm}}{\boldsymbol{\gamma}'\boldsymbol{\Phi}\boldsymbol{\gamma} + tr(\mathbf{T\Sigma}) + \mathbf{m}'\mathbf{Tm} + \sigma^2}$	Proportion of total outcome variance explained by *cluster-specific outcome means* (beyond that attributable to predictors via fixed slopes and random slope variation/covariation)
$R_t^{2(fv)} = \dfrac{\boldsymbol{\gamma}'\boldsymbol{\Phi}\boldsymbol{\gamma} + tr(\mathbf{T\Sigma})}{\boldsymbol{\gamma}'\boldsymbol{\Phi}\boldsymbol{\gamma} + tr(\mathbf{T\Sigma}) + \mathbf{m}'\mathbf{Tm} + \sigma^2}$	Proportion of total outcome variance explained by *predictors via fixed slopes and random slope variation/covariation*
$R_t^{2(fvm)} = \dfrac{\boldsymbol{\gamma}'\boldsymbol{\Phi}\boldsymbol{\gamma} + tr(\mathbf{T\Sigma}) + \mathbf{m}'\mathbf{Tm}}{\boldsymbol{\gamma}'\boldsymbol{\Phi}\boldsymbol{\gamma} + tr(\mathbf{T\Sigma}) + \mathbf{m}'\mathbf{Tm} + \sigma^2}$	Proportion of total outcome variance explained by *predictors via fixed slopes and random slope variation/covariation* and by *cluster-specific outcome means*

Note. These measures are described only in the Extensions and Future Directions section of our Discussion. These measures use the Appendix A Equation A11 model-implied variance in their denominators, which is applicable when not cluster-mean-centering level-1 predictors. The article provides reasons why we recommend using the framework of measures in Table 1, together with fitting a MLM with cluster-mean-centered level-1 predictors, rather than using the supplemental formulas in Table 5, together with fitting a MLM without cluster-mean-centered level-1 predictors.

Table 1, though the interpretation of v and m is less clean: v can now explain variance at both levels (not just within-cluster), and m now reflects between-cluster variance that is not attributable to f or v. At the end of each Appendix B Sections B1–B6 we include additional derivations showing relationships between each published total R^2 and our corresponding total R^2 when not cluster-mean-centering level-1 predictors.

Second, some pre-existing R^2 measures were not discussed or related to our framework in the current article. For generality, we excluded measures that required fitted models to have no random slopes. Perhaps the most common example of such a measure that assumes only random intercepts is that developed by Snijders and Bosker (1999 [Equation 7.2]). Snijders and Bosker (1999) noted that, if one has a random slope model of interest, one could refit the model without random slopes (i.e., fitting a more constrained model) to compute the measure. In line with Jaeger et al. (2017) and Johnson (2014) (see also Gurka, Edwards, & Muller, 2011) we feel, however, that it is preferable to have an R^2 measure that fully reflects the structure of the fitted model. Nonetheless, it can be shown that this Snijders and Bosker (1999) measure is analogous to $R_t^{2(f)}$, even when the full fitted model has random slopes and the measure is computed by constraining slope variances to 0. Additionally, to maintain a manageable scope (Appendix B already contains 10 sections of derivations) we did not relate our framework to existing individual measures that were not commonly used in MLM applications. Future research can relate special case measures from our framework to these other existing individual measures that are, for instance, computed under a Bayesian framework using posterior simulation (Gelman & Pardoe, 2006), derived from likelihood ratio tests or Wald tests (Edwards et al., 2008; Jaeger et al., 2017; Magee, 1990), represented as proportion reduction in deviance or in entropy (Zheng, 2000), or represented as a weighted least squares average of cluster-specific R^2's (Roberts et al., 2011).

Third, though all R^2 measures in this article were specifically denoted for the most commonly modeled multilevel structure—a two-level hierarchical design—our approach to variance partitioning could be implemented in models with more than two levels. For instance, in a three-level context, variance can be explained by level-1, level-2, and level-3 predictors via fixed components of slopes. A three-level total R^2 could be computed with the three-level model-implied total outcome variance in the denominator and one (or some combination of) model-implied variance(s) attributable to a given source (or combination of sources) of explained variance in the numerator. Level-specific R^2 measures could similarly be computed, but with the model-implied outcome variance only at level-1, or level-2, or level-3 in the denominator.

Fourth, we focused on MLM with the most commonly used type of outcome distribution—normal outcomes—though these measures could be extended to a generalized linear mixed model (GLMM) framework with, for instance, binary outcomes. This has been done for selected R^2 measures from our framework by Nakagawa and Schielzeth (2013) and Johnson (2014) (see also Jaeger et al., 2017). The approach in the former two publications is similar to our own in that it involves partitioning of model-implied variance to form both the outcome variance (denominator) and the explained variance (numerator). The difference is that, for GLMM, the definition of the level-1 residual variance is particular to the error distribution and link function used. For instance, to obtain our framework of R^2 measures for GLMMs with binary outcomes using the binary logit link, σ^2 would be replaced by $\pi^2/3$ in our Table 1.

Fifth, all measures in the current article were defined and demonstrated for evaluation of a single hypothesized model at a time, which is currently the most common use of MLM R^2 (Gelman & Pardoe, 2006). Nonetheless, researchers may be interested in how a MLM R^2 increases with the addition of parameters. Researchers can compute any of the measures in Table 1 for two models under comparison and compute the difference in each measure. For instance, one can compare a reduced Model A to a full Model B that adds a fixed slope of a single level-1 predictor. If one is interested in how much more variance is explained in Model B by level-1 predictors via fixed slopes one can compute the increase in $R_t^{2(f_1)}$ and $R_w^{2(f_1)}$. However, a full treatment of how to interpret and use these differences in practice, and how they relate to prior literature and recommendations on model selection with MLM R^2's, is outside of the scope of the present article.

Lastly, in the current article, we focused on obtaining point estimates of proportions of variance explained rather than on the computation of standard errors or confidence intervals. Our focus mirrors that of the current MLM R^2 literature more broadly. To our knowledge, only one MLM R^2 measure is accompanied by an analytic solution for computing confidence intervals (Edwards et al., 2008); however, it is limited in that it evaluates only the fixed effects as it is computed from a Wald test of the fixed components. More broadly, the estimates of the measures in Table 1 (which together evaluate the importance of both fixed and random effects) would each have their own unique degree of precision in a sample. Measure-specific confidence intervals could be computed, for instance, via one of several available kinds of multilevel bootstrapping (see Goldstein, 2011) or with Bayesian estimation via Markov-Chain Monte Carlo (MCMC) sampling. Both approaches have been suggested as viable options that should be investigated (LaHuis et al., 2014; Nakagawa & Schielzeth, 2013). Note that, when using bootstrapping in this context, researchers would want to ensure that level-1 predictors remain cluster-mean-centered within every bootstrap resample.

Conclusions

This article responds to widespread methodological recommendations to move beyond consideration of only statistical significance when fitting models (APA, 2009; Cumming, 2012; Harlow et al., 1997; Kelley & Preacher, 2012; Panter & Sterba, 2011; Wilkinson & APA Task Force on Statistical Inference, 1999). With the general framework for computing MLM R^2 measures provided here, researchers have the flexibility to choose among the different types of variance that can be explained (total vs. within-cluster vs. between-cluster) and the different sources that can contribute to explanation. Moreover, researchers can graphically visualize this expanded set of options simultaneously. This leads to an integrated interpretation of measures from our framework. It is our hope that, by extending and relating existing measures across disciplines, this unifying framework of R^2 effect size measures will aid researchers in efforts to convey practical significance.

References

Aguinis, H., & Culpepper, S. A. (2015). An expanded decision-making procedure for examining cross-level interaction effects with multilevel modeling. *Organizational Research Methods, 18,* 155–176. http://dx.doi.org/10.1177/1094428114563618

American Psychological Association. (2009). *Publication manual of the American Psychological Association* (6th ed.). Washington, DC: American Psychological Association.

Bickel, R. (2007). *Multilevel analysis for applied research. It's just regression!* New York, NY: Guilford Press.

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

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Erlbaum.

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

Cumming, G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. New York, NY: Routledge.

Demidenko, E., Sargent, J., & Onega, T. (2012). Random effects coefficient of determination for mixed and meta-analysis models. *Communications in Statistics Theory and Methods, 41,* 953–969. http://dx.doi.org/10.1080/03610926.2010.535631

Edwards, L. J., Muller, K. E., Wolfinger, R. D., Qaqish, B. F., & Schabenberger, O. (2008). An R2 statistic for fixed effects in the linear mixed model. *Statistics in Medicine, 27,* 6137–6157. http://dx.doi.org/10.1002/sim.3429

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. http://dx.doi.org/10.1037/1082-989X.12.2.121

Engert, V., Plessow, F., Miller, R., Kirschbaum, C., & Singer, T. (2014). Cortisol increase in empathic stress is modulated by emotional closeness and observation modality. *Psychoneuroendocrinology, 45,* 192–201. http://dx.doi.org/10.1016/j.psyneuen.2014.04.005

Fitzmaurice, G. M., Laird, N. M., & Ware, J. H. (2011). *Applied longitudinal analysis* (2nd ed.). Hoboken, N: Wiley.

Gelman, A., & Hill, J. (2007). *Data analysis using regression and hierarchical/multilevel models*. New York, NY: Cambridge.

Gelman, A., & Pardoe, I. (2006). Bayesian measures of explained variance and pooling in multilevel (hierarchical) models. *Technometrics, 48,* 241–251. http://dx.doi.org/10.1198/004017005000000517

Goldstein, H. (2011). Bootstrapping in multilevel models. In J. J. Hox & J. K. Roberts (Eds.), *Handbook of advanced multilevel analysis* (pp. 163–172). New York, NY: Routledge.

