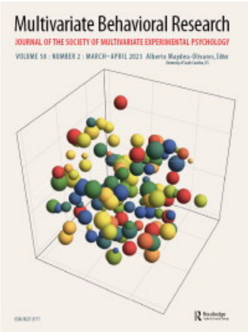


Multivariate Behavioral Research

ISSN: 0027-3171 (Print) 1532-7906 (Online) Journal homepage: www.tandfonline.com/journals/hmbr20

R-squared Measures for Multilevel Models with Three or More Levels

Jason D. Rights & Sonya K. Sterba

To cite this article: Jason D. Rights & Sonya K. Sterba (2023) R-squared Measures for Multilevel Models with Three or More Levels, Multivariate Behavioral Research, 58:2, 340-367, DOI: 10.1080/00273171.2021.1985948

To link to this article: https://doi.org/10.1080/00273171.2021.1985948

MULTIVARIATE BEHAVIORAL RESEARCH
2023, VOL. 58, NO. 2, 340–367
https://doi.org/10.1080/00273171.2021.1985948

R-squared Measures for Multilevel Models with Three or More Levels

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

[a]Department of Psychology, University of British Columbia; [b]Dept. of Psychology & Human Development; Vanderbilt University

ABSTRACT
Applications of multilevel models (MLMs) with three or more levels have increased alongside expanding software capability and dataset availability. Though researchers often express interest in R-squared measures as effect sizes for MLMs, R-squareds previously proposed for MLMs with three or more levels cover a limited subset of choices for how to quantify explained variance in these models. Additionally, analytic relationships between total and level-specific versions of MLM R-squared measures have not been clarified, despite such relationships becoming increasingly important to understand when there are more levels. Furthermore, the impact of predictor centering strategy on R-squared computation and interpretation has not been explicated for MLMs with any number of levels. To fill these gaps, we extend the Rights and Sterba two-level MLM R-squared framework to three or more levels, providing a general set of measures that includes preexisting three-level measures as special cases and yields additional results not obtainable from existing measures. We mathematically and pedagogically relate total and level-specific R-squareds, and show how all total and level-specific R-squared measures in our framework can be computed under any centering strategy. Finally, we provide and empirically demonstrate software (available in the *r2mlm* R package) to compute measures and graphically depict results.

KEYWORDS
Multilevel modeling; effect size; mixed effects modeling; R-squared; hierarchical linear modeling

Social science researchers are increasingly fitting multilevel models (MLMs) with three or more levels as software capability and dataset availability expand (e.g., Chen, Zhu, & Zhou, 2015; Curran, McGinley, Serrano, & Burfeind, 2012; Dollard et al., 2012; Gong, Kim, Lee, & Zhu, 2013; Liu, Liao, & Loi, 2012; Maier, Vitiello, & Greenfield, 2012; Van den Noortgate et al., 2013). An example three-level data structure involves students (at level-1) nested within classrooms (at level-2) nested within schools (at level-3). Though researchers often want to report R-squared (R^2) measures as effect sizes for MLMs (e.g., Bickel, 2007; Edwards et al., 2008; Jaeger et al., 2017; Johnson, 2014; Kramer, 2005; LaHuis et al., 2014; Lorah, 2018; Nakagawa & Schielzeth, 2013; Orelien & Edwards, 2008; Recchia, 2010; Roberts et al., 2011; Wang & Schaalje, 2009; Xu, 2003; Zheng, 2000), MLM applications with three or more levels rarely do so. In contrast, applications of two-level models commonly report R^2 (LaHuis et al., 2014).

Unfortunately, R^2 measures previously proposed for MLMs with three or more levels cover a limited subset of the choices for how to quantify explained variance in these models. This may be a contributing factor to their underutilization in practice. Another contributing factor may be that precise analytic relationships between total and level-specific R^2 measures—which become increasingly important to understand when there are more levels, and hence more level-specific measures—have gone unaddressed. To fill these gaps, we extend the two-level MLM R^2 framework of Rights and Sterba (2019) to three or more levels, providing a comprehensive set of measures that not only includes preexisting three-level measures as special cases but also yields additional substantively meaningful results that cannot be obtained using existing measures. Further, we newly delineate mathematical relationships between total and level-specific R^2's and explain why it is important to understand such relationships in applied practice. Additionally, we derive and explicate the impact of centering strategy for predictors (e.g., cluster-mean-centering versus centering-by-a-constant) on R^2 computation and interpretation for MLMs with any number of levels, and we show how to quantify total and level-specific variance explained under any centering

CONTACT Jason Rights jrights@psych.ubc.ca Department of Psychology, University of British Columbia, 2136 West Mall, Vancouver, BC, Canada V6T1Z4

Supplemental data for this article is available online at at https://doi:10.1080/00273171.2021.1985948.

strategy for MLMs with any number of levels. We also explain how (any-level) MLM R^2 results from our framework can be computed and graphically visualized in a simultaneous and integrated fashion, thereby avoiding the previous concern that "the use of these [R^2] techniques, however, can become confusing … as the models become more complex" (Raudenbush & Bryk, 2002, p. 149). We provide software (now also available in the R package *r2mlm*; Shaw, Rights, Sterba, & Flake, 2020) to compute measures in our framework and produce graphical depictions of results, and we illustrate its utility with an empirical example. We conclude with guidance for how to use R-squared differences (ΔR^2) to compute the unique contribution of individual terms in such models. Our methodological developments support researchers' efforts to convey practical significance by facilitating their understanding of, and ability to report, R^2s in MLMs with clustering beyond two levels.

We begin by reviewing the utility of reporting R^2 measures in MLMs. Second, we review the three-level MLM. Third, we discuss existing R^2 measures that have been proposed for three (or more) level models, describe their limitations, and explain how our framework overcomes these limitations. Next we present our framework of R^2 measures for three-level MLMs. Subsequently, we describe relationships among total and level-specific R^2 measures and explain their substantive implications. We then empirically illustrate the use of our framework of R^2s in a three-level application studying math achievement, provide general recommendations for practice, and describe software implementation for three-level MLMs. Subsequently, we provide and discuss the extension of our framework to MLMs with any number of levels.

Throughout the majority of the article, we follow widespread methodological recommendations to center all *lower-level predictors* (i.e., all predictors below the highest level of nesting) using cluster-mean-centering so that they contain variance at only one level (e.g., Algina & Swaminathan, 2011; Curran et al. 2012; Curran & Bauer, 2011; Enders & Tofighi, 2007; Hofmann & Gavin, 1998; Preacher, Zyphur, & Zhang, 2010; Raudenbush & Bryk, 2002; Rights, Preacher, & Cole, 2020; Snijders & Bosker, 2012), which prevents conflation of level-specific effects. However, in a later section we show how all total and level-specific R^2 measures in our framework can nonetheless be computed for MLMs that do not use cluster-mean-clustering for some or all lower-level predictors, and we discuss the impact of centering choice on R^2 computation and interpretation.

Utility of reporting R^2 for three-level models

In general, R^2s are useful indications of effect size in that they assess the proportion of outcome variance explained (e.g., Cohen et al., 2003; Gelman & Pardoe, 2006) on an interpretable metric with meaningful bounds of 0 and 1. The complication in three-level models, over that of single- or two-level models, is that variance can be explained by sources at level-1, level-2, *and* level-3. This makes the use of R^2 measures—in particular, those that distinguish among alternative sources of explained variance—particularly important to consider for three-level contexts.

To appreciate the utility of R^2 measures for three-level models, consider the following. In an analysis in which students are nested within classrooms within schools, there is outcome variability across students within classrooms (i.e., variance at level-1), across classrooms within schools (i.e., variance at level-2), and across schools (i.e., variance at level-3). To understand the mechanisms through which the outcome can be explained, one must then consider the extent to which each of these different levels contributes. This type of question can be difficult to assess even in two-level contexts, but becomes even more burdensome as the number of levels/potential sources of variability increase, particularly if one is examining only regression coefficients, random effect (co)variances, and associated p-values (as is common in practice; Roberts et al., 2011; Xu, 2003). R^2 measures for three-level contexts help alleviate this burden by allowing researchers to quantify, on an interpretable proportion metric, the amount of variance explained by sources *at each level* of the hierarchical structure. If, for instance, it turns out that a non-negligible proportion of the total outcome variance is explained only by student-level characteristics, then a researcher can better understand that the outcome is not predicted by classroom-level or school-level characteristics included in the model and, instead, the student-level characteristics being modeled are likely most important to consider.

Three-level multilevel model

The three-level MLM for the outcome y_{ijk}, for level-1 unit i, level-2 unit j, and level-3 unit k, can be represented as:

$$y_{ijk} = \gamma_{000} + \mathbf{x}'_{1ijk}\boldsymbol{\gamma}_1 + \mathbf{x}'_{2jk}\boldsymbol{\gamma}_2 + \mathbf{x}'_{3k}\boldsymbol{\gamma}_3 + \mathbf{w}'_{1*2ijk}\mathbf{u}_{1*2jk} + \mathbf{w}'_{1*3ijk}\mathbf{q}_{1*3k} + \mathbf{z}'_{2*3jk}\mathbf{q}_{2*3k} + e_{ijk}$$

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

$$\begin{bmatrix} \mathbf{u}_{1*2jk} \\ \mathbf{q}_{1*3k} \\ \mathbf{q}_{2*3k} \end{bmatrix} \sim MVN\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathrm{T}_{1*2} & & \\ \mathbf{0} & \mathrm{T}_{1*3} & \\ \mathbf{0} & \mathrm{T}_{12*3} & \mathrm{T}_{2*3} \end{bmatrix} \right) \tag{1}$$

The first four terms denote the fixed portion of the model, with the first term (γ_{000}) representing the fixed component of the intercept, and the next three terms ($\mathbf{x}'_{1ijk}\boldsymbol{\gamma}_1$ and $\mathbf{x}'_{2jk}\boldsymbol{\gamma}_2$ and $\mathbf{x}'_{3k}\boldsymbol{\gamma}_3$) representing fixed components of slopes at each of the three levels. Specifically, $\mathbf{x}_{1ijk}$ denotes the vector of level-1 predictors with $\boldsymbol{\gamma}_1$ denoting the vector of the level-1 fixed components of slopes; $\mathbf{x}_{2jk}$ denotes the vector of the level-2 predictors with $\boldsymbol{\gamma}_2$ denoting the vector of level-2 fixed components of slopes; and $\mathbf{x}_{3k}$ denotes the vector of level-3 predictors with $\boldsymbol{\gamma}_3$ denoting the vector of level-3 fixed components of slopes.[1] Note that $\mathbf{x}_{1ijk}$ can contain cross-level interaction terms (i.e., product terms) between level-1 and level-2 or -3 variables (as these product terms will vary only at level-1; Rights & Sterba, 2019), whereas $\mathbf{x}_{2jk}$ can contain cross-level interaction terms between level-2 and level-3 variables (as these product terms will vary only at level-2).

The next term ($\mathbf{w}'_{1*2ijk}\mathbf{u}_{1*2jk}$) represents the random portion across level-2 units. Specifically, $\mathbf{w}_{1*2ijk}$ denotes a vector whose first element is 1 (for the level-2 random component of the intercept) and whose subsequent elements are level-1 predictors with randomly varying slopes across level-2 units; $\mathbf{u}_{1*2jk}$ denotes a vector of the corresponding level-2 random effect residuals (note that the "*" in the subscripts denotes "across," and hence the "1*2" subscript in $\mathbf{w}_{1*2ijk}$ means it contains *level-1* variables with slopes varying *across level-2* units). The next two terms ($\mathbf{w}'_{1*3ijk}\mathbf{q}_{1*3k}$ and $\mathbf{z}'_{2*3jk}\mathbf{q}_{2*3k}$) represent the random portion across level-3 units. Specifically, $\mathbf{w}_{1*3ijk}$ denotes a vector of 1 (for the level-3 random component of the intercept) and level-1 predictors with randomly varying slopes across level-3 units (where again the "*" subscript notation denotes "across"), $\mathbf{q}_{1*3k}$ denotes a vector of the corresponding level-3 random effect residuals (with the first element being the level-3 random component of the intercept), $\mathbf{z}_{2*3jk}$ denotes a vector of level-2 predictors with randomly varying slopes across level-3 units, and $\mathbf{q}_{2*3k}$ denotes a vector of the corresponding level-3 random effect residuals. Note that a given level-1 predictor can be in both $\mathbf{w}_{1*2ijk}$ and $\mathbf{w}_{1*3ijk}$ if it has a random slope varying across both level-2 and level-3 units. All random effect residuals in $\mathbf{u}_{1*2jk}$, $\mathbf{q}_{1*3k}$, and $\mathbf{q}_{2*3k}$ are assumed multivariate normally distributed, with corresponding covariance matrices T_{1*2}, T_{1*3}, and T_{2*3}, respectively, with T_{12*3} containing covariances between terms in $\mathbf{q}_{1*3k}$ and $\mathbf{q}_{2*3k}$; because residuals across levels are orthogonal, all across-level correlations are 0. The last term in the model (e_{ijk}) denotes the level-1 residual, which is normally distributed with variance σ^2.

To avoid conflating level-specific effects in the Equation (1) model, each level-1 variable is decomposed into purely level-1, -2, and -3 components via cluster-mean-centering (often called group-mean-centering) and each level-2 variable is decomposed into purely level-2 and -3 components via cluster-mean-centering (e.g., Brincks et al., 2017). More specifically, for a level-1 variable, x_{ijk}, the level-3 component (given as $x_{\cdot\cdot k}$) is computed[2] as the level-3 cluster mean of x_{ijk}; the level-2 component (given as $x_{\cdot jk} - x_{\cdot\cdot k}$) is computed by subtracting $x_{\cdot\cdot k}$ from x_{ijk}'s level-2 cluster mean (given as $x_{\cdot jk}$); and the level-1 component is $x_{ijk} - x_{\cdot jk}$. For a level-2 variable w_{jk}, the level-3 component (given as $w_{\cdot k}$) is computed as the level-3 cluster mean of w_{jk} and the level-2 component is $w_{jk} - w_{\cdot k}$ (e.g., Brincks et al., 2017). Note that any level-3 predictor can be grand-mean-centered or left uncentered. If there is interest in potential between-cluster effects of level-1 predictors for a particular application, the level-2 and/or level-3 means of the level-1 variables can be included as separate predictors; however, there is no requirement that these be added. Similarly, one can include a level-3 cluster-mean of a level-2 variable as a separate predictor if interest lies in between-level-3-cluster effects of the level-2 variable. For specific guidance on interpreting between- and within-cluster effects of categorial predictors (with two or more categories), see Yaremych, Preacher, and Hedeker (2020; submitted).

Existing three-level R^2 measures

In this section we define the R^2s that have been developed for three-level contexts and describe their limitations. In the subsequent section, we summarize how our framework overcomes these limitations. Note that certain preexisting R^2s quantify variance explained via both fixed and random effects (commonly termed *conditional measures*), whereas others quantify variance explained via only fixed effects (commonly termed *marginal measures*; Nakagawa & Schielzeth,

[1]Letting P_x, P_w, and P_z denote the number of level-1, level-2, and level-3 predictors, note that $\mathbf{x}_{1ijk}$ is $P_x \times 1$, $\boldsymbol{\gamma}_1$ is $P_x \times 1$, $\mathbf{x}_{2jk}$ is $P_w \times 1$, $\boldsymbol{\gamma}_2$ is $P_w \times 1$, $\mathbf{x}_{3k}$ is $P_z \times 1$, and $\boldsymbol{\gamma}_3$ is $P_z \times 1$.

[2]Here we are assuming that the level-3 cluster-mean is computed as the mean value of the level-1 observations of x_{ijk} within level-3 cluster k, which we denote $x_{\cdot\cdot k}$.

2013; Vonesh & Chinchilli, 1997). Additionally, some preexisting R^2s have been previously defined as the reduction in residual variances (going from a null model to the full model) whereas others have been defined as the squared correlation between observed and predicted outcomes. Nonetheless we show that each of these kinds of measures are special cases of our general framework and can be understood as estimating some underlying proportion of variance explained.

Snijders and Bosker's (2012) measure

A *total* R^2—quantifying the proportion of total outcome variance explained—was developed for three-level MLMs by Snijders and Bosker (1999, p. 104; 2012, p. 113):

$$R^2_{S\&B} = 1 - \frac{\varphi_{000} + \tau_{000} + \sigma^2}{\varphi_{000(null)} + \tau_{000(null)} + \sigma^2_{(null)}} \quad (2)$$

Here φ_{000} denotes level-3 intercept variance, τ_{000} denotes level-2 intercept variance, and σ^2 denotes the level-1 residual variance from the fitted model. The $\varphi_{000(null)}$, $\tau_{000(null)}$, and $\sigma^2_{(null)}$ denote variance components from a random-intercept-only null model, i.e., a model with no predictors and an intercept that varies randomly across level-2 and level-3 units. Consequently, $\varphi_{000(null)}$ represents overall outcome variance at level-3, $\tau_{000(null)}$ overall outcome variance at level-2, and $\sigma^2_{(null)}$ overall outcome variance at level-1. $R^2_{S\&B}$ is a *total* R^2 measure in that the sum of the latter three terms (in the denominator) represents the total outcome variance.[3]

Note that this $R^2_{S\&B}$ measure requires constraining the random slope variances to 0. Though it is possible that a researcher may not be interested in slope variance specifically, methodologists have noted that it is preferable to have the measure reflect the structure of the full model of interest (e.g., Gurka et al., 2011; Jaeger et al., 2017; Johnson, 2014). Additionally, $R^2_{S\&B}$ involves a two-model fitting approach that can yield negative values (e.g., if due to chance fluctuation in estimates, the sum of the null model variance component estimates is smaller than the sum of the fitted model variance component estimates; Snijders & Bosker, 1994). Further, $R^2_{S\&B}$ assesses only variance explained by predictors via fixed effects; this kind of measure is often termed a *marginal* total R^2 (in essence, predicted scores are "marginalized" across random effects). Thus, it is not possible to use $R^2_{S\&B}$ to assess the degree to which the outcome can be explained by any other source (e.g., via random effects, as is the case with other measures discussed shortly). Finally, $R^2_{S\&B}$ provides an overall summary of variance explained marginally by *all* predictors; hence, it tells a researcher nothing about the relative importance of predictors at different levels (e.g., student- vs. classroom- vs. school-level predictors).

Nakagawa and Schielzeth's (2013) measures with Johnson's (2014) extension

Two other *total* R^2 measures were developed for three-level MLMs by Nakagawa and Schielzeth (2013, p. 137). They also required constraining random slope variances to 0; however, they were both recently extended to allow for random slope variation by Johnson (2014, p. 945). For generality we discuss Johnson's (2014) latter measures only.

The first measure, $R^2_{NSJ(m)}$, is a *marginal* total R^2, meaning that it quantifies the proportion of total variance explained by the predictors via fixed effects only, like $R^2_{S\&B}$:

$$R^2_{NSJ(m)} = \frac{\sigma^2_f}{\sigma^2_f + \overline{\sigma^2_{L2}} + \overline{\sigma^2_{L3}} + \sigma^2} \quad (3)$$

Here, σ^2_f denotes the variance attributable to predictors via fixed components of slopes. $\overline{\sigma^2_{L2}}$ and $\overline{\sigma^2_{L3}}$ denote the mean random effect variance across observations for level-2 and level-3 random effects, respectively.[4] The second measure, $R^2_{NSJ(c)}$, is termed a *conditional* total R^2 in that it quantifies the proportion of total variance explained via both fixed *and* random effects (in essence, predicted scores are "conditioned" on random effects) and is given as:

$$R^2_{NSJ(c)} = \frac{\sigma^2_f + \overline{\sigma^2_{L2}} + \overline{\sigma^2_{L3}}}{\sigma^2_f + \overline{\sigma^2_{L2}} + \overline{\sigma^2_{L3}} + \sigma^2} \quad (4)$$

Note that, similar to $R^2_{S\&B}$, neither $R^2_{NSJ(m)}$ nor $R^2_{NSJ(c)}$ can distinguish among variance explained by

[3]Though here we define $R^2_{S\&B}$ as a "total" measure, Snijders and Bosker (2012, p. 113) call this a "level-1 measure" in the sense that it quantifies variance explained across all observations (i.e., across all level-1 units). However, variance across all observations includes not just purely level-1 variance (i.e., purely within-cluster variance) but also includes level-2 and level-3 variance (i.e., across-cluster variance). In other words, the denominator of $R^2_{S\&B}$ contains the *total* outcome variance. As such, we feel it is most appropriate to refer to $R^2_{S\&B}$ as a total measure (quantifying the proportion of total outcome variance explained), and use the term "level-1 measure" to refer to measures with only level-1 outcome variance in the denominator. Our definition of total and level-specific measures is consistent with the MLM R^2 literature (see e.g., Lahuis et al., 2014).

[4]For instance, with a random intercept and one level-1 predictor with a random slope across level-2 and level-3 units, the observation-specific across-level-2 random effect variance can be written as $\sigma^2_{L2ijk} = \text{var}(u_{0jk} + u_{1jk}x_{ijk})$ and the observation-specific across-level-3 random effect variance as $\sigma^2_{L3ijk} = \text{var}(u_{00k} + u_{1k}x_{ijk})$. The quantities $\overline{\sigma^2_{L2}}$ and $\overline{\sigma^2_{L3}}$ are the expected values of these observation-specific random effect variances (see Johnson, 2014, for further detail).

level-1, -2, and -3 predictors. Second, though it is possible to assess the degree to which any random effect variability is important in understanding the outcome (i.e., by comparing $R^2_{NSJ(m)}$ and $R^2_{NSJ(c)}$), the $R^2_{NSJ(c)}$ measure provides no way of further distinguishing between contributions of predictors via random *slope* variation vs. contributions of cluster means via random *intercept* variation. Hence, these measures do not help researchers deeming random slope variation of substantive interest but random intercept variation not of substantive interest.