Gurka, M. J., Edwards, L. J., & Muller, K. E. (2011). Avoiding bias in mixed model inference for fixed effects. *Statistics in Medicine, 30,* 2696–2707. http://dx.doi.org/10.1002/sim.4293

Harlow, L. L., Mulaik, S. A., & Steiger, J. H. (Eds.). (1997). *What if there were no significance tests?* Mahwah, NJ: Erlbaum.

Hedeker, D., & Gibbons, R. D. (2006). *Longitudinal data analysis*. Hoboken, NJ: Wiley.

Holland, J. M., & Neimeyer, R. A. (2011). Separation and traumatic distress in prolonged grief: The role of cause of death and relationship to the deceased. *Journal of Psychopathology and Behavioral Assessment, 33,* 254–263. http://dx.doi.org/10.1007/s10862-010-9214-5

Hox, J. J. (2002). *Multilevel analyses: Techniques and applications*. Mahwah, NJ: Erlbaum.

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

Jaeger, B. C., Edwards, L. J., Das, K., & Sen, P. K. (2017). An R^2 statistic for fixed effects in the generalized linear mixed model. *Journal of Applied Statistics, 44,* 1086–1105. http://dx.doi.org/10.1080/02664763.2016.1193725

Johnson, P. C. (2014). Extension of Nakagawa & Schielzeth's R^2_{GLMM} to random slopes models. *Methods in Ecology and Evolution, 5,* 944–946. http://dx.doi.org/10.1111/2041-210X.12225

Kelley, K., & Preacher, K. J. (2012). On effect size. *Psychological Methods, 17,* 137–152. http://dx.doi.org/10.1037/a0028086

King, G. (1986). How not to lie with statistics: Avoiding common mistakes in quantitative political science. *American Journal of Political Science, 30,* 666–687. http://dx.doi.org/10.2307/2111095

Kramer, M. (2005). R^2 statistics for mixed models. *2005 Proceedings of the Conference on Applied Statistics in Agriculture* (pp. 148–160). Manhattan, KS: Kansas State University.

Kreft, I. G., & de Leeuw, J. (1998). *Introducing multilevel modeling*. Thousand Oaks, CA: Sage. http://dx.doi.org/10.4135/9781849209366

Kvålseth, T. O. (1985). Cautionary note about R^2. *The American Statistician, 39,* 279–285.

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

Lusby, C. M., Goodman, S. H., Yeung, E. W., Bell, M. A., & Stowe, Z. N. (2016). Infant EEG and temperament negative affectivity: Coherence of vulnerabilities to mothers' perinatal depression. *Development and Psychopathology, 28,* 895–911. http://dx.doi.org/10.1017/S0954579416000614

Maas, C. J. M., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology, 1,* 86–92. http://dx.doi.org/10.1027/1614-2241.1.3.86

Magee, L. (1990). R^2 measures based on Wald and likelihood ratio joint significance tests. *The American Statistician, 44,* 250–253.

McCoach, D., & Black, A. (2008). Assessing model adequacy. In A. O'Connell & D. McCoach (Eds.), *Multilevel modeling of educational data* (pp. 245–272). Charlotte, NC: Information Age.

McCrae, C. S., McNamara, J. P., Rowe, M. A., Dzierzewski, J. M., Dirk, J., Marsiske, M., & Craggs, J. G. (2008). Sleep and affect in older adults: Using multilevel modeling to examine daily associations. *Journal of Sleep Research, 17,* 42–53. http://dx.doi.org/10.1111/j.1365-2869.2008.00621.x

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,* 133–142. http://dx.doi.org/10.1111/j.2041-210x.2012.00261.x

Orelien, J. G., & Edwards, L. J. (2008). Fixed-effect variable selection in linear mixed models using R^2 statistics. *Computational Statistics & Data Analysis, 52,* 1896–1907. http://dx.doi.org/10.1016/j.csda.2007.06.006

Panter, A. T., & Sterba, S. K. (2011). *Handbook of ethics in quantitative methodology*. New York, NY: Routledge.

Payne, B. R., Lee, C. L., & Federmeier, K. D. (2015). Revisiting the incremental effects of context on word processing: Evidence from single-word event-related brain potentials. *Psychophysiology, 52,* 1456–1469. http://dx.doi.org/10.1111/psyp.12515

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

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

Recchia, A. (2010). R-squared measures for two-level hierarchical linear models using SAS. *Journal of Statistical Software, 32,* 1–9. http://dx.doi.org/10.18637/jss.v032.c02

Rights, J. D., & Sterba, S. K. (2016). The relationship between multilevel models and non-parametric multilevel mixture models: Discrete approximation of intraclass correlation, random coefficient distributions, and residual heteroscedasticity. *British Journal of Mathematical & Statistical Psychology, 69,* 316–343. http://dx.doi.org/10.1111/bmsp.12073

Rights, J. D., & Sterba, S. K. (2017). A framework of R-squared measures for single-level and multilevel regression mixture models. *Psychological Methods*. Advance online publication. http://dx.doi.org/10.1037/met0000139

Roberts, J. K., Monaco, J. P., Stovall, H., & Foster, V. (2011). Explained variance in multilevel models. In J. J. Hox & J. K. Roberts (Eds.), *Handbook of advanced multilevel analysis* (pp. 219–230). New York, NY: Routledge.

Sasidharan, S., Santhanam, R., Brass, D. J., & Sambamurthy, V. (2012). The effects of social network structure on enterprise systems success: A longitudinal multilevel analysis. *Information Systems Research, 23,* 658–678. http://dx.doi.org/10.1287/isre.1110.0388

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

Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London, UK: Sage.

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

Stolz, E. (2015). Cross-national variation in quality of life of care-dependent elders in Europe: A two-step approach combining multilevel regression and fuzzy-Set QCA. *International Journal of Sociology, 45,* 286–308. http://dx.doi.org/10.1080/00207659.2015.1098177

Vonesh, E. F., & Chinchilli, V. M. (1997). *Linear and nonlinear models for the analysis of repeated measurements*. New York, NY: Marcel Dekker.

Wang, J., & Schaalje, G. B. (2009). Model selection for linear mixed models using predictive criteria. *Communications in Statistics Simulation and Computation, 38,* 788–801. http://dx.doi.org/10.1080/03610910802645362

Wang, J., Xie, H., & Fisher, J. H. (2011). *Multilevel models: Applications using SAS*. Göttingen, Germany: Walter de Gruyter. http://dx.doi.org/10.1515/9783110267709

Wells, J. M., & Krieckhaus, J. (2006). Does national context influence democratic satisfaction? A multi-level analysis. *Political Research Quarterly, 59,* 569–578. http://dx.doi.org/10.1177/106591290605900406

Wilkinson, L., & The American Psychological Association Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanation. *American Psychologist, 54,* 594–604. http://dx.doi.org/10.1037/0003-066X.54.8.594

Xu, R. (2003). Measuring explained variation in linear mixed effects models. *Statistics in Medicine, 22,* 3527–3541. http://dx.doi.org/10.1002/sim.1572

Zheng, B. (2000). Summarizing the goodness of fit of generalized linear models for longitudinal data. *Statistics in Medicine, 19,* 1265–1275. http://dx.doi.org/10.1002/(SICI)1097-0258(20000530)19:10<1265::AID-SIM486>3.0.CO;2-U

(*Appendices follow*)

Appendix A

Model-Implied Outcome Variance Decomposition

Section A1: Model-Implied Outcome Variance With Cluster-Mean-Centered Level-1 Predictors

The data model for the two-level model with cluster-mean-centered level-1 predictors was given in manuscript Equation 5 and is restated here as Appendix Equation A1. Symbols used in this data model were defined in the manuscript.

$$\begin{gathered} y_{ij} = \mathbf{x}_{ij}^{w\prime}\boldsymbol{\gamma}^{w} + \mathbf{x}_{j}^{b\prime}\boldsymbol{\gamma}^{b} + \mathbf{w}_{ij}^{\prime}\mathbf{u}_{j} + e_{ij} \\ \mathbf{u}_{j} \sim MVN(\mathbf{0}, \mathbf{T}) \\ e_{ij} \sim N(0, \sigma^{2}) \end{gathered} \tag{A1}$$

Here we use Equation A1 to compute the model-implied total variance of y_{ij}:

$$\text{var}(y_{ij}) = \text{var}(\mathbf{x}_{ij}^{w\prime}\boldsymbol{\gamma}^{w} + \mathbf{x}_{j}^{b\prime}\boldsymbol{\gamma}^{b} + \mathbf{w}_{ij}^{\prime}\mathbf{u}_{j} + e_{ij}) \tag{A2}$$

Given independence of residuals and fixed effects as well as that of level-1 and level-2 predictors:

$$\text{var}(y_{ij}) = \text{var}(\mathbf{x}_{ij}^{w\prime}\boldsymbol{\gamma}^{w}) + \text{var}(\mathbf{x}_{j}^{b\prime}\boldsymbol{\gamma}^{b}) + \text{var}(\mathbf{w}_{ij}^{\prime}\mathbf{u}_{j}) + \text{var}(e_{ij}) \tag{A3}$$