Raudenbush and Bryk's (2002) measures

Another set of three-level MLM R^2 measures, developed by Raudenbush and Bryk (2002, p. 79, Eqn. 4.20 and p. 85, Eqn. 4.24) (see also Bryk & Raudenbush, 1992), are *level-specific*, rather than total, measures (meaning that they quantify the proportion of level-specific outcome variance that is explained). These are commonly called *proportion reduction in residual variance* measures.[5] For level-1, this is given as

$$R^2_{R\&B,1} = \frac{\sigma^2_{(null)} - \sigma^2}{\sigma^2_{(null)}} \tag{5}$$

and for level-2:

$$R^2_{R\&B,2} = \frac{\tau_{000(null)} - \tau_{000}}{\tau_{000(null)}} \tag{6}$$

and for level-3:

$$R^2_{R\&B,3} = \frac{\varphi_{000(null)} - \varphi_{000}}{\varphi_{000(null)}} \tag{7}$$

Symbols in Equations (5)–(7) have the same definitions as in Equation (2). Though not apparent from Equations (5)–(7), it will be shown later analytically that actually $R^2_{R\&B,1}$ is a *conditional* R^2 at level-1, whereas $R^2_{R\&B,3}$ is a *marginal* R^2 at level-3, and $R^2_{R\&B,2}$ is a *hybrid* of a conditional and marginal measure at level-2.

Note that, going from a random-intercept-only null model to the full fitted model of interest, the residual variance at level-1 or level-2 can decrease *both* due to predictors via fixed components of slopes and due to predictors via random slope variation. Raudenbush and Bryk's (2002) level-1 and level-2 measures ($R^2_{R\&B,1}$ and $R^2_{R\&B,2}$) both combine these two potential sources of explained variance with no way of decomposing them, making these measures uninterpretable blends of sources, similar to $R^2_{NSJ(c)}$ discussed above. Second, similar to $R^2_{S\&B}$, the $R^2_{R\&B,1}$, $R^2_{R\&B,2}$, and $R^2_{R\&B,3}$ measures all have the potential to be negative. Lastly, the $R^2_{R\&B,1}$, $R^2_{R\&B,2}$, and $R^2_{R\&B,3}$ measures focus only on level-specific variance, to the exclusion of total variance. As will be discussed in an upcoming section, examining only level-specific measures can yield misleading results regarding the overall importance of sources at different levels.

[5]These are also often called *pseudo-R^2* measures. Although they can be computed for any null model (Hoffman, 2015), they are interpretable as the overall level-specific variance explained only when using a random-intercept-only null (explained further in Rights & Sterba, 2020).

Singer and Willett's (2003) and Peugh and Heck's (2017) measure

A final existing *total* R^2 measure for three-level MLMs we will discuss (Singer & Willett, 2003, p. 102; also disseminated by Peugh & Heck, 2017, p. 47) involves several steps: first, using the MLM-estimated fixed effects, one computes marginal predicted scores (i.e., not including random effect residuals; $\hat{y}^{(marg)}_{ijk} = \hat{\gamma}_{000} + \mathbf{x}'_{1ijk}\hat{\boldsymbol{\gamma}}_1 + \mathbf{x}'_{2jk}\hat{\boldsymbol{\gamma}}_2 + \mathbf{x}'_{3k}\hat{\boldsymbol{\gamma}}_3$), then computes the Pearson correlation between these predicted scores and the observed scores, and lastly squares this correlation to obtain the $R^2_{S\&W}$:

$$R^2_{S\&W} = corr(\hat{y}^{(marg)}_{ijk}, y_{ijk})^2 \tag{8}$$

Note that, like $R^2_{S\&B}$ and $R^2_{NSJ(m)}$, the $R^2_{S\&W}$ is also a *marginal* total R^2 measure and thus does not help assess relative importance of random effects. Additionally, similar to the other total measures ($R^2_{S\&B}$, $R^2_{NSJ(m)}$, and $R^2_{NSJ(c)}$), the $R^2_{S\&W}$ does not distinguish among variance explained by level-1, -2, and -3 predictors.

One other measure to note was provided by Gelman and Pardoe (2006) but we do not focus on it here because it is specific to Bayesian estimation (e.g., requires computation using posterior simulation draws) whereas the vast majority of MLMs are fit using frequentist estimation methods. However, when computing their measure at the observation-level, it is analogous to a conditional total R^2 (Johnson's [2014] $R^2_{NSJ(c)}$, defined above).

How our framework overcomes limitations of existing three-level R^2 measures

Our framework overcomes limitations of existing three-level R^2s by providing a more comprehensive and flexible set of measures—all obtainable from a single fitted model that can include random slopes, and all related analytically and graphically to facilitate a cohesive interpretation. More specifically, compared with existing R^2s for three-level models, our framework of three-level R^2 measures has the following

Table 1. Definitions of three-level MLM R^2 measures in framework for cluster-mean-centered models*: Interpret as a set of single-source-of-explained-variance measures or sum measures with the same denominator to form a combination-source measure that quantifies variance explained by a combination of sources.

Measure	Definition (Interpretation)
Total MLM R^2 measures: Can be used as a set or in combination	
$R_t^{2(f_1)} = \frac{\gamma'_1\Phi_1\gamma_1}{\text{Eqn. 9}}$	Proportion of total outcome variance explained by level-1 predictors via fixed components of slopes
$R_t^{2(f_2)} = \frac{\gamma'_2\Phi_2\gamma_2}{\text{Eqn. 9}}$	Proportion of total outcome variance explained by level-2 predictors via fixed components of slopes
$R_t^{2(f_3)} = \frac{\gamma'_3\Phi_3\gamma_3}{\text{Eqn. 9}}$	Proportion of total outcome variance explained by level-3 predictors via fixed components of slopes
$R_t^{2(v_{1*2})} = \frac{tr(\Sigma_{1*2}\mathrm{T}_{1*2})}{\text{Eqn. 9}}$	Proportion of total outcome variance explained by level-1 predictors via random slope variation/covariation across level-2 units
$R_t^{2(v_{1*3})} = \frac{tr(\Sigma_{1*3}\mathrm{T}_{1*3})}{\text{Eqn. 9}}$	Proportion of total outcome variance explained by level-1 predictors via random slope variation/covariation across level-3 units
$R_t^{2(v_{2*3})} = \frac{tr(\Sigma_{2*3}\mathrm{T}_{2*3})}{\text{Eqn. 9}}$	Proportion of total outcome variance explained by level-2 predictors via random slope variation/covariation across level-3 units
$R_t^{2(m_2)} = \frac{\tau_{000}}{\text{Eqn. 9}}$	Proportion of total outcome variance explained by cluster-specific outcome means via intercept variation across level-2 units
$R_t^{2(m_3)} = \frac{\varphi_{000}}{\text{Eqn. 9}}$	Proportion of total outcome variance explained by cluster-specific outcome means via intercept variation across level-3 units
Level-1 MLM R^2 measures: Can be used as a set or in combination	
$R_1^{2(f_1)} = \frac{\gamma'_1\Phi_1\gamma_1}{\text{Eqn. 20}}$	Proportion of level-1 outcome variance explained by level-1 predictors via fixed components of slopes
$R_1^{2(v_{1*2})} = \frac{tr(\Sigma_{1*2}\mathrm{T}_{1*2})}{\text{Eqn. 20}}$	Proportion of level-1 outcome variance explained by level-1 predictors via random slope variation/covariation across level-2 units
$R_1^{2(v_{1*3})} = \frac{tr(\Sigma_{1*3}\mathrm{T}_{1*3})}{\text{Eqn. 20}}$	Proportion of level-1 outcome variance explained by level-1 predictors via random slope variation/covariation across level-3 units
Level-2 MLM R^2 measures: Can be used as a set or in combination	
$R_2^{2(f_2)} = \frac{\gamma'_2\Phi_2\gamma_2}{\text{Eqn. 21}}$	Proportion of level-2 outcome variance explained by level-2 predictors via fixed components of slopes
$R_2^{2(v_{2*3})} = \frac{tr(\Sigma_{2*3}\mathrm{T}_{2*3})}{\text{Eqn. 21}}$	Proportion of level-2 outcome variance explained by level-2 predictors via random slope variation/covariation across level-3 units
$R_2^{2(m_2)} = \frac{\tau_{000}}{\text{Eqn. 21}}$	Proportion of level-2 outcome variance explained by cluster-specific outcome means via intercept variation across level-2 units
Level-3 MLM R^2 measures: Can be used as a set	
$R_3^{2(f_3)} = \frac{\gamma'_3\Phi_3\gamma_3}{\text{Eqn. 22}}$	Proportion of level-3 outcome variance explained by level-3 predictors via fixed components of slopes
$R_3^{2(m_3)} = \frac{\varphi_{000}}{\text{Eqn. 22}}$	Proportion of level-3 outcome variance explained by cluster-specific outcome means via intercept variation across level-3 units

Notes. Terms in this table are defined in manuscript text. *Cluster-mean-centered model = model in which *all* lower-level predictors (here, all level-1 and level-2 predictors) are cluster-mean-centered. See Table 5 for corresponding table of measures for three-level *non*-cluster-mean-centered models (i.e., models in which *at least one* lower-level predictor is not cluster-mean-centered).

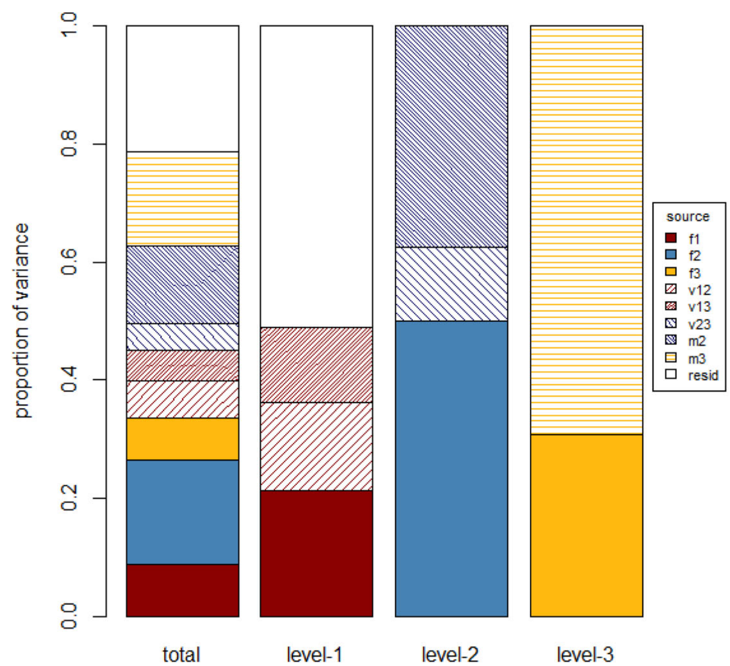
benefits: (1) It provides more generality by subsuming both new and old measures (Online Appendix A provides proofs showing that the separately developed previous R^2s for three-level MLMs [$R^2_{S\&B}$, $R^2_{NSJ(c)}$, $R^2_{NSJ(m)}$, $R^2_{S\&W}$, $R^2_{R\&B,1}$, $R^2_{R\&B,2}$, $R^2_{R\&B,3}$] each correspond, in the population, to a special-case measure computable from our framework, as elaborated later with reference to Table 2). (2) Our framework also provides more flexibility by supplying more options for source(s) contributing to explanation given a particular type of outcome variance—total vs. level-specific (as elaborated later with reference to Table 1). (3) Our framework provides more coherence through graphical visualizations of relationships among its constituent measures, enabling them to be viewed and interpreted as an integrated set (elaborated later with reference to Figure 1). Further details on the first two benefits of the framework are as follows.

First, unlike the *total* R^2 measures previously proposed for three-level models ($R^2_{S\&B}$, $R^2_{NSJ(c)}$, $R^2_{NSJ(m)}$, $R^2_{S\&W}$), our framework provides total measures that allow researchers to distinguish among variance explained by: (1) predictors via fixed components of slopes, vs. (2) predictors via random slope variation, vs. (3) cluster-specific outcome means via intercept variation. Previous total R^2s for three-level models did not distinguish (2) and (3), yet it is foreseeable that researchers could be interested in considering (2) but

Table 2. Relating existing three-level R^2 measures to those in our framework (supporting proofs of population equivalencies are provided in Online Appendix B).

Existing Measure	Equation in current paper	Citation	Using our framework in Table 1, the same population quantity is measured by …
$R^2_{S\&B}$ †	Equation (2) †	Snijders and Bosker (2012) †	$R_t^{2(f_1)} + R_t^{2(f_2)} + R_t^{2(f_3)}$ †
$R^2_{NSJ(m)}$	Equation (3)	Johnson's (2014) extension* of Nakagawa and Schielzeth (2013)	$R_t^{2(f_1)} + R_t^{2(f_2)} + R_t^{2(f_3)}$
$R^2_{NSJ(c)}$	Equation (4)	Johnson's (2014) extension* of Nakagawa and Schielzeth (2013)	$R_t^{2(f_1)} + R_t^{2(f_2)} + R_t^{2(f_3)} + R_t^{2(v_{1\circ 2})} + R_t^{2(v_{1\circ 3})} + R_t^{2(v_{2\circ 3})} + R_t^{2(m_2)} + R_t^{2(m_3)}$
$R^2_{R\&B,1}$	Equation (5)	Raudenbush and Bryk (2002)	$R_1^{2(f_1)} + R_1^{2(v_{1\circ 2})} + R_1^{2(v_{1\circ 3})}$
$R^2_{R\&B,2}$	Equation (6)	Raudenbush and Bryk (2002)	$R_2^{2(f_2)} + R_2^{2(v_{2\circ 3})}$
$R^2_{R\&B,3}$	Equation (7)	Raudenbush and Bryk (2002)	$R_3^{2(f_3)}$
$R^2_{S\&W}$	Equation (8)	Singer and Willett (2003); Peugh and Heck (2017)	$R_t^{2(f_1)} + R_t^{2(f_2)} + R_t^{2(f_3)}$

Notes: † = The $R^2_{S\&B}$ measure requires constraining all random slope variances to 0, but the corresponding measure from our framework does not require this constraint. Hence our measure and $R^2_{S\&B}$ would be estimating the same population quantity if random slope variances were constrained to 0. Because this constraint is neither necessary nor recommended using our framework, this constraint is not demonstrated in this manuscript.
* = The $R^2_{NSJ(m)}$ and $R^2_{NSJ(c)}$ measures for random intercept models proposed by Nakagawa & Schielzeth (2013) were extended to random slope models by Johnson (2014); here we relate Johnson's extended version to our framework.



Notes. Shorthand labels for each source of explained variance are given in the legend; corresponding full definitions of each source are provided in manuscript Equations 10-18. MLM= multilevel model.

Figure 1. Visualization of three-level MLM R^2 measures in framework: Proportions of total variance (column 1), level-1 variance (column 2), level-2 variance (column 3), and level-3 variance (column 4) attributable to each source of explained variance.

not (3) as explained variance (see rationales in Aguinis & Culpepper, 2015; Rights & Sterba, 2018, 2019). Further, again unlike the *total* R^2s previously proposed for three-level models ($R^2_{S\&B}$, $R^2_{NSJ(c)}$, $R^2_{NSJ(m)}$, $R^2_{S\&W}$), our framework provides total R^2s that further unpack (1), from above, by distinguishing among variance explained by: (1a) level-1 predictors vs. (1b) level-2 predictors vs. (1c) level-3 predictors via fixed components of slopes—making it possible to separately assess the contribution of each source individually from one fitted model. Additionally, unlike all of the previously-proposed *total* R^2s for three-level models, our framework provides total R^2s that further unpack (2), from above, by distinguishing among variance attributable to predictors via different kinds of random slope variation: (2a) variation of level-1 slopes across level-2 units, (2b) variation of level-1 slopes across level-3 units, and (2c) variation of level-2 slopes across level-3 units.[6] Our framework of measures accomplishes these generalizations using a more complete partitioning of model-implied total outcome variance.

Unlike the *level-specific* R^2 measures previously proposed for three-level models ($R^2_{R\&B,1}, R^2_{R\&B,2}$, and $R^2_{R\&B,3}$), our framework supplies level-specific R^2s using only a single fitted model. This is not only more convenient than a two-model-fitting approach but, importantly, it also ensures that no measures will be negative (provided a proper solution is obtained). Our framework accomplishes this by providing the first set of model-implied expressions for level-1-specific, level-2-specific, and level-3-specific measures. Further, again unlike the level-specific R^2s previously proposed for three-level models, level-specific measures in our framework distinguish between variance explained by predictors via fixed components of slopes vs. via random slope variation.

Full partitioning of variance used to create a more general framework of R^2 measures

Employing a more general and complete partitioning of model-implied variance than used previously crucially affords us greater flexibility in creating R^2s for three-level MLMs. Taking the variance of Equation (1) yields the following model-implied outcome variance for a three-level MLM:

$$\begin{aligned}\text{var}(y_{ijk}) &= \text{var}(\gamma_{000} + \mathbf{x'}_{1ijk}\boldsymbol{\gamma}_1 + \mathbf{x'}_{2jk}\boldsymbol{\gamma}_2 + \mathbf{x'}_{3k}\boldsymbol{\gamma}_3 \\ &\quad + \mathbf{w'}_{1*2ijk}\mathbf{u}_{1*2jk} + \mathbf{w'}_{1*3ijk}\mathbf{q}_{1*3k} + \mathbf{z'}_{2*3jk}\mathbf{q}_{2*3k} + e_{ijk}) \\ &= \boldsymbol{\gamma'}_1\boldsymbol{\Phi}_1\boldsymbol{\gamma}_1 + \boldsymbol{\gamma'}_2\boldsymbol{\Phi}_2\boldsymbol{\gamma}_2 + \boldsymbol{\gamma'}_3\boldsymbol{\Phi}_3\boldsymbol{\gamma}_3 + tr(\boldsymbol{\Sigma}_{1*2}\mathrm{T}_{1*2}) \\ &\quad + tr(\boldsymbol{\Sigma}_{1*3}\mathrm{T}_{1*3}) + tr(\boldsymbol{\Sigma}_{2*3}\mathrm{T}_{2*3}) + \varphi_{000} + \tau_{000} + \sigma^2\end{aligned} \tag{9}$$

(see derivation in Appendix A, supplementary material) with $\boldsymbol{\Phi}_1, \boldsymbol{\Phi}_2$, and $\boldsymbol{\Phi}_3$ denoting covariance matrices of level-1, level-2, and level-3 predictors, respectively, and $\boldsymbol{\Sigma}_{1*2}$, $\boldsymbol{\Sigma}_{1*3}$, and $\boldsymbol{\Sigma}_{2*3}$ denoting covariance matrices of elements in $\mathbf{w'}_{1*2ijk}$, $\mathbf{w'}_{1*3ijk}$, and $\mathbf{z'}_{2*3jk}$, respectively. Other symbols were defined above.

Each term in Equation (9) reflects variance attributable to one of nine different *sources*:

$$\boldsymbol{\gamma'}_1\boldsymbol{\Phi}_1\boldsymbol{\gamma}_1 = \text{variance attributable to level-1 predictors via fixed components of slopes (shorthand: variance attributable to "}f_1\text{")} \tag{10}$$

$$\boldsymbol{\gamma'}_2\boldsymbol{\Phi}_2\boldsymbol{\gamma}_2 = \text{variance attributable to level-2 predictors via fixed components of slopes (shorthand: variance attributable to "}f_2\text{")} \tag{11}$$

$$\boldsymbol{\gamma'}_3\boldsymbol{\Phi}_3\boldsymbol{\gamma}_3 = \text{variance attributable to level-3 predictors via fixed components of slopes (shorthand: variance attributable to "}f_3\text{")} \tag{12}$$

$$tr(\boldsymbol{\Sigma}_{1*2}\mathrm{T}_{1*2}) = \text{variance attributable to level-1 predictors via slope variation across level-2 units (shorthand: variance attributable to "}v_{1*2}\text{")} \tag{13}$$

$$tr(\boldsymbol{\Sigma}_{1*3}\mathrm{T}_{1*3}) = \text{variance attributable to level-1 predictors via slope variation across level-3 units (shorthand: variance attributable to "}v_{1*3}\text{")} \tag{14}$$

$$tr(\boldsymbol{\Sigma}_{2*3}\mathrm{T}_{2*3}) = \text{variance attributable to level-2 predictors via slope variation across level-3 units (shorthand: variance attributable to "}v_{2*3}\text{")} \tag{15}$$

$$\varphi_{000} = \text{variance attributable to level-3-cluster-specific outcome means via intercept variation (shorthand: variance attributable to "}m_3\text{")} \tag{16}$$

$$\tau_{000} = \text{variance attributable to level-2-cluster-specific outcome means via intercept variation (shorthand: variance attributable to "}m_2\text{")} \tag{17}$$

$$\sigma^2 = \text{variance attributable to level-1 residuals (shorthand: variance attributable to "resid")} \tag{18}$$

Shorthand descriptors of each source (e.g., "f_1") are listed in quotes alongside Equations (10)–(18).

[6]Even when slope variability is not considered substantively interesting in and of itself, variance explained via slope variability could, in future modeling, instead be explained by cross-level interactions via fixed effects (Aguinis & Culpepper, 2015). Breaking down slope variability into component parts can aid researchers in understanding specifically what type of cross-level interactions (e.g., level-1 × level-2, level-1 × level-3, and/or level-2 × level-3) could be considered.