The first term in Equation A3 is equal to

$$\text{var}(\mathbf{x}_{ij}^{w\prime}\boldsymbol{\gamma}^{w}) = \boldsymbol{\gamma}^{w\prime}\,\text{var}(\mathbf{x}_{ij}^{w\prime})\boldsymbol{\gamma}^{w} = \boldsymbol{\gamma}^{w\prime}\boldsymbol{\Phi}^{w}\boldsymbol{\gamma}^{w} \tag{A4}$$

where $\boldsymbol{\Phi}^{w}$ denotes the covariance matrix of all elements of $\mathbf{x}_{ij}^{w}$. The second term in Equation A3 is

$$\text{var}(\mathbf{x}_{j}^{b\prime}\boldsymbol{\gamma}^{b}) = \boldsymbol{\gamma}^{b\prime}\,\text{var}(\mathbf{x}_{j}^{b\prime})\boldsymbol{\gamma}^{b} = \boldsymbol{\gamma}^{b\prime}\boldsymbol{\Phi}^{b}\boldsymbol{\gamma}^{b} \tag{A5}$$

where $\boldsymbol{\Phi}^{b}$ denotes the covariance matrix of all elements of $\mathbf{x}_{j}^{b}$. The fourth term in Equation A3 is

$$\text{var}(e_{ij}) = \sigma^{2} \tag{A6}$$

The third term in Equation A3, using the law of total variance, is

$$\begin{aligned} \text{var}(\mathbf{w}_{ij}^{\prime}\mathbf{u}_{j}) &= E[\text{var}(\mathbf{w}_{ij}^{\prime}\mathbf{u}_{j}|\mathbf{u}_{j})] + \text{var}(E[\mathbf{w}_{ij}^{\prime}\mathbf{u}_{j}|\mathbf{u}_{j}]) \\ &= E[\mathbf{u}_{j}^{\prime}\boldsymbol{\Sigma}\mathbf{u}_{j}] + \text{var}(E[\mathbf{w}_{ij}^{\prime}]\mathbf{u}_{j}) \\ &= E[\mathbf{u}_{j}^{\prime}\boldsymbol{\Sigma}\mathbf{u}_{j}] + \text{var}(\mathbf{m}^{\prime}\mathbf{u}_{j}) \\ &= E[\mathbf{u}_{j}^{\prime}\boldsymbol{\Sigma}\mathbf{u}_{j}] + \mathbf{m}^{\prime}\text{var}(\mathbf{u}_{j})\mathbf{m} \\ &= E[\mathbf{u}_{j}^{\prime}\boldsymbol{\Sigma}\mathbf{u}_{j}] + \mathbf{m}^{\prime}\mathbf{T}\mathbf{m} \\ &= E[\mathbf{u}_{j}^{\prime}\boldsymbol{\Sigma}\mathbf{u}_{j}] + \tau_{00} \end{aligned} \tag{A7}$$

where $\boldsymbol{\Sigma}$ denotes the covariance matrix of all elements of $\mathbf{w}_{ij}$, $\mathbf{m}$ denotes a vector of means of all elements of $\mathbf{w}_{ij}$, and τ_{00} denotes the random intercept variance. Note that $\mathbf{m}^{\prime}\mathbf{T}\mathbf{m} = \tau_{00}$ because, when cluster-mean-centering all level-1 predictors, the mean of each level-1 predictor is equal to 0.

The first term in Equation A7 can be re-expressed as

$$\begin{aligned} E[\mathbf{u}_{j}\boldsymbol{\Sigma}\mathbf{u}_{j}] &= E[tr(\mathbf{u}_{j}^{\prime}\boldsymbol{\Sigma}\mathbf{u}_{j})] \\ &= E[tr(\mathbf{u}_{j}\mathbf{u}_{j}^{\prime}\boldsymbol{\Sigma})] \\ &= tr(E[\mathbf{u}_{j}\mathbf{u}_{j}^{\prime}]\boldsymbol{\Sigma}) \\ &= tr(\mathbf{T}\boldsymbol{\Sigma}) \end{aligned} \tag{A8}$$

Combining Equations A4–A6 and Equations A9 and A11 yields

$$\text{var}(y_{ij}) = \boldsymbol{\gamma}^{w\prime}\boldsymbol{\Phi}^{w}\boldsymbol{\gamma}^{w} + \boldsymbol{\gamma}^{b\prime}\boldsymbol{\Phi}^{b}\boldsymbol{\gamma}^{b} + tr(\mathbf{T}\boldsymbol{\Sigma}) + \tau_{00} + \sigma^{2} \tag{A9}$$

Section A2: Model-Implied Outcome Variance With Non-Cluster-Mean-Centered Level-1 Predictors

In the context of *non*-cluster-mean-centered level-1 variables (see *Extensions* section of manuscript Discussion), we can express the MLM as in manuscript Equation 4

$$\begin{gathered} y_{ij} = \mathbf{x}_{ij}^{\prime}\boldsymbol{\gamma} + \mathbf{w}_{ij}^{\prime}\mathbf{u}_{j} + e_{ij} \\ \mathbf{u}_{j} \sim MVN(\mathbf{0}, \mathbf{T}) \\ e_{ij} \sim N(0, \sigma^{2}) \end{gathered} \tag{A10}$$

where $\mathbf{x}_{ij}$ denotes a vector of 1 and all predictors and $\boldsymbol{\gamma}$ denotes a vector of all fixed effects. Other terms retain the definitions from Equation A1. Following the same procedure as for the model in Equation A1, the model-implied variance of y_{ij} for the non-cluster-mean-centered model is

$$\begin{aligned} \text{var}(y_{ij}) &= \text{var}(\mathbf{x}_{ij}^{\prime}\boldsymbol{\gamma} + \mathbf{w}_{ij}^{\prime}\mathbf{u}_{j} + e_{ij}) \\ &= \text{var}(\mathbf{x}_{ij}^{\prime}\boldsymbol{\gamma}) + \text{var}(\mathbf{w}_{ij}^{\prime}\mathbf{u}_{j}) + \text{var}(e_{ij}) \\ &= \boldsymbol{\gamma}^{\prime}\boldsymbol{\Phi}\boldsymbol{\gamma} + tr(\mathbf{T}\boldsymbol{\Sigma}) + \mathbf{m}^{\prime}\mathbf{T}\mathbf{m} + \sigma^{2} \end{aligned} \tag{A11}$$

where $\boldsymbol{\Phi}$ denotes the covariance matrix of $\mathbf{x}_{ij}$ (other terms retain the same definitions as above).

(*Appendices continue*)

Appendix B

Analytically Relating Previous Authors' MLM R^2's to Integrative Framework

Appendix B provides derivations showing how 5 of our R^2's reflect the same population quantities as those of 10 measures previously developed by Snijders and Bosker (1999, 2012), Raudenbush and Bryk (2002 [see also 1992 edition]), Vonesh and Chinchilli (1997), Xu (2003), Aguinis and Culpepper (2015), and Johnson (2014 [an extension of Nakagawa & Schielzeth, 2013]), as overviewed in manuscript Table 3. In some cases, certain of our measures are not only equivalent to pre-existing measures in the population, but are also equivalent in the sample (for Johnson's [2014] extension of Nakagawa & Schielzeth [2013] and Snijders & Bosker [2012]). In these cases, derivations involve showing that terms in pre-existing measures are equivalent to terms in our measures. In other cases, denoted by "*" in Appendix B, equivalencies between our measures and pre-existing measures hold in the population, but not necessarily in a given sample (for Aguinis & Culpepper [2015]; Raudenbush & Bryk [2002]; Vonesh & Chinchilli [1997]; and Xu [2003]). In these cases, derivations involve showing that replacing certain terms in the pre-existing measures with what the MLM implies for these terms (i.e., the population quantity based on MLM parameters) yield equivalencies to our measures.

All sections of Appendix B (sections B1-B10) show population equivalencies among pre-existing measures and measures from our framework for cluster-mean-centered MLMs (manuscript Table 1); these measures from our framework utilize the cluster-mean-centered model-implied variance expression from Appendix A Section A1 Equation A9. Additionally, Appendix B sections B1-B6 also show population equivalencies among pre-existing total measures and the *non*-cluster-mean-centered versions of our total measures (manuscript Table 5); these latter measures utilize the non-cluster-mean-centered model-implied variance expression from Appendix A Section A2 Equation A11. Note that the manuscript provides reasons why we recommend using the framework of measures in Table 1, together with fitting a MLM with cluster-mean-centered level-1 predictors, rather than using the supplemental formulas in Table 5, together with fitting a MLM without cluster-mean-centered level-1 predictors.

Section B1: Correspondence Between Vonesh and Chinchilli's (1997) Measure and $R_t^{2(fvm)}$

Vonesh and Chinchilli's (1997, p. 422 Eqn. 8.3.7) conditional R^2 with an implied fixed intercept null model, denoted $R^2_{c(FInull)}$, is given as

$$R^2_{c(FInull)} = 1 - \frac{\sum_{j=1}^{J}\sum_{i=1}^{n_j}(y_{ij}-\hat{y}_{ij})^2}{\sum_{j=1}^{J}\sum_{i=1}^{n_j}(y_{ij}-\bar{y})^2} \tag{B1}$$

Here J denotes the number of clusters, n_j denotes cluster size, $\bar{y}$ denotes the grand mean of y_{ij}, and $\hat{y}_{ij} = \mathbf{x}_{ij}^{w\prime}\boldsymbol{\gamma}^{w} + \mathbf{x}_{j}^{b\prime}\boldsymbol{\gamma}^{b} + \mathbf{w}_j\mathbf{u}_j$. Note that it is not necessary to actually fit a fixed intercept null model to get $\bar{y}$ so we do not refer to this measure as requiring a "two-model fitting approach" in the manuscript. With N denoting the total sample size (i.e., all observations), B1 can be re-expressed as

$$\begin{aligned} R^2_{c(FInull)} &= 1 - \frac{\dfrac{\sum_{j=1}^{J}\sum_{i=1}^{n_j}(y_{ij}-\hat{y}_{ij})^2}{N}}{\dfrac{\sum_{j=1}^{J}\sum_{i=1}^{n_j}(y_{ij}-\bar{y})^2}{N}} \\ &= \frac{\dfrac{\sum_{j=1}^{J}\sum_{i=1}^{n_j}(y_{ij}-\bar{y})^2}{N}}{\dfrac{\sum_{j=1}^{J}\sum_{i=1}^{n_j}(y_{ij}-\bar{y})^2}{N}} - \frac{\dfrac{\sum_{j=1}^{J}\sum_{i=1}^{n_j}(y_{ij}-\hat{y}_{ij})^2}{N}}{\dfrac{\sum_{j=1}^{J}\sum_{i=1}^{n_j}(y_{ij}-\bar{y})^2}{N}}. \\ &= \frac{\dfrac{\sum_{j=1}^{J}\sum_{i=1}^{n_j}(y_{ij}-\bar{y})^2}{N} - \dfrac{\sum_{j=1}^{J}\sum_{i=1}^{n_j}(y_{ij}-\hat{y}_{ij})^2}{N}}{\dfrac{\sum_{j=1}^{J}\sum_{i=1}^{n_j}(y_{ij}-\bar{y})^2}{N}}. \end{aligned} \tag{B2}$$

Here, the denominator estimates the total variance, and the numerator estimates the total variance minus the residual variance (i.e., the explained variance). Thus, $R^2_{c(FInull)}$ takes the general form of

$$R^2_{c(FInull)} = \frac{\text{var}(\hat{y}_{ij})}{\text{var}(y_{ij})}. \tag{B3}$$

Replacing these terms with model-implied variances from Equation A9 yields

(Appendices continue)

$$R^{2*}_{c(FInull)} = \frac{\text{var}\,(\mathbf{x}^{w\prime}_{ij}\,\boldsymbol{\gamma}^w + \mathbf{x}^{b\prime}_{j}\,\boldsymbol{\gamma}^b + \mathbf{w}'_{ij}\mathbf{u}_j)}{\text{var}\,(y_{ij})} = \frac{\boldsymbol{\gamma}^{w\prime}\,\boldsymbol{\Phi}^w\boldsymbol{\gamma}^w + \boldsymbol{\gamma}^{b\prime}\,\boldsymbol{\Phi}^b\boldsymbol{\gamma}^b + tr(\mathbf{T\Sigma}) + \tau_{00}}{\boldsymbol{\gamma}^{w\prime}\,\boldsymbol{\Phi}^w\boldsymbol{\gamma}^w + \boldsymbol{\gamma}^{b\prime}\,\boldsymbol{\Phi}^b\boldsymbol{\gamma}^b + tr(\mathbf{T\Sigma}) + \tau_{00} + \sigma^2} = R_t^{2(fvm)} \tag{B4}$$

When predictors are not cluster-mean-centered, we can instead use the expression in Equation A11 and replace Equation B4 with Equation B5:

$$R^{2*}_{c(FInull)} = \frac{\text{var}\,(\mathbf{x}'_{ij}\boldsymbol{\gamma} + \mathbf{w}'_{ij}\mathbf{u}_j)}{\text{var}\,(y_{ij})} = \frac{\boldsymbol{\gamma}'\,\boldsymbol{\Phi}\boldsymbol{\gamma} + tr(\mathbf{T\Sigma}) + \mathbf{m}'\,\mathbf{Tm}}{\boldsymbol{\gamma}'\,\boldsymbol{\Phi}\boldsymbol{\gamma} + tr(\mathbf{T\Sigma}) + \mathbf{m}'\,\mathbf{Tm} + \sigma^2} = R_t^{2(fvm)} \tag{B5}$$

Section B2: Correspondence Between Xu's (2003) Measure and $R_t^{2(fvm)}$

Xu's (2003, p. 3530 Eqn. 6) Ω_0^2 is given as

$$\Omega_0^2 = 1 - \frac{\sigma^2}{\text{var}\,(y_{ij})}. \tag{B6}$$

Replacing var(y_{ij}) with the model-implied variance of y_{ij} from Equation A9, this is

$$\begin{aligned}\Omega_0^{2*} &= 1 - \frac{\sigma^2}{\boldsymbol{\gamma}^{w\prime}\,\boldsymbol{\Phi}^w\boldsymbol{\gamma}^w + \boldsymbol{\gamma}^{b\prime}\,\boldsymbol{\Phi}^b\boldsymbol{\gamma}^b + tr(\mathbf{T\Sigma}) + \tau_{00} + \sigma^2}\\ &= \frac{\boldsymbol{\gamma}^{w\prime}\,\boldsymbol{\Phi}^w\boldsymbol{\gamma}^w + \boldsymbol{\gamma}^{b\prime}\,\boldsymbol{\Phi}^b\boldsymbol{\gamma}^b + tr(\mathbf{T\Sigma}) + \tau_{00} + \sigma^2}{\boldsymbol{\gamma}^{w\prime}\,\boldsymbol{\Phi}^w\boldsymbol{\gamma}^w + \boldsymbol{\gamma}^{b\prime}\,\boldsymbol{\Phi}^b\boldsymbol{\gamma}^b + tr(\mathbf{T\Sigma}) + \tau_{00} + \sigma^2} - \frac{\sigma^2}{\boldsymbol{\gamma}^{w\prime}\,\boldsymbol{\Phi}^w\boldsymbol{\gamma}^w + \boldsymbol{\gamma}^{b\prime}\,\boldsymbol{\Phi}^b\boldsymbol{\gamma}^b + tr(\mathbf{T\Sigma}) + \tau_{00} + \sigma^2}\\ &= \frac{\boldsymbol{\gamma}^{w\prime}\,\boldsymbol{\Phi}^w\boldsymbol{\gamma}^w + \boldsymbol{\gamma}^{b\prime}\,\boldsymbol{\Phi}^b\boldsymbol{\gamma}^b + tr(\mathbf{T\Sigma}) + \tau_{00}}{\boldsymbol{\gamma}^{w\prime}\,\boldsymbol{\Phi}^w\boldsymbol{\gamma}^w + \boldsymbol{\gamma}^{b\prime}\,\boldsymbol{\Phi}^b\boldsymbol{\gamma}^b + tr(\mathbf{T\Sigma}) + \tau_{00} + \sigma^2}\\ &= R_t^{2(fvm)}.\end{aligned} \tag{B7}$$

When predictors are not cluster-mean-centered, we can replace var(y_{ij}) with Equation A11 and replace Equation B7 with Equation B8:

$$\begin{aligned}\Omega_0^{2*} &= 1 - \frac{\sigma^2}{\boldsymbol{\gamma}'\,\boldsymbol{\Phi}\boldsymbol{\gamma} + tr(\mathbf{T\Sigma}) + \mathbf{m}'\,\mathbf{Tm} + \sigma^2}\\ &= \frac{\boldsymbol{\gamma}'\,\boldsymbol{\Phi}\boldsymbol{\gamma} + tr(\mathbf{T\Sigma}) + \mathbf{m}'\,\mathbf{Tm} + \sigma^2}{\boldsymbol{\gamma}'\,\boldsymbol{\Phi}\boldsymbol{\gamma} + tr(\mathbf{T\Sigma}) + \mathbf{m}'\,\mathbf{Tm} + \sigma^2} - \frac{\sigma^2}{\boldsymbol{\gamma}'\,\boldsymbol{\Phi}\boldsymbol{\gamma} + tr(\mathbf{T\Sigma}) + \mathbf{m}'\,\mathbf{Tm} + \sigma^2}\\ &= \frac{\boldsymbol{\gamma}'\,\boldsymbol{\Phi}\boldsymbol{\gamma} + tr(\mathbf{T\Sigma}) + \mathbf{m}'\,\mathbf{Tm}}{\boldsymbol{\gamma}'\,\boldsymbol{\Phi}\boldsymbol{\gamma} + tr(\mathbf{T\Sigma}) + \mathbf{m}'\,\mathbf{Tm} + \sigma^2}\\ &= R_t^{2(fvm)}\end{aligned} \tag{B8}$$

Section B3: Correspondence Between Johnson's (2014) Measure (Extending Nakagawa & Schielzeth [2013]) and $R_t^{2(fvm)}$

The linear mixed model (LMM) version of Johnson's (2014) measure is denoted $R^2_{LMM(c)}$ and is an extension of Nakagawa and Schielzeth's (2013) measure to allow random slopes. This measure (Johnson, 2014, p. 945 Eqn 2 after replacing $\sum_{l=1}^{u}\sigma_l^2$ with Johnson's Eqn. 10 and dropping σ_d^2, which is irrelevant for LMM) is given as

$$R^2_{LMM(c)} = \frac{\sigma_f^2 + \overline{\sigma_l^2}}{\sigma_f^2 + \overline{\sigma_l^2} + \sigma_\varepsilon^2} \tag{B9}$$

with σ_f^2 denoting the variance attributable to predictors via fixed effects, $\overline{\sigma_l^2}$ denoting the mean random effect variance across observations, and σ_ε^2 denoting the level-1 residual variance. Note that σ_f^2 is the same as our $\boldsymbol{\gamma}^{w\prime}\boldsymbol{\Phi}^w\boldsymbol{\gamma}^w + \boldsymbol{\gamma}^{b\prime}\boldsymbol{\Phi}^b\boldsymbol{\gamma}^b$ and σ_ε^2 is the same as our σ^2. We will show that $\overline{\sigma_l^2}$ is equivalent to $tr(\mathbf{T\Sigma}) + \tau_{00}$. $\overline{\sigma_l^2}$ is given by

$$\overline{\sigma_l^2} = tr(\mathbf{ZTZ}')/N \tag{B10}$$

with N denoting total sample size and $\mathbf{Z}$ denoting a $N \times (p+1)$ design matrix for random effects (i.e., a column of N 1's for the intercept, and a column for each predictor with a random slope) and $\mathbf{T}$ denoting the random effect covariance matrix.