General framework of three-level R^2 measures

Measures in our framework can be computed by fitting a single MLM and take the form:

$$R^2 = \frac{\text{model-implied explained variance}}{\text{model-implied outcome variance}} \tag{19}$$

For *two-level* models, there are three options for what to consider *outcome variance* in the denominator of an R^2: total, level-1 (within-cluster) or level-2 (between-cluster) variance. Also, for two-level models, there are four sources that could potentially contribute to explained variance in the numerator of an R^2 (variance can be explained by predictors via fixed effects at level-1 and/or level-2, by predictors via random slope variation, and/or by cluster-specific outcome means via intercept variation; Rights & Sterba, 2019). For *three-level* models, there are not only more options for what to consider outcome variance in the denominator (total, level-1 [within-level-2-cluster], level-2 [between-level-2-within-level-3-cluster], or level-3 [between-level-3-cluster] variance) but also there are more sources that could potentially contribute to explained variance in the numerator (variance can be explained by predictors via fixed effects at level-1, level-2, and/or level-3, by predictors via random slope variation across level-2 or level-3 units, and/or by cluster-specific outcome means via intercept variation across level-2 or level-3 units).

Having just defined Equations (10)–(17) each as variance attributable to a single *source,* next we describe total and level-specific R^2 s for three-level MLMs that consider each of these separately as single sources of explained variance; we term these the *single-source* measures in our framework. Subsequently we discuss combining them to form *combination-source* measures using our framework.

R^2 measures utilizing a single source of explained variance

Total R^2 measures

Single-source total R^2 measures are created by dividing Equations (10)–(17) each by Equation (9) to yield the proportion of total variance attributable to a single source. Table 1 defines total R^2 measures that each contain a single source of explained variance in the numerator; each measure is represented by one of the shaded segments in the left-most column of Figure 1. Each source is denoted in the Figure 1 legend by the shorthand listed above in Equations (10)–(18). For instance, in the hypothetical Figure 1 example, 9% of the total variance is attributable to level-1 predictors via fixed components of slopes (shorthand "f_1").

Level-specific R^2 measures

We next partition the total model-implied outcome variance from Equation (9) into three levels in the following manner (see derivations in Appendix A, , supplementary material):

$$\begin{aligned}\text{level-1 variance} &= \text{var}_{i|jk}(y_{ijk}) \\ &= \boldsymbol{\gamma}'_1\boldsymbol{\Phi}_1\boldsymbol{\gamma}_1 + tr(\boldsymbol{\Sigma}_{1*2}\mathrm{T}_{1*2}) \\ &\quad + tr(\boldsymbol{\Sigma}_{1*3}\mathrm{T}_{1*3}) + \sigma^2\end{aligned} \tag{20}$$

$$\begin{aligned}\text{level-2 variance} &= \text{var}_{j|k}(y_{ijk}) \\ &= \boldsymbol{\gamma}'_2\boldsymbol{\Phi}_2\boldsymbol{\gamma}_2 + tr(\boldsymbol{\Sigma}_{2*3}\mathrm{T}_{2*3}) + \tau_{000}\end{aligned} \tag{21}$$

$$\text{level-3 variance} = \text{var}_k(y_{ijk}) = \boldsymbol{\gamma}'_3\boldsymbol{\Phi}_3\boldsymbol{\gamma}_3 + \varphi_{000} \tag{22}$$

Equation (20) denotes the level-1 outcome variance (i.e., variance of y_{ijk} across i given j and k), Equation (21) the level-2 outcome variance (i.e., variance of y_{ijk} across j given k), and Equation (22) the level-3 outcome variance (i.e., variance of y_{ijk} across k). Single-source level-1, -2, and -3 R^2 s are listed in Table 1 and are created using the relevant level-specific variance (from Equations (20)–(22)) in the denominator, and one source of explained variance at a time in the numerator. The set of three level-1-specific measures are represented by shaded segments of the second column in Figure 1; likewise, the set of three level-2-specific measures and the set of two level-3-specific measures are represented by shaded segments in the third and fourth columns of Figure 1, respectively. For instance, in the Figure 1 example, 21% of level-1 variance is attributable to level-1 predictors via fixed components of slopes ("f_1"), as seen in the size of its corresponding segment in the second column. Similarly, inspecting the third and fourth columns reveals that 50% of level-2 variance is attributable to level-2 predictors via fixed components of slopes ("f_2") whereas 31% of level-3 variance is attributable to level-3 predictors via fixed components of slopes ("f_3").

Each of the measures in Table 1 can be used as a quantitative effect size to help supplement qualitative interpretation of results. As one example of how the framework aids interpretation of MLM results, one may have a model wherein slopes of level-1 predictors vary across level-3 units, and the degree to which these slopes vary might be meaningful. However, the specific *magnitude* of this slope heterogeneity would be difficult to quantify if one were examining only parameter estimates and associated *p*-values. With $R_t^{2(v_{1*3})}$ and $R_1^{2(v_{1*3})}$ however, we can easily quantify this magnitude and report, respectively, the proportion of total variance and the proportion of level-1 variance explained by level-1 predictors via random slope variation across level-3 units. As another example, the set of all single-source measures in Table 1 can be useful in determining which *types of sources* (*f*, *v*, and/or *m*) and *which levels* (1, 2, and/or 3) are most important in understanding the outcome. For instance, if interest lies in assessing the relative importance of fixed components of slopes for predictors at each level, one could compute and compare $R_t^{2(f_1)}$, $R_t^{2(f_2)}$, and $R_t^{2(f_3)}$.

R^2 measures combining sources of explained variance

One might also be interested in the cumulative impact of multiple sources and thus desire a summary measure of explained variance that combines certain single-source measures in Table 1. To this end, one option is to substantively justify which sources should contribute to "explained" variance and combine these to form a relevant R^2. Here we give three such examples of combination-source measures, though researchers could develop others for their own purposes.

A first option for creating a combination-source measure is to quantify variance explained by *any* predictors via fixed components of slopes; the relevant *total* R^2 measure can then be computed by summing $R_t^{2(f_1)} + R_t^{2(f_2)} + R_t^{2(f_3)}$. As indicated in Table 2, this combination yields a so-called *marginal* total R^2 which is equivalent in the population to Nakagawa-Schielzeth/Johnson's $R^2_{NSJ(m)}$ and Singer and Willett's $R^2_{S\&W}$ (and to Snijders and Bosker's $R^2_{S\&B}$ if slope variability were constrained to 0). See Table 2 and Online Appendix A (supplementary material) for proofs of these population equivalencies. The *marginal* level-specific measures would not require any summation, as there is only one "*f*" measure within each. For instance, a marginal level-3 R^2 is equivalent in the population to Raudenbush and Bryk's $R^2_{R\&B,3}$ (see Table 2 and see Online Appendix A for proof).

Another option for creating a combination-source measure is to consider any potential source in Equations (10)–(17) as explained variance. This approach yields a so-called *conditional* total R^2 (summing all total measures in Table 1) which is equivalent in the population to $R^2_{NSJ(c)}$ (see Table 2 and see Online Appendix A for proof). Likewise, at level-1, this approach yields a *conditional level-1 specific* R^2 measure (summing all level-1 measures in Table 1) which is equivalent in the population to Raudenbush and Bryk's $R^2_{R\&B,1}$ (see Table 2 and see online Appendix A for proof).[7]

A third option for creating a combination-source measure is to quantify variance explained by predictors via both fixed components of slopes and random slope variation (but not by cluster- specific outcome means via intercept variation). This new total measure combines $R_t^{2(f_1)}+R_t^{2(f_2)}+ R_t^{2(f_3)}+R_t^{2(v_{1*2})}+R_t^{2(v_{1*3})}+R_t^{2(v_{2*3})}$. This approach was motivated by Rights and Sterba (2018, 2019) as a *hybrid* measure that serves as a compromise between conditional and marginal approaches. The marginal vs. conditional R^2 distinction had previously been framed as all-or-nothing (i.e., counting as explained all or none of the variance attributable to predictors and cluster means via random effects). However, researchers sometimes want to simultaneously evaluate the importance of fixed and random effects (e.g., Edwards et al., 2008; Jaeger et al., 2017; Kramer, 2005) but are only interested in slope not intercept heterogeneity (e.g., Aguinis & Culpepper, 2015). The corresponding level-2 hybrid measure combines $R_t^{2(f_2)}+R_t^{2(v_{2*3})}$ and is equivalent in the population to Raudenbush and Bryk's $R^2_{R\&B,2}$ (see Table 2 and see Online Appendix A for proof; hence, our framework clarifies that the Raudenbush and Bryk [2002] set of measures are composed of a conditional measure at level-1 ($R^2_{R\&B,1}$), a hybrid measure at level-2 ($R^2_{R\&B,2}$), and a marginal measure at level-3 ($R^2_{R\&B,3}$)).

We have now presented several example combination-source R^2s that researchers could construct using our framework. We emphasize that researchers should use combination-source measures as a supplement to, rather than instead of, single-source measures in Table 1. Otherwise, it is difficult to assess which sources are most important and at which levels variance is primarily explained.

Relationships between total and level-specific R^2 measures

Our Table 1 framework includes level-specific and total measures that have the same source(s) of explained variance (i.e., the same numerator). An example is $R_t^{2(f_1)}$ vs. $R_1^{2(f_1)}$. As the number of levels

[7]This approach would not be useful for level-2 and -3 measures given that combining all potential sources for higher-level R^2 yields a value of 1.

increases, there are increasing numbers of such parallel measures, making it particularly important in three-level (or higher-level) models to understand how these measures are mathematically related when interpreting them simultaneously. Because such analytic relationships have not previously been shown for three-level (nor two-level, nor higher-level) models, we derive these relationships in Online Appendix B while describing them, and their practical implications, here.

Total R^2 and level-specific R^2 measures with the same source(s) of explained variance have a conditionally linear, positive relationship; however, this relationship is *moderated* by the amount of cluster dependency. We operationalize the amount of cluster dependency with the *intraclass correlation* (ICC), defined as the proportion of total variance that is across clusters (Hox, 2010). Specifically, the proportion of outcome variance that is at level-2 is given as:

$$ICC_2 = \frac{\text{level-2 variance}}{\text{total variance}} = \frac{\gamma'_2\Phi_2\gamma_2 + tr(\Sigma_{2*3}T_{2*3}) + \tau_{000}}{\gamma'_1\Phi_1\gamma_1 + \gamma'_2\Phi_2\gamma_2 + \gamma'_3\Phi_3\gamma_3 + tr(\Sigma_{1*2}T_{1*2}) + tr(\Sigma_{1*3}T_{1*3}) + tr(\Sigma_{2*3}T_{2*3}) + \varphi_{000} + \tau_{000} + \sigma^2} \quad (23)$$

Similarly, the proportion of outcome variance that is at level-3 is:

$$ICC_3 = \frac{\text{level-3 variance}}{\text{total variance}} = \frac{\gamma'_3\Phi_3\gamma_3 + \varphi_{000}}{\gamma'_1\Phi_1\gamma_1 + \gamma'_2\Phi_2\gamma_2 + \gamma'_3\Phi_3\gamma_3 + tr(\Sigma_{1*2}T_{1*2}) + tr(\Sigma_{1*3}T_{1*3}) + tr(\Sigma_{2*3}T_{2*3}) + \varphi_{000} + \tau_{000} + \sigma^2} \quad (24)$$

And the proportion of outcome variance that is across both level-2 and level-3 clusters is:

$$ICC_{23} = ICC_2 + ICC_3 \quad (25)$$

With these formulas,[8] we can now establish the following relationships:

1. Consider a level-1 measure (generically denoted $R_1^{2(s)}$) and total measure (generically denoted $R_t^{2(s)}$) with the same generic source "s" of explained variance at level-1. These total and level-1 measures are related purely through ICC_{23}, as such:

$$R_1^{2(s)} = \frac{R_t^{2(s)}}{1 - ICC_{23}} \quad (26)$$

Equation (26) implies that, as higher-level clustering increases from ICC_{23}=0 to ICC_{23}=1, the level-1 measure $R_1^{2(s)}$ exponentially increases from its lower bound (where it is exactly equal to the total measure, i.e., $R_1^{2(s)}=R_t^{2(s)}$) to its upper bound (where $R_1^{2(s)}$ =1). This relationship is illustrated in Figure 2 Panel A.