$$\begin{aligned}tr(\mathbf{ZTZ}')/N &= tr(\mathbf{TZ}'\,\mathbf{Z})/N\\ &= N \times tr\left(\mathbf{T}\frac{\mathbf{Z}'\,\mathbf{Z}}{N}\right)/N\\ &= tr\left(\mathbf{T}\frac{\mathbf{Z}'\,\mathbf{Z}}{N}\right)\end{aligned} \tag{B11}$$

Note that $\frac{\mathbf{Z'Z}}{N}$ denotes a $(p+1) \times (p+1)$ matrix with the means of the squares of each element of $\mathbf{Z}$ across all N observations on the diagonal, and the means of pairwise products of each nonredundant element of $\mathbf{Z}$ across all N observations on the off-diagonals. We will call this matrix $\overline{\mathbf{Z}^2}$. Note also that, by definition of variance, $\boldsymbol{\Sigma}$ (the covariance matrix of elements of $\mathbf{w}_{ij}$) can be given as $\overline{\mathbf{Z}^2} - \overline{\mathbf{Z}}^2$, with $\overline{\mathbf{Z}}^2$ denoting a matrix with the squared means of each element of $\mathbf{Z}$ across all N observations on the diagonal and the pairwise products of means of each nonredundant element of $\mathbf{Z}$ across all N observations on the off-diagonals. Thus,

$$\begin{aligned}\overline{\sigma_l^2} &= tr\left(\mathbf{T}\overline{\mathbf{Z}^2}\right)\\ &= tr\left(\mathbf{T}\left(\boldsymbol{\Sigma} + \overline{\mathbf{Z}}^2\right)\right)\\ &= tr\left(\mathbf{T\Sigma} + \mathbf{T}\overline{\mathbf{Z}}^2\right)\\ &= tr(\mathbf{T\Sigma}) + tr\left(\mathbf{T}\overline{\mathbf{Z}}^2\right)\end{aligned} \tag{B12}$$

(*Appendices continue*)

With cluster-mean-centered level-1 predictors, $\overline{\mathbf{Z}}^2$ contains the intercept variance as the first element and all other elements are 0. Thus,

$$\begin{aligned}\overline{\sigma_l^2} &= tr(\mathbf{T\Sigma}) + tr(\mathbf{T}\overline{\mathbf{Z}}^2)\\ &= tr(\mathbf{T\Sigma}) + \tau_{00}\end{aligned} \tag{B13}$$

We can then express $R^2_{LMM(c)}$ as

$$\begin{aligned}R^2_{LMM(c)} &= \frac{\boldsymbol{\gamma}^{w\prime}\,\boldsymbol{\Phi}^w\boldsymbol{\gamma}^w + \boldsymbol{\gamma}^{b\prime}\,\boldsymbol{\Phi}^b\boldsymbol{\gamma}^b + tr(\mathbf{T\Sigma}) + \tau_{00}}{\boldsymbol{\gamma}^{w\prime}\,\boldsymbol{\Phi}^w\boldsymbol{\gamma}^w + \boldsymbol{\gamma}^{b\prime}\,\boldsymbol{\Phi}^b\boldsymbol{\gamma}^b + tr(\mathbf{T\Sigma}) + \tau_{00} + \sigma^2}\\ &= R_t^{2(fvm)}\end{aligned} \tag{B14}$$

When predictors are not cluster-mean-centered, note that σ_f^2 is the same as our $\boldsymbol{\gamma}'\boldsymbol{\Phi\gamma}$ and again σ_ε^2 is the same as our σ^2. Equation B13 can be re-expressed as

$$\begin{aligned}\overline{\sigma_l^2} &= tr(\mathbf{T\Sigma}) + tr(\mathbf{T}\overline{\mathbf{Z}}^2)\\ &= tr(\mathbf{T\Sigma}) + tr(\mathbf{Tmm}')\\ &= tr(\mathbf{T\Sigma}) + tr(\mathbf{m}'\,\mathbf{Tm})\\ &= tr(\mathbf{T\Sigma}) + \mathbf{m}'\,\mathbf{Tm}\end{aligned} \tag{B15}$$

Thus, when not cluster-mean-centering we can express $R^2_{LMM(c)}$ as

$$\begin{aligned}R^2_{LMM(c)} &= \frac{\boldsymbol{\gamma}'\,\boldsymbol{\Phi\gamma} + tr(\mathbf{T\Sigma}) + \mathbf{m}'\,\mathbf{Tm}}{\boldsymbol{\gamma}'\,\boldsymbol{\Phi\gamma} + tr(\mathbf{T\Sigma}) + \mathbf{m}'\,\mathbf{Tm} + \sigma^2}\\ &= R_t^{2(fvm)}\end{aligned} \tag{B16}$$

Section B4: Correspondence Between Snijders and Bosker (1999, 2012) Measure and $R_t^{2(f)}$

Snijders and Bosker's (1999, 2012) $R^2_{S\&B}$ (Snijders & Bosker, 2012, p. 117 with their Eqn. 7.9 as the denominator, and with the first term in their Eqn 7.9, $\boldsymbol{\gamma}_X'\boldsymbol{\Sigma}_X\boldsymbol{\gamma}_X$, as the numerator) is based on the model-implied decomposition of variance in random slope models and is given by:

$$R^2_{S\&B} = \frac{\boldsymbol{\gamma}_X'\boldsymbol{\Sigma}_X\boldsymbol{\gamma}_X}{\boldsymbol{\gamma}_X'\boldsymbol{\Sigma}_X\boldsymbol{\gamma}_X + \tau_{00} + 2\boldsymbol{\mu}_X'\boldsymbol{\tau}_{10} + \boldsymbol{\mu}_X'\mathbf{T}_{11}\boldsymbol{\mu}_X + tr(\mathbf{T}_{11}\boldsymbol{\Sigma}_X) + \sigma^2} \tag{B17}$$

with $\boldsymbol{\gamma}_X$ denoting a vector of fixed effects (slopes) for each predictor, $\boldsymbol{\Sigma}_X$ denoting a covariance matrix of all predictors, τ_{00} denoting the random intercept variance, $\boldsymbol{\mu}_X$ denoting a vector of means for each predictor, $\boldsymbol{\tau}_{10}$ denoting a vector of intercept-slope covariances associated with each predictor, $\mathbf{T}_{11}$ denoting a covariance matrix for the random slopes, and σ^2 denoting the level-1 residual variance. Note that this expression assumes all predictors have both fixed and random effects, but could be easily modified by having separate matrices/vectors for fixed and random components.

We show that our model-implied variance in Equation A9 is equivalent to the denominator expression in Equation B17. First, we will show that our $\boldsymbol{\gamma}^{w\prime}\boldsymbol{\Phi}^w\boldsymbol{\gamma}^w + \boldsymbol{\gamma}^{b\prime}\boldsymbol{\Phi}^b\boldsymbol{\gamma}^b$ is equal to Snijders and Bosker's $\boldsymbol{\gamma}_X'\boldsymbol{\Sigma}_X\boldsymbol{\gamma}_X$ term. Note that the combination of $\boldsymbol{\gamma}^b$ and $\boldsymbol{\gamma}^w$ in our expression contains all fixed parts of effects, whereas $\boldsymbol{\gamma}_X$ excludes the intercept. The combination of $\boldsymbol{\Phi}^w$ and $\boldsymbol{\Phi}^b$ in our expression and $\boldsymbol{\Sigma}_X$ both contain the variances and covariances of all predictors, but our $\boldsymbol{\Phi}^b$ also has a first row and column consisting entirely of 0's (because the first element of $\mathbf{x}_{ij}^b$ is a constant). We can thus set

$$\begin{aligned}\boldsymbol{\gamma}^{w\prime}\,\boldsymbol{\Phi}^w\boldsymbol{\gamma}^{w\prime} + \boldsymbol{\gamma}^{b\prime}\,\boldsymbol{\Phi}^b\boldsymbol{\gamma}^{b\prime} &= (\boldsymbol{\gamma}^b \quad \boldsymbol{\gamma}^w)'\begin{pmatrix}\boldsymbol{\Phi}^b & \mathbf{0}\\ \mathbf{0} & \boldsymbol{\Phi}^w\end{pmatrix}\begin{pmatrix}\boldsymbol{\gamma}^b\\ \boldsymbol{\gamma}^w\end{pmatrix}\\ &= (\gamma_{00} \quad \boldsymbol{\gamma}_X)'\begin{pmatrix}0 & \mathbf{0}_p'\\ \mathbf{0}_p & \boldsymbol{\Sigma}_X\end{pmatrix}\begin{pmatrix}\gamma_{00}\\ \boldsymbol{\gamma}_X\end{pmatrix}\\ &= (0 \quad \boldsymbol{\gamma}_X'\boldsymbol{\Sigma}_X)\begin{pmatrix}\gamma_{00}\\ \boldsymbol{\gamma}_X\end{pmatrix}\\ &= \boldsymbol{\gamma}_X'\boldsymbol{\Sigma}_X\boldsymbol{\gamma}_X\end{aligned} \tag{B18}$$

with $\mathbf{0}_p$ denoting a vector of p (number of predictors) 0's and γ_{00} the fixed part of the intercept.

Next, we will show that $tr(\mathbf{T\Sigma}) = tr(\mathbf{T}_{11}\boldsymbol{\Sigma}_X)$. Note that the first row and column of $\boldsymbol{\Sigma}$ consist solely of 0's (as the first element of $\mathbf{w}_{ij}$ is a constant) and the remaining rows and columns consist of variances and covariances of predictors with random slopes (i.e., $\boldsymbol{\Sigma}_X$). Thus,

$$\begin{aligned}tr(\mathbf{T\Sigma}) &= tr\left\{\begin{pmatrix}\tau_{00} & \boldsymbol{\tau}_{10}'\\ \boldsymbol{\tau}_{10} & \mathbf{T}_{11}\end{pmatrix}\begin{pmatrix}0 & \mathbf{0}_p'\\ \mathbf{0}_p & \boldsymbol{\Sigma}_X\end{pmatrix}\right\}\\ &= tr\left\{\begin{pmatrix}0 & \boldsymbol{\tau}_{10}'\boldsymbol{\Sigma}_X\\ 0 & \mathbf{T}_{11}\boldsymbol{\Sigma}_X\end{pmatrix}\right\}\\ &= tr\{\mathbf{T}_{11}\boldsymbol{\Sigma}_X\}\end{aligned} \tag{B19}$$

Next, note that when level-1 predictors with random slopes are cluster-mean-centered, the expression $\tau_{00} + 2\boldsymbol{\mu}_X'\boldsymbol{\tau}_{10} + \boldsymbol{\mu}_X'\mathbf{T}_{11}\boldsymbol{\mu}_X$ simplifies to τ_{00}, as the means of each such predictor equal 0. Thus,

$$\begin{aligned}R^2_{S\&B} &= \frac{\boldsymbol{\gamma}_X'\boldsymbol{\Sigma}_X\boldsymbol{\gamma}_X}{\boldsymbol{\gamma}_X'\boldsymbol{\Sigma}_X\boldsymbol{\gamma}_X + \tau_{00} + 2\boldsymbol{\mu}'\,\boldsymbol{\tau}_{10} + \boldsymbol{\mu}_X'\mathbf{T}_{11}\boldsymbol{\mu}_X + tr(\mathbf{T}_{11}\boldsymbol{\Sigma}_X) + \sigma^2}\\ &= \frac{\boldsymbol{\gamma}^{w\prime}\,\boldsymbol{\Phi}^w\boldsymbol{\gamma}^w + \boldsymbol{\gamma}^{b\prime}\,\boldsymbol{\Phi}^b\boldsymbol{\gamma}^b}{\boldsymbol{\gamma}^{w\prime}\,\boldsymbol{\Phi}^w\boldsymbol{\gamma}^w + \boldsymbol{\gamma}^{b\prime}\,\boldsymbol{\Phi}^b\boldsymbol{\gamma}^b + tr(\mathbf{T\Sigma}) + \tau_{00} + \sigma^2}\\ &= R_t^{2(f)}\end{aligned} \tag{B20}$$

(*Appendices continue*)

When level-1 predictors are not cluster-mean-centered, note first that our $tr(\mathbf{T\Sigma})$ is equal to $tr(\mathbf{T}_{11}\mathbf{\Sigma}_X)$, as shown above (nothing in the above derivation is specific to cluster-mean-centered contexts). Second, note that our $\boldsymbol{\gamma}'\mathbf{\Phi}\boldsymbol{\gamma}$ is equal to $\boldsymbol{\gamma}'_X\mathbf{\Sigma}_X\boldsymbol{\gamma}_X$.

$$\begin{aligned}\boldsymbol{\gamma}'\,\mathbf{\Phi}\boldsymbol{\gamma} &= (\gamma_{00} \quad \boldsymbol{\gamma}_X)'\begin{pmatrix}0 & \mathbf{0}'_p\\ \mathbf{0}_p & \mathbf{\Sigma}_X\end{pmatrix}\begin{pmatrix}\gamma_{00}\\ \boldsymbol{\gamma}_X\end{pmatrix}\\ &= (0 \quad \boldsymbol{\gamma}'_X\mathbf{\Sigma}_X)\begin{pmatrix}\gamma_{00}\\ \boldsymbol{\gamma}_X\end{pmatrix}\\ &= \boldsymbol{\gamma}'_X\mathbf{\Sigma}_X\boldsymbol{\gamma}_X\end{aligned} \tag{B21}$$

Next, note that $\mathbf{m}'\mathbf{Tm} = \tau_{00} + 2\boldsymbol{\mu}'_X\boldsymbol{\tau}_{10} + \boldsymbol{\mu}'_X\mathbf{T}_{11}\boldsymbol{\mu}_X$.

$$\begin{aligned}\mathbf{m}'\,\mathbf{Tm} &= tr(\mathbf{m}'\,\mathbf{Tm})\\ &= tr(\mathbf{Tmm}')\\ &= tr\left\{\begin{pmatrix}\tau_{00} & \boldsymbol{\tau}'_{10}\\ \boldsymbol{\tau}_{10} & \mathbf{T}_{11}\end{pmatrix}\begin{pmatrix}1\\ \boldsymbol{\mu}_X\end{pmatrix}(1 \quad \boldsymbol{\mu}'_X)\right\}\\ &= tr\left\{\begin{pmatrix}\tau_{00} + \boldsymbol{\tau}'_{10}\boldsymbol{\mu}_X\\ \boldsymbol{\tau}_{10} + \mathbf{T}_{11}\boldsymbol{\mu}_X\end{pmatrix}(1 \quad \boldsymbol{\mu}'_X)\right\}\\ &= tr\begin{Bmatrix}\tau_{00} + \boldsymbol{\tau}'_{10}\boldsymbol{\mu}_X & \tau_{00}\boldsymbol{\mu}'_X + \boldsymbol{\tau}'_{10}\boldsymbol{\mu}_X\boldsymbol{\mu}'_X\\ \boldsymbol{\tau}_{10} + \mathbf{T}_{11}\boldsymbol{\mu}_X & \boldsymbol{\tau}_{10}\boldsymbol{\mu}'_X + \mathbf{T}_{11}\boldsymbol{\mu}_X\boldsymbol{\mu}'_X\end{Bmatrix}\\ &= \tau_{00} + \boldsymbol{\tau}'_{10}\boldsymbol{\mu}_X + tr(\boldsymbol{\tau}_{10}\boldsymbol{\mu}'_X + \mathbf{T}_{11}\boldsymbol{\mu}_X\boldsymbol{\mu}'_X)\\ &= \tau_{00} + \boldsymbol{\tau}'_{10}\boldsymbol{\mu}_X + tr(\boldsymbol{\tau}_{10}\boldsymbol{\mu}'_X) + tr(\mathbf{T}_{11}\boldsymbol{\mu}_X\boldsymbol{\mu}'_X)\\ &= \tau_{00} + \boldsymbol{\tau}'_{10}\boldsymbol{\mu}_X + tr(\boldsymbol{\mu}'_X\boldsymbol{\tau}_{10}) + tr(\boldsymbol{\mu}'_X\mathbf{T}_{11}\boldsymbol{\mu}_X)\\ &= \tau_{00} + \boldsymbol{\tau}'_{10}\boldsymbol{\mu}_X + \boldsymbol{\mu}'_X\boldsymbol{\tau}_{10} + \boldsymbol{\mu}'_X\mathbf{T}_{11}\boldsymbol{\mu}_X\\ &= \tau_{00} + 2\boldsymbol{\mu}'_X\boldsymbol{\tau}_{10} + \boldsymbol{\mu}'_X\mathbf{T}_{11}\boldsymbol{\mu}_X\end{aligned} \tag{B22}$$

Thus, when not cluster-mean-centering level-1 predictors,

$$\begin{aligned}R^2_{S\&B} &= \frac{\boldsymbol{\gamma}'_X\mathbf{\Sigma}_X\boldsymbol{\gamma}_X}{\boldsymbol{\gamma}'_X\mathbf{\Sigma}_X\boldsymbol{\gamma}_X + \tau_{00} + 2\boldsymbol{\mu}'_X\boldsymbol{\tau}_{10} + \boldsymbol{\mu}'_X\mathbf{T}_{11}\boldsymbol{\mu}_X + tr(\mathbf{T}_{11}\mathbf{\Sigma}_X) + \sigma^2}\\ &= \frac{\boldsymbol{\gamma}'\,\mathbf{\Phi}\boldsymbol{\gamma}}{\boldsymbol{\gamma}'\,\mathbf{\Phi}\boldsymbol{\gamma} + tr(\mathbf{T\Sigma}) + \mathbf{m}'\,\mathbf{Tm} + \sigma^2}\\ &= R_t^{2(f)}.\end{aligned} \tag{B23}$$

Section B5: Correspondence Between Vonesh and Chinchilli's (1997) Measure and $R_t^{2(f)}$

Vonesh and Chinchilli's (1997) marginal R^2 (p. 422 Eqn. 8.3.7), denoted R^2_m, is given by

$$R^2_m = 1 - \frac{\sum_{j=1}^{J}\sum_{i=1}^{n_j}(y_{ij} - \hat{y}_{ij})^2}{\sum_{j=1}^{J}\sum_{i=1}^{n_j}(y_{ij} - \bar{y})^2} \tag{B24}$$

with $\hat{y}_{ij} = \mathbf{x}^{w\prime}_{ij}\boldsymbol{\gamma}^w + \mathbf{x}^{b\prime}_j\boldsymbol{\gamma}^b$. As shown in Appendix B Section B1, this can be reexpressed as

$$R^2_m = \frac{\mathrm{var}\,(\hat{y}_{ij})}{\mathrm{var}\,(y_{ij})}. \tag{B25}$$