2. Similarly, the total measure ($R_t^{2(s)}$) and *level-2* measure ($R_2^{2(s)}$) for a generic source "s" of explained variance *at level-2* are related purely through the degree of clustering at level-2 (i.e., ICC_2):

$$R_2^{2(s)} = \frac{R_t^{2(s)}}{ICC_2} \quad (27)$$

Equation (27) implies that as between-level-2 cluster dependency increases from ICC_2=0 to ICC_2=1, the level-2 measure will exponentially increase from its lower bound (where $R_2^{2(s)} = R_t^{2(s)}$) to its upper bound (where $R_2^{2(s)} = 1$). This relationship is illustrated in Figure 2 Panel B.

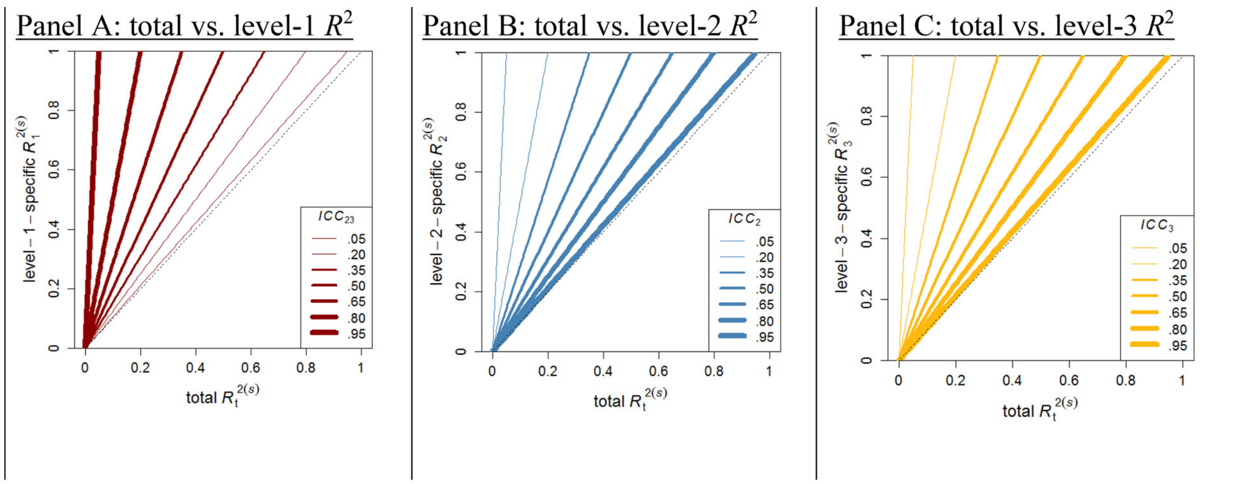
3. Likewise, the total measure ($R_t^{2(s)}$) and *level-3* measure ($R_3^{2(s)}$) for a generic source "s" of explained variance *at level-3* are related purely through between-level-3 cluster dependency (i.e., ICC_3) in the same manner as total and level-2 R^2 s were (above) related through ICC_2 . This relationship is illustrated in Figure 2 Panel C.[9]

$$R_3^{2(s)} = \frac{R_t^{2(s)}}{ICC_3} \quad (28)$$

As a concrete example of why these relationships between level-specific and total measures are important to understand in practice, suppose two studies report estimates of .80 for a level-1 measure with a particular source, s, of explained variance at level-1. Based on this information, it is tempting to think that source s is equally important in both studies. However, the total measure for source s could be .70 in one study but

[8]Note that these ICC formulas in Equations (23), (24), and (25) apply for any three-level model in which all level-1 and level-2 predictors are cluster-mean-centered in the manner described earlier. For models in which there is at least one level-1 or level-2 predictor that is not cluster-mean-centered (regardless of the specific alternative centering method employed), the formula for ICC_2 would be given as (Equation B24)/(Equation B22), ICC_3 would be (Equation B25/Equation B22), and ICC_{23} would be (Equation B24 + Equation B25)/(Equation B22); see Appendix B.

[9]These same general relationships illustrated in Figure 2 hold for MLMs involving more or fewer than three levels. Denoting the levels of a MLM as l=1 … L, relationship #1 from above would hold for an L-level model when replacing mentions of "level-2 and level-3" with "all levels > 1." Similarly, relationships #2 and #3 would hold for an L-level model when replacing mentions of "level-2" or "level-3," with "level-l" (where $1 < l \leq L$).



Notes. The set of lines varying in thickness denote alternative degrees of clustering. ICC_{23}, ICC_2, and ICC_3 were defined in manuscript text.

Figure 2. The mathematical relationships between total R^2 and level-specific R^2 for a generic source s of explained variance, where in Panel A source s is at level-1, in Panel B source s is at level-2, and in Panel C, source s is at level-3.

only .05 in the other study (both possibilities are shown in Figure 2 Panel A). Hence, it would be problematic to *solely* report the level-specific measure, as readers could be misled into thinking source s is highly important in understanding the outcome, when actually a large level-specific R^2 could be more attributable to high level-2 and/or level-3 cluster dependency rather than the actual amount of total variance being explained. Consequently, considering *only a level-specific measure* conveys little about how much variance is explained by the source *relative to the total variance*, and vice versa. We therefore recommend considering level-specific and total measures simultaneously.

Empirical example

To illustrate our approach for a three-level model, we analyzed data on first-grade students from the Study of Instructional Improvement by Hill, Rowan, and Ball (2005). These data consist of 1081 students nested in 285 teachers nested in 105 schools, and are given as an example for three-level models in a popular MLM textbook (Rabe-Hesketh & Skrondal, 2008). Here we show how the use of our R^2 measures can aid in substantive interpretation. The substantive focus of this application is in predicting students' math achievement difference scores from kindergarten to first grade.[10] Of primary interest is whether or not teachers' math preparation and content knowledge has an effect on these difference scores. The student-level predictors (centered both by teacher-mean and by school-mean) are math achievement score from kindergarten, sex (coded 1 for girl), and socioeconomic status (SES). The teacher-level predictors (each school-mean-centered) are the first-grade teacher's math preparation (a score based on number of math content and methods courses), the first-grade teacher's math content knowledge (based on a 30-item scale), and, as a control variable, teacher's years of experience. School-level predictors are school means of each of the aforementioned student-level and classroom-level predictors, namely, school means of kindergarten math, sex, SES, and teacher's math preparation, math content knowledge, and years of experience. The fixed portion of the model included a fixed component of the intercept and fixed component of each predictor's slope. For the random portion of the model, the intercept and the slope of kindergarten math achievement varied across *both* classrooms and schools (yielding two random intercept and two random slope components). These random slope components were included to account for the possibility that kindergarten math score is less predictive of math gains for certain classrooms and schools. For instance, some teachers or schools may teach material at the same level for all students, whereas others provide more personalized instruction to students commensurate with individual math ability; in the latter case, kindergarten math score would likely be less predictive of math gains.

The MLM was fit using the *lmer* function in the R package *lme4* (Bates, Maechler, Bolker, & Walker, 2014) with restricted maximum likelihood (REML)

[10]For didactic purposes we slightly modified (including using cluster-mean-centering and random slopes) the model described in Rabe-Hasketh and Skrondal (2008) and Hill et al. (2005).

Table 3. Empirical example parameter estimate results: Predicting gains in math scores from kindergarten to first grade.

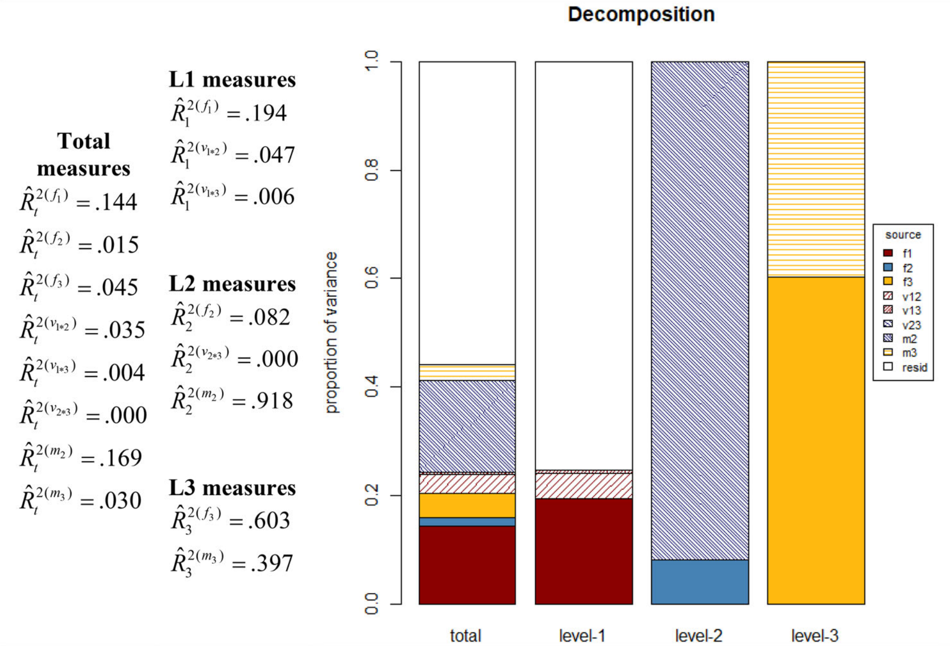
	Est	*SE*	*t*
Fixed effects			
intercept	57.766	1.368	42.229*
kindergarten math score	−0.431	0.033	−13.128*
sex (1 = girl; 0 = boy)	−2.112	1.896	−1.114
socioeconomic status (SES)	5.804	1.451	4.000*
teacher math preparation	1.688	1.571	1.075
teacher math knowledge	5.935	1.722	3.447*
teacher years of experience	−0.053	0.168	−0.313
school-mean kindergarten math score	−0.325	0.067	−4.828*
school-mean sex (proportion girls)	2.074	8.434	0.246
school-mean SES	5.025	4.024	1.249
school-mean teacher math preparation	2.460	2.281	1.078
school-mean teacher math knowledge	−1.403	1.815	−0.773
school-mean teacher years of experience	0.406	0.208	1.957
Variance components (class-level)			
	Est	χ^2 †	
Intercept	204.338	35.785*	
kindergarten math score	0.044	8.380*	
intercept-math-score covariance	−0.661	0.895	
Variance components (school-level)			
	Est	χ^2 †	
intercept	36.227	1.619	
kindergarten math score	0.005	0.269	
intercept-math-score covariance	−0.193	0.180	
Residual (student-level)			
	Est		
level-1 residual	674.940		

Notes: *$p < .001$ † Each random intercept and slope variance was tested using a mixture chi-square LRT comparing a model without the random component to a model with the random component, allowing all random effects to covary at the same level (Ke & Wang, 2015). The null reference distribution for each of these tests was a 50:50 mixture of $\chi^2_{df=1}$ and $\chi^2_{df=2}$ (Stram & Lee, 1994). Each random effect covariance was individually tested using an LRT with null distribution $\chi^2_{df=1}$. For an overview of alternative random effect testing approaches, see Rights and Sterba (2016).

estimation. MLM parameter estimates are presented in Table 3 and R^2 results for each measure in Table 1 are presented in Figure 3, along with the corresponding barchart.[11] As seen in Table 3, fixed effects of student-level kindergarten math and SES, teacher-level math knowledge, and school-mean kindergarten math score were all significant, as were the across-classroom random intercept and slope variances. Here we give three illustrations of how R^2 results from our framework provide useful, supplemental indications of practical significance that would be difficult to ascertain by simply examining the MLM parameter estimates and associated p-values in Table 3, as typically done in practice (e.g., Hill et al., 2005).

As a first illustration, recall that one primary substantive research question was whether teacher-level characteristics were predictive of student math gains on average. From the estimated MLM fixed coefficients, there is evidence that this is the case, particularly in that teacher math knowledge has a positive relationship with math gains ($p < .05$), consistent with results from Hill et al. (2005). Furthermore, as discussed by Hill et al. (2005), it is also consistent with earlier research suggesting that more knowledgeable teachers have a better understanding of connections among underlying elementary math concepts and also provide better math explanations to their students (e.g., Borko et al., 1992; Carpenter et al., 1989; Leinhardt & Smith, 1985; Ma, 1999; Thompson & Thompson, 1994). However, despite this statistically significant finding, examining the R^2 results in Table 3 reveals the supplementary information that the teacher-level predictors explain only 1% of the total variance ($\hat{R}_t^{2(f_2)} = .015$), which is reflected by the small solid [blue] section in the leftmost stacked bar of Figure 3. Even when isolating variance to teacher-level variance, only 8% is explained by level-2 predictors via fixed effects ($\hat{R}_2^{2(f_2)} = .082$), leaving 92% to be explained by teacher-level characteristics not included as predictors in the model ($\hat{R}_2^{2(m_2)} = .918$). Though this does not preclude the possibility that

[11]As an example calculation, in Figure 3, the estimate of $R_1^{2(f_1)}$ (i.e., the proportion of level-1 variance explained by level-1 predictors via fixed components of slopes) is computed as the ratio of the estimate of Equation 10, $[-0.431 \quad -2.112 \quad 5.804]\widehat{\text{var}}(\mathbf{x}_{1ijk})[-0.431 \quad -2.112 \quad 5.804]'$, to the estimate of Equation 20, $[-0.431 \quad -2.112 \quad 5.804]\widehat{\text{var}}(\mathbf{x}_{1ijk})[-0.431 \quad -2.112 \quad 5.804]' + \widehat{\text{var}}(mathkind)\,0.044 + \widehat{\text{var}}(mathkind)0.005 + 674.940$, which equals .144 (where $\widehat{\text{var}}(\mathbf{x}_{1ijk})$ is the estimated covariance matrix of the three level-1 predictors, and $\widehat{\text{var}}(mathkind)$ the estimated variance of school-and-teacher-mean-centered kindergarten math).



Notes. These R^2 estimates supplement empirical example multilevel model (MLM) parameter estimate results in Table 3. Shorthand labels for each source of explained variance are given in the legend; corresponding full definitions of each source were in manuscript Equations 10-18.

Figure 3. Empirical example three-level model R^2 results.

future interventions serving to increase teachers' math knowledge could benefit student scores, results here suggest that the observed variation in teacher knowledge accounts for little of the variation in math scores. This result would have been missed had we relied exclusively on preexisting three-level R^2 measures that combine variance explained by predictors at every level; doing so would have indicated a sizable portion of variance explained (driven primarily by student-level and school-level predictors).

As a second illustration, another useful way to consider MLM parameter estimates together with R^2s from our framework regards understanding the impact of the random slopes. Examining random slope variance component estimates and associated p-values might lead one to think that the slope heterogeneity is unimportant, given the small point estimates (0.044 and 0.005). In the example, however, slope heterogeneity across teachers explains much more of the total variance than do the teacher-level predictors via fixed effects ($\hat{R}_t^{2(v_{1*2})} = .035$), which were of primary interest. Thus, across-teacher variability in the degree to which kindergarten math scores explain math gains might be important to consider. Even if a researcher is not interested in slope heterogeneity in and of itself, its presence elucidates potential cross-level interactions that can be included in future modeling. Had we solely relied on preexisting three-level R^2s that combine variance attributable to both fixed and random effects, the contribution of these random slopes would not have been clear.

As a final illustration, to supplement the suite of single-source measures, suppose the researcher also wants to compute a combination-source measure to get a sense of how much total variance can be explained jointly by all sources considered substantively meaningful (as desired in Johnson, 2014; Nakagawa & Schielzeth, 2013; Peugh & Heck, 2017; Singer & Willett, 2003; Snijders & Bosker, 2012). To do so, we simply combine single-source measures (i.e., individual segments from the first barchart in Figure 3). For instance, if a researcher were interested in total variance explained by all predictors via fixed effects (as in Nakagawa & Schielzeth/Johnson's and Singer & Willett's "marginal" measures $R^2_{NSJ(m)}$ and $R^2_{S\&W}$) then combining ($\hat{R}_t^{2(f_1)} + \hat{R}_t^{2(f_2)} + \hat{R}_t^{2(f_3)}$) from Figure 3 indicates that 20% of total variance is meaningfully explained (of which we know from earlier that teacher-level predictors contribute little). If a researcher were interested in specifically quantifying the impact of terms at a subset of levels they could, for instance, instead quantify the proportion of variance explained by just teacher-level and school-level predictors via fixed effects as $\hat{R}_t^{2(f_2)} + \hat{R}_t^{2(f_3)} = .159$. Rather than focusing solely on the fixed effects,

Table 4. R^2 measures for L-level cluster-mean-centered models*: Interpret as a set of single-source-of-explained-variance measures or sum measures with the same denominator to quantify variance explained by a combination of sources.

Measure	Definition (Interpretation)
Total MLM R^2 measures	
$R_t^{2(f_l)} = \frac{\gamma'_l \Phi_l \gamma_l}{\text{Eqn. A9}}$	Proportion of total outcome variance explained by level-l predictors via fixed components of slopes
$R_t^{2(v_{l*q})} = \frac{tr(\Sigma_{l*q} T_{l*q})}{\text{Eqn. A9}}$	Proportion of total outcome variance explained by level-l predictors via random slope variation across level-q units (where $q > l$)
$R_t^{2(m_l)} = \frac{\tau_l}{\text{Eqn. A9}}$	Proportion of total outcome variance explained by level-l (where $l > 1$) cluster-specific outcome means via random intercept variation
Level-1 MLM R^2 measures	
$R_1^{2(f_1)} = \frac{\gamma'_1 \Phi_1 \gamma_1}{\text{Eqn. A10}}$	Proportion of level-1 outcome variance explained by level-1 predictors via fixed components of slopes
$R_1^{2(v_{1*q})} = \frac{tr(\Sigma_{1*q} T_{1*q})}{\text{Eqn. A10}}$	Proportion of level-1 outcome variance explained by level-1 predictors via random slope variation across level-q units (where $q > 1$)
Level-M (intermediate level in which $1 < M < L$) MLM R^2 measures	
$R_M^{2(f_M)} = \frac{\gamma'_M \Phi_M \gamma_M}{\text{Eqn. A11}}$	Proportion of level-M outcome variance explained by level-M predictors via fixed components of slopes
$R_M^{2(v_{M*q})} = \frac{tr(\Sigma_{M*q} T_{M*q})}{\text{Eqn. A11}}$	Proportion of level-M outcome variance explained by level-M predictors via random slope variation across level-q units (where $q > M$)
$R_M^{2(m_M)} = \frac{\tau_M}{\text{Eqn. A11}}$	Proportion of level-M outcome variance explained by level-M cluster-specific outcome means via random intercept variation
Level-L MLM R^2 measures	
$R_L^{2(f_L)} = \frac{\gamma'_L \Phi_L \gamma_L}{\text{Eqn. A12}}$	Proportion of level-L outcome variance explained by level-L predictors via fixed components of slopes
$R_L^{2(m_L)} = \frac{\tau_L}{\text{Eqn. A12}}$	Proportion of level-L outcome variance explained by level-L cluster-specific outcome means via random intercept variation

Notes. Terms in this table are defined in Appendix A. *Cluster-mean-centered model = model in which *all* lower-level (i.e., below level-L) predictors are cluster-mean-centered. See Table 6 for corresponding table of measures for L-level *non*-cluster-mean-centered models (i.e., models in which *at least one* lower-level predictor is not cluster-mean-centered). Also note that L is the integer denoting the highest level, M is > 1 and $< L$, l can be an integer from 1 to L, and q can be an integer from 2 to L.

another researcher might be interested in all potential sources of explained variance at all levels (as in Nakagawa & Schielzeth/Johnson's "conditional" measure $R^2_{NSJ(c)}$), combining $(\hat{R}_t^{2(f_1)} + \hat{R}_t^{2(f_2)} + \hat{R}_t^{2(f_3)} + \hat{R}_t^{2(v_{1*2})} + \hat{R}_t^{2(v_{1*3})} + \hat{R}_t^{2(v_{2*3})} + \hat{R}_t^{2(m_2)} + \hat{R}_t^{2(m_3)})$ from Figure 3 indicates that 44% of the total variance is explained. Finally, if instead a researcher were interested in total variance explained by predictors via fixed effects and random slope variation (a compromise we offer between "marginal" and "conditional" perspectives) they could combine $(\hat{R}_t^{2(f_1)} + \hat{R}_t^{2(f_2)} + \hat{R}_t^{2(f_3)} + \hat{R}_t^{2(v_{1*2})} + \hat{R}_t^{2(v_{1*3})} + \hat{R}_t^{2(v_{2*3})})$ to report that 24% of total variance is explained. This number may be useful in succinctly summarizing the overall impact of all predictors, without including the impact of random intercept variation (which may not be of substantive interest).