Using our model-implied variance approach from Equation A9, this becomes

$$\begin{aligned}R^{2*}_m &= \frac{\mathrm{var}\,(\mathbf{x}^{w\prime}_{ij}\,\boldsymbol{\gamma}^w + \mathbf{x}^{b\prime}_j\,\boldsymbol{\gamma}^b)}{\mathrm{var}\,(y_{ij})}\\ &= \frac{\boldsymbol{\gamma}^{w\prime}\,\mathbf{\Phi}^w\boldsymbol{\gamma}^w + \boldsymbol{\gamma}^{b\prime}\,\mathbf{\Phi}^b\boldsymbol{\gamma}^b}{\boldsymbol{\gamma}^{w\prime}\,\mathbf{\Phi}^w\boldsymbol{\gamma}^w + \boldsymbol{\gamma}^{b\prime}\,\mathbf{\Phi}^b\boldsymbol{\gamma}^b + tr(\mathbf{T\Sigma}) + \tau_{00} + \sigma^2}\\ &= R_t^{2(f)}.\end{aligned} \tag{B26}$$

When level-1 predictors are not cluster-mean-centered, we use Equation A11 to yield

$$\begin{aligned}R^{2*}_m &= \frac{\mathrm{var}\,(\mathbf{x}'_{ij}\boldsymbol{\gamma})}{\mathrm{var}\,(y_{ij})}\\ &= \frac{\boldsymbol{\gamma}'\,\mathbf{\Phi}\boldsymbol{\gamma}}{\boldsymbol{\gamma}'\,\mathbf{\Phi}\boldsymbol{\gamma} + tr(\mathbf{T\Sigma}) + \mathbf{m}'\,\mathbf{Tm} + \sigma^2}\\ &= R_t^{2(f)}.\end{aligned} \tag{B27}$$

Section B6: Correspondence Between Johnson's (2014) Measure (Extending Nakagawa and Schielzeth [2013]) and $R_t^{2(f)}$

The linear mixed model (LMM) version of Johnson's (2014) measure, denoted $R^2_{LMM(m)}$ (Johnson, 2014, p. 945 Eqn 1, after replacing $\Sigma^u_{l=1}\sigma^2_l$ with Johnson's Eqn 10 and dropping σ^2_d, which is irrelevant for LMM), is an extension of Nakagawa and Schielzeth's (2013) measure to allow random slopes and is given as

$$R^2_{LMM(m)} = \frac{\sigma^2_f}{\sigma^2_f + \overline{\sigma^2_l} + \sigma^2_\varepsilon}. \tag{B28}$$

Replacing these terms with our model-implied notation that involves cluster-mean-centered predictors, as shown in Appendix B Section B3, this becomes

$$\begin{aligned}R^{2*}_{LMM(m)} &= \frac{\boldsymbol{\gamma}^{w\prime}\,\mathbf{\Phi}^w\boldsymbol{\gamma}^w + \boldsymbol{\gamma}^{b\prime}\,\mathbf{\Phi}^b\boldsymbol{\gamma}^b}{\boldsymbol{\gamma}^{w\prime}\,\mathbf{\Phi}^w\boldsymbol{\gamma}^w + \boldsymbol{\gamma}^{b\prime}\,\mathbf{\Phi}^b\boldsymbol{\gamma}^b + tr(\mathbf{T\Sigma}) + \tau_{00} + \sigma^2}\\ &= R_t^{2(f)}\end{aligned} \tag{B29}$$

When level-1 predictors are not cluster-mean-centered,

$$\begin{aligned}R^{2*}_{LMM(m)} &= \frac{\boldsymbol{\gamma}'\,\mathbf{\Phi}\boldsymbol{\gamma}}{\boldsymbol{\gamma}'\,\mathbf{\Phi}\boldsymbol{\gamma} + tr(\mathbf{T\Sigma}) + \mathbf{m}'\,\mathbf{Tm} + \sigma^2}\\ &= R_t^{2(f)}\end{aligned} \tag{B30}$$

(*Appendices continue*)

Section B7: Correspondence Between Raudenbush and Bryk's (2002) Measure and $R_w^{2(f_1v)}$

Raudenbush and Bryk's (2002, p. 79 Eqn. 4.20; also in Bryk & Raudenbush, 1992; Hox 2002, 2010; Kreft & de Leeuw, 1998) level-1 R^2 (which uses a two-model-fitting approach in which the null model is a random-intercept-only model) is denoted $R^2_{L1,RInull}$ and is given as

$$R^2_{L1,RInull} = \frac{\sigma^2_{RInull} - \sigma^2}{\sigma^2_{RInull}} \tag{B31}$$

with σ^2_{RInull} denoting the level-1 residual variance from the random-intercept-only null model. Note that the model-implied variance of y_{ij} from this null model can be denoted

$$\text{var}\,(y_{ij}) = \sigma^2_{RInull} + \tau^*_{00} \tag{B32}$$

with τ^*_{00} denoting the intercept variance from the random-intercept-only null model. Thus, we can set

$$R^{2*}_{L1,RInull} = \frac{(\text{var}\,(y_{ij}) - \tau^*_{00}) - \sigma^2}{(\text{var}\,(y_{ij}) - \tau^*_{00})} \tag{B33}$$

Using our model-implied variance that involves cluster-mean-centered predictors, this becomes

$$\begin{aligned} R^{2*}_{L1,RInull} &= \frac{(\boldsymbol{\gamma}^{w\prime}\,\boldsymbol{\Phi}^w\boldsymbol{\gamma}^w + \boldsymbol{\gamma}^{b\prime}\,\boldsymbol{\Phi}^b\boldsymbol{\gamma}^b + tr(\mathbf{T\Sigma}) + \tau_{00} + \sigma^2 - \tau^*_{00}) - \sigma^2}{\boldsymbol{\gamma}^{w\prime}\,\boldsymbol{\Phi}^w\boldsymbol{\gamma}^w + \boldsymbol{\gamma}^{b\prime}\,\boldsymbol{\Phi}^b\boldsymbol{\gamma}^b + tr(\mathbf{T\Sigma}) + \tau_{00} + \sigma^2 - \tau^*_{00}} \\ &= \frac{\boldsymbol{\gamma}^{w\prime}\,\boldsymbol{\Phi}^w\boldsymbol{\gamma}^w + \boldsymbol{\gamma}^{b\prime}\,\boldsymbol{\Phi}^b\boldsymbol{\gamma}^b + tr(\mathbf{T\Sigma}) + \tau_{00} - \tau^*_{00}}{\boldsymbol{\gamma}^{w\prime}\,\boldsymbol{\Phi}^w\boldsymbol{\gamma}^w + \boldsymbol{\gamma}^{b\prime}\,\boldsymbol{\Phi}^b\boldsymbol{\gamma}^b + tr(\mathbf{T\Sigma}) + \tau_{00} + \sigma^2 - \tau^*_{00}} \end{aligned} \tag{B34}$$

Note that τ^*_{00} reflects all between-cluster variance, and thus

$$\tau^*_{00} = \boldsymbol{\gamma}^{b\prime}\,\boldsymbol{\Phi}^b\boldsymbol{\gamma}^b + \tau_{00} \tag{B35}$$

Thus,

$$\begin{aligned} R^{2*}_{L1,RInull} &= \frac{\boldsymbol{\gamma}^{w\prime}\,\boldsymbol{\Phi}^w\boldsymbol{\gamma}^w + \boldsymbol{\gamma}^{b\prime}\,\boldsymbol{\Phi}^b\boldsymbol{\gamma}^b + tr(\mathbf{T\Sigma}) + \tau_{00} - (\boldsymbol{\gamma}^{b\prime}\,\boldsymbol{\Phi}^b\boldsymbol{\gamma}^b + \tau_{00})}{\boldsymbol{\gamma}^{w\prime}\,\boldsymbol{\Phi}^w\boldsymbol{\gamma}^w + \boldsymbol{\gamma}^{b\prime}\,\boldsymbol{\Phi}^b\boldsymbol{\gamma}^b + tr(\mathbf{T\Sigma}) + \tau_{00} + \sigma^2 - (\boldsymbol{\gamma}^{b\prime}\,\boldsymbol{\Phi}^b\boldsymbol{\gamma}^b + \tau_{00})} \\ &= \frac{\boldsymbol{\gamma}^{w\prime}\,\boldsymbol{\Phi}^w\boldsymbol{\gamma}^w + tr(\mathbf{T\Sigma})}{\boldsymbol{\gamma}^{w\prime}\,\boldsymbol{\Phi}^w\boldsymbol{\gamma}^w + tr(\mathbf{T\Sigma}) + \sigma^2} \\ &= R_w^{2(f_1v)} \end{aligned} \tag{B36}$$

Section B8: Correspondence Between Vonesh and Chinchilli's (1997) Measure and $R_w^{2(f_1v)}$

Vonesh and Chinchilli's (1997) conditional R^2 with a random-intercept-only null model involves a two-model-fitting approach and is denoted $R^2_{c(RInull)}$ (p. 422 Eqn 8.3.7 replacing $\bar{y}$ with $\hat{y}^{null}_{ij}$ from Eqn. 8.3.9). It is given by

$$R^2_{c(RInull)} = 1 - \frac{\sum_{j=1}^{J}\sum_{i=1}^{n_j}(y_{ij} - \hat{y}_{ij})^2}{\sum_{j=1}^{J}\sum_{i=1}^{n_j}(y_{ij} - \hat{y}^{null}_{ij})^2} \tag{B37}$$

The full model has $\hat{y}_{ij} = \mathbf{x}'_{ij}\boldsymbol{\gamma} + \mathbf{w}_j\mathbf{u}_j$ and null model has $\hat{y}^{null}_{ij} = \gamma^{null}_{00} + u^{null}_{0j}$. It can be re-expressed as

$$\begin{aligned} R^2_{c(RInull)} &= 1 - \frac{\dfrac{\sum_{j=1}^{J}\sum_{i=1}^{n_j}(y_{ij} - \hat{y}_{ij})^2}{N}}{\dfrac{\sum_{j=1}^{J}\sum_{i=1}^{n_j}(y_{ij} - \hat{y}^{null}_{ij})^2}{N}} \\ &= \frac{\dfrac{\sum_{j=1}^{J}\sum_{i=1}^{n_j}(y_{ij} - \hat{y}^{null}_{ij})^2}{N}}{\dfrac{\sum_{j=1}^{J}\sum_{i=1}^{n_j}(y_{ij} - \hat{y}^{null}_{ij})^2}{N}} - \frac{\dfrac{\sum_{j=1}^{J}\sum_{i=1}^{n_j}(y_{ij} - \hat{y}_{ij})^2}{N}}{\dfrac{\sum_{j=1}^{J}\sum_{i=1}^{n_j}(y_{ij} - \hat{y}^{null}_{ij})^2}{N}} \\ &= \frac{\dfrac{\sum_{j=1}^{J}\sum_{i=1}^{n_j}(y_{ij} - \hat{y}^{null}_{ij})^2}{N} - \dfrac{\sum_{j=1}^{J}\sum_{i=1}^{n_j}(y_{ij} - \hat{y}_{ij})^2}{N}}{\dfrac{\sum_{j=1}^{J}\sum_{i=1}^{n_j}(y_{ij} - \hat{y}^{null}_{ij})^2}{N}}. \end{aligned} \tag{B38}$$

Here, the denominator estimates the residual variance from the random-intercept-only null model, and the numerator estimates the residual variance from the random-intercept-only null model minus the residual variance from the full model. Thus, $R^2_{c(RInull)}$ takes the general form of

(*Appendices continue*)

$$R^{2*}_{c(RInull)} = \frac{(\text{var}\,(y_{ij}) - \text{var}\,(\hat{y}^{null}_{ij})) - (\text{var}\,(y_{ij}) - \text{var}\,(\hat{y}_{ij}))}{\text{var}\,(y_{ij}) - \text{var}\,(\hat{y}^{null}_{ij})} \tag{B39}$$

Using our model-implied variance approach that involves cluster-mean-centered predictors, this is

$$\begin{aligned} R^{2*}_{c(RInull)} &= \frac{(\text{var}\,(y_{ij}) - \tau^*_{00}) - (\text{var}\,(y_{ij}) - \text{var}\,(\mathbf{x}'_{ij}\boldsymbol{\gamma} + \mathbf{w}_j\mathbf{u}_j))}{\text{var}\,(y_{ij}) - \tau^*_{00}} \\ &= \frac{(\boldsymbol{\gamma}^{w\prime}\,\boldsymbol{\Phi}^w\boldsymbol{\gamma}^w + \boldsymbol{\gamma}^{b\prime}\,\boldsymbol{\Phi}^b\boldsymbol{\gamma}^b + tr(\mathbf{T}\boldsymbol{\Sigma}) + \tau_{00} + \sigma^2) - (\boldsymbol{\gamma}^{b\prime}\,\boldsymbol{\Phi}^b\boldsymbol{\gamma}^b + \tau_{00})}{(\boldsymbol{\gamma}^{w\prime}\,\boldsymbol{\Phi}^w\boldsymbol{\gamma}^w + \boldsymbol{\gamma}^{b\prime}\,\boldsymbol{\Phi}^b\boldsymbol{\gamma}^b + tr(\mathbf{T}\boldsymbol{\Sigma}) + \tau_{00} + \sigma^2) - (\boldsymbol{\gamma}^{b\prime}\,\boldsymbol{\Phi}^b\boldsymbol{\gamma}^b + \tau_{00})} \\ &\quad - \frac{(\boldsymbol{\gamma}^{w\prime}\,\boldsymbol{\Phi}^w\boldsymbol{\gamma}^w + \boldsymbol{\gamma}^{b\prime}\,\boldsymbol{\Phi}^b\boldsymbol{\gamma}^b + tr(\mathbf{T}\boldsymbol{\Sigma}) + \tau_{00} + \sigma^2) - (\boldsymbol{\gamma}^{w\prime}\,\boldsymbol{\Phi}^w\boldsymbol{\gamma}^w + \boldsymbol{\gamma}^{b\prime}\,\boldsymbol{\Phi}^b\boldsymbol{\gamma}^b + tr(\mathbf{T}\boldsymbol{\Sigma}) + \tau_{00})}{(\boldsymbol{\gamma}^{w\prime}\,\boldsymbol{\Phi}^w\boldsymbol{\gamma}^w + \boldsymbol{\gamma}^{b\prime}\,\boldsymbol{\Phi}^b\boldsymbol{\gamma}^b + tr(\mathbf{T}\boldsymbol{\Sigma}) + \tau_{00} + \sigma^2) - (\boldsymbol{\gamma}^{b\prime}\,\boldsymbol{\Phi}^b\boldsymbol{\gamma}^b + \tau_{00})} \\ &= \frac{\boldsymbol{\gamma}^{w\prime}\,\boldsymbol{\Phi}^w\boldsymbol{\gamma}^w + tr(\mathbf{T}\boldsymbol{\Sigma}) + \sigma^2 - \sigma^2}{\boldsymbol{\gamma}^{w\prime}\,\boldsymbol{\Phi}^w\boldsymbol{\gamma}^w + tr(\mathbf{T}\boldsymbol{\Sigma}) + \sigma^2} \\ &= \frac{\boldsymbol{\gamma}^{w\prime}\,\boldsymbol{\Phi}^w\boldsymbol{\gamma}^w + tr(\mathbf{T}\boldsymbol{\Sigma})}{\boldsymbol{\gamma}^{w\prime}\,\boldsymbol{\Phi}^w\boldsymbol{\gamma}^w + tr(\mathbf{T}\boldsymbol{\Sigma}) + \sigma^2} \\ &= R^{2(f_1v)}_w \end{aligned} \tag{B40}$$

Section B9: Correspondence Between Raudenbush and Bryk's (2002) Measure and $R^{2(f_2)}_b$

Raudenbush and Bryk's (2002, p. 74 Eqn 4.12; also in Bryk & Raudenbush, 1992; Hox 2002, 2010; Kreft & de Leeuw, 1998) proportion of variance explained in β_{0j} (the intercept) (which uses a two-model-fitting approach in which the null model is a random-intercept-only model) is denoted $R^2_{\beta_{0j}}$ and is given as

$$R^2_{\beta_{0j}} = \frac{\tau^*_{00} - \tau_{00}}{\tau^*_{00}} \tag{B41}$$

Using our model-implied variances involving cluster-mean-centered predictors, this is

$$\begin{aligned} R^{2*}_{\beta_{0j}} &= \frac{\boldsymbol{\gamma}^{b\prime}\,\boldsymbol{\Phi}^b\boldsymbol{\gamma}^b + \tau_{00} - \tau_{00}}{\boldsymbol{\gamma}^{b\prime}\,\boldsymbol{\Phi}^b\boldsymbol{\gamma}^b + \tau_{00}} \\ &= \frac{\boldsymbol{\gamma}^{b\prime}\,\boldsymbol{\Phi}^b\boldsymbol{\gamma}^b}{\boldsymbol{\gamma}^{b\prime}\,\boldsymbol{\Phi}^b\boldsymbol{\gamma}^b + \tau_{00}} \\ &= R^{2(f_2)}_b \end{aligned} \tag{B42}$$

Section B10: Correspondence Between Aguinis and Culpepper's (2015) Measure and $R^{2(v)}_t$

Aguinis and Culpepper's (2015) ICC beta, ρ_β (p. 945 Eqn 2), is given by

$$\rho_\beta = tr\left(\mathbf{T}\frac{\mathbf{X}'_c\,\mathbf{X}_c}{N-1}\right)S^{-2} \tag{B43}$$

with $\mathbf{X}_c$ denoting a $N \times (p + 1)$ design matrix with group-mean-centered predictors and S^2 denoting the total sample variance of y_{ij}. First note that the term $tr\left(\mathbf{T}\frac{\mathbf{X}'_c\mathbf{X}_c}{N-1}\right)$ is equivalent to our $tr(\mathbf{T}\boldsymbol{\Sigma})$ in the group-mean-centered version of our decomposition of total variance. By definition, $\frac{\mathbf{X}'_c\mathbf{X}_c}{N-1}$ is the sample estimate of the covariance matrix of all group-mean-centered predictors (including 1 for the intercept), i.e., the sample estimate of $\boldsymbol{\Sigma}$. Next, substituting the sample variance S^2 with our model-implied variance of y_{ij} involving group-mean-centered predictors, this yields

$$\begin{aligned} \rho^*_\beta &= \frac{tr\left(\mathbf{T}\frac{\mathbf{X}'_c\,\mathbf{X}_c}{N-1}\right)}{\boldsymbol{\gamma}^{w\prime}\,\boldsymbol{\Phi}^w\boldsymbol{\gamma}^w + \boldsymbol{\gamma}^{b\prime}\,\boldsymbol{\Phi}^b\boldsymbol{\gamma}^b + tr(\mathbf{T}\boldsymbol{\Sigma}) + \tau_{00} + \sigma^2} \\ &= \frac{tr(\mathbf{T}\boldsymbol{\Sigma})}{\boldsymbol{\gamma}^{w\prime}\,\boldsymbol{\Phi}^w\boldsymbol{\gamma}^w + \boldsymbol{\gamma}^{b\prime}\,\boldsymbol{\Phi}^b\boldsymbol{\gamma}^b + tr(\mathbf{T}\boldsymbol{\Sigma}) + \tau_{00} + \sigma^2} \\ &= R^{2(v)}_t \end{aligned} \tag{B44}$$

Received May 20, 2017
Revision received January 18, 2018
Accepted February 2, 2018 ■