Generalizing the framework of MLM R^2 to any number of levels

Until now we have restricted focus to three-level models for simplicity of illustration. However, noting that models with more than three levels may become more common in the future, we additionally provide measures that can quantify variance explained for multilevel models with any number of levels. In Appendix A we first generalize the three-level model expression in Equation (1) to L-levels (in which L is any integer > 1). We then derive the model-implied total outcome variance for a *L-level cluster-mean-centered model*—defined as a model in which *all* predictors below level-L are cluster-mean-centered—providing a decomposition involving the following possible *sources* of explained variance at a generic level-l (in which l is any integer from 1 to L):

- level-l predictors via fixed components of slopes (shorthand: variance attributable to "f_l")

- level-l predictors via random slope variation across level-q units (in which q is an integer from $l+1$ to L) (shorthand: variance attributable to "v_{l*q}")
- level-l cluster-specific outcome means via random intercept variation (for each $l>1$) (shorthand: variance attributable to "m_l")

From the decomposition of total outcome variance for L-level models provided in Appendix A researchers can quantify total variance explained by f_l (with $R_t^{2(f_l)}$), v_{l*q} (with $R_t^{2(v_{l*q})}$), and m_l (with $R_t^{2(m_l)}$). A full list of R^2 measure definitions for cluster-mean-centered L-level models are provided in Table 4.

Additionally in Appendix A we derive model-implied level-specific outcome variances for L-level cluster-mean-centered models, and in Table 4, we define the available level-specific R^2 measures for L-level cluster-mean-centered models. At level-1, researchers can quantify variance explained by f_1 (with $R_1^{2(f_1)}$) and v_{1*q} (with $R_1^{2(v_{1*q})}$). At any intermediate level-M (i.e., where M is >1 and $<L$), researchers can quantify variance explained by f_M (with $R_M^{2(f_M)}$), v_{M*q} (with $R_M^{2(v_{M*q})}$), and m_M (with $R_M^{2(m_M)}$). At the highest level, researchers can quantify variance explained by f_L (with $R_L^{2(f_L)}$) and m_L (with $R_L^{2(m_L)}$). Next we derive and describe all corresponding total and level-specific R^2 measures for *non*-cluster-mean-centered L-level models (in Appendix B, and Table 6).

Impact of centering choice on R^2 measures

We thus far have focused on R^2s for MLMs that use cluster-mean-centering for all lower-level predictors (i.e., all predictors below the highest level), which ensures that each of these predictors varies only at a single level. However, in some contexts, researchers might prefer to fit MLMs that do not utilize cluster-mean-centering for at least some lower-level predictors. The typical alternative to cluster-mean-centering is *centering-at-a-constant* (which subsumes centering lower-level predictors by the grand-mean or the origin, or even leaving them uncentered). For instance, sometimes researchers include a non-cluster-mean-centered level-1 predictor even though it has between-level-2-cluster and/or between-level-3-cluster variance, because they are comfortable assuming that its within-cluster effect is equivalent to the between-cluster effect(s) (Snijders & Bosker, 2012). As another example, for a three-level longitudinal growth model (e.g., repeated measures nested within students nested within schools), sometimes researchers want to center the level-1 predictor *time* at a constant such that 0 = baseline, so that they can interpret the intercept as reflecting baseline levels of the outcome (e.g., Biesanz et al., 2004; Hoffman, 2015).

For such researcher interests, we here derive a full decomposition[12] of variance for L-level *non*-cluster-mean-centered MLMs in Appendix B. We define a *non-cluster-mean-centered model* as one in which *at least one* lower-level (i.e., below level-L) predictor is not cluster-mean-centered. Hence, if a researcher specified a model that included some cluster-mean-centered lower-level predictors and some non-cluster-mean-centered lower-level predictors, they would use the decomposition for *non*-cluster-mean-centered MLMs. In Table 5, we provide the formulas and definitions for the full set of total and level-specific R^2 measures specifically for three-level *non*-cluster-mean-centered MLMs. The Table 5 R^2 measures for three-level non-cluster-mean-centered MLMs are the counterpart to the Table 1 R^2 measures for three-level cluster-mean-centered MLMs. Similarly, in Table 6 we provide general R^2 formulas and definitions for L-level non-cluster-mean-centered models, which are the counterpart of Table 4's general R^2 formulas and definitions for L-level cluster-mean-centered models. Note that level-specific measures for non-cluster-mean-centered MLMs had not previously been derived or incorporated into the framework, even for two-level MLMs in Rights and Sterba (2019).

Having provided one set of R^2 measures assuming cluster-mean-centering and another set of R^2 measures *not* assuming cluster-mean-centering, researchers may wonder how R^2 interpretations would differ across alternative centering options. The first interpretational difference to note is that, in non-cluster-mean-centered models, if lower-level predictors have any across-cluster variance, they can explain variance at multiple levels (whereas in cluster-mean-centered models each individual predictor can only explain variance at a single level). For instance, in a three-level non-cluster-mean-centered model, if a level-1 predictor has mean variation across level-2 units within level-3 units, it can explain level-2 outcome variance, and if it has mean variation across level-3

[12]Note that when researchers do not utilize cluster-mean-centering and have predictors with random slopes that do not have a mean of 0 (e.g., centering-at the origin for *time* with a random slope of *time*), the variance attributable to source *m* (i.e., the between-cluster variance not attributable to *f* or *v*) necessarily has a different expression than under cluster-mean-centering (see Equations 16-17 vs. Equation B17) in order for the decomposition of variance to add up to the model-implied outcome variance. However, if all predictors with random slopes have a mean of 0 in a non-cluster-mean-centered model, the expression for the variance attributable to source *m* (Equation B17) is equivalent to (i.e., simplifies to) the expression for cluster-mean-centered models (Equation 16/17).

Table 5. Definitions of three-level MLM R^2 measures in framework for *non*-cluster-mean-centered models*: Interpret as a set of single-source-of-explained-variance measures or sum measures with the same denominator to quantify variance explained by a combination of sources.

Measure	Definition (Interpretation)
Total MLM R^2 measures: Can be used as a set or in combination	
$R_t^{2(f_1)} = \frac{\gamma'\Phi_1\gamma}{\text{Eqn. B22}}$	Proportion of total outcome variance explained by the level-1-varying portion of predictors via fixed components of slopes
$R_t^{2(f_2)} = \frac{\gamma'\Phi_2\gamma}{\text{Eqn. B22}}$	Proportion of total outcome variance explained by the level-2-varying portion of predictors via fixed components of slopes
$R_t^{2(f_3)} = \frac{\gamma'\Phi_3\gamma}{\text{Eqn. B22}}$	Proportion of total outcome variance explained by the level-3-varying portion of predictors via fixed components of slopes
$R_t^{2(v_{1*2})} = \frac{tr(\Sigma_{1*2}T_2)}{\text{Eqn. B22}}$	Proportion of total outcome variance explained by the level-1-varying portion of predictors via random slope variation/covariation across level-2 units
$R_t^{2(v_{1*3})} = \frac{tr(\Sigma_{1*3}T_3)}{\text{Eqn. B22}}$	Proportion of total outcome variance explained by the level-1-varying portion of predictors via random slope variation/covariation across level-3 units
$R_t^{2(v_{2*2})} = \frac{tr(\Sigma_{2*2}T_2)}{\text{Eqn. B22}}$†	Proportion of total outcome variance explained by the level-2-varying portion of predictors via random slope variation/covariation across level-2 units
$R_t^{2(v_{2*3})} = \frac{tr(\Sigma_{2*3}T_3)}{\text{Eqn. B22}}$	Proportion of total outcome variance explained by the level-2-varying portion of predictors via random slope variation/covariation across level-3 units
$R_t^{2(v_{3*2})} = \frac{tr(\Sigma_{3*2}T_2)}{\text{Eqn. B22}}$†	Proportion of total outcome variance explained by the level-3-varying portion of predictors via random slope variation/covariation across level-2 units
$R_t^{2(v_{3*3})} = \frac{tr(\Sigma_{3*3}T_3)}{\text{Eqn. B22}}$†	Proportion of total outcome variance explained by the level-3-varying portion of predictors via random slope variation/covariation across level-3 units
$R_t^{2(m_2)} = \frac{\sum_{q=1}^{3} \mathbf{m}'_{q*2}T_2\mathbf{m}_{q*2}}{\text{Eqn. B22}}$	Proportion of total outcome variance explained by level-2 cluster-specific outcome means via intercept variation at the mean of all predictors with random slopes
$R_t^{2(m_3)} = \frac{\sum_{q=1}^{3} \mathbf{m}'_{q*3}T_3\mathbf{m}_{q*3}}{\text{Eqn. B22}}$	Proportion of total outcome variance explained by level-3 cluster-specific outcome means via intercept variation at the mean of all predictors with random slopes
Level-1 MLM R^2 measures: Can be used as a set or in combination	
$R_1^{2(f_1)} = \frac{\gamma'\Phi_1\gamma}{\text{Eqn. B23}}$	Proportion of level-1 outcome variance explained by the level-1-varying portion of predictors via fixed components of slopes
$R_1^{2(v_{1*2})} = \frac{tr(\Sigma_{1*2}T_2)}{\text{Eqn. B23}}$	Proportion of level-1 outcome variance explained by the level-1-varying portion of predictors via random slope variation/covariation across level-2 units
$R_1^{2(v_{1*3})} = \frac{tr(\Sigma_{1*3}T_3)}{\text{Eqn. B23}}$	Proportion of level-1 outcome variance explained by the level-1-varying portion of predictors via random slope variation/covariation across level-3 units
Level-2 MLM R^2 measures: Can be used as a set or in combination	
$R_2^{2(f_2)} = \frac{\gamma'\Phi_2\gamma}{\text{Eqn. B24}}$	Proportion of level-2 outcome variance explained by the level-2-varying portion of predictors via fixed components of slopes
$R_2^{2(v_{2*2})} = \frac{tr(\Sigma_{2*2}T_2)}{\text{Eqn. B24}}$†	Proportion of level-2 outcome variance explained by the level-2-varying portion of predictors via random slope variation/covariation across level-2 units
$R_2^{2(v_{2*3})} = \frac{tr(\Sigma_{2*3}T_3)}{\text{Eqn. B24}}$	Proportion of level-2 outcome variance explained by the level-2-varying portion of predictors via random slope variation/covariation across level-3 units
$R_2^{2(m_2)} = \frac{\sum_{q=1}^{3} \mathbf{m}'_{q*2}T_2\mathbf{m}_{q*2}}{\text{Eqn. B24}}$	Proportion of level-2 outcome variance explained by level-2 cluster-specific outcome means via intercept variation at the mean of all predictors with random slopes
Level-3 MLM R^2 measures: Can be used as a set or in combination	
$R_3^{2(f_3)} = \frac{\gamma'\Phi_3\gamma}{\text{Eqn. B25}}$	Proportion of level-3 outcome variance explained by the level-3-varying portion of predictors via fixed components of slopes
$R_3^{2(v_{3*2})} = \frac{tr(\Sigma_{3*2}T_2)}{\text{Eqn. B25}}$†	Proportion of level-3 outcome variance explained by the level-3-varying portion of predictors via random slope variation/covariation across level-2 units
$R_3^{2(v_{3*3})} = \frac{tr(\Sigma_{3*3}T_3)}{\text{Eqn. B25}}$†	Proportion of level-3 outcome variance explained by the level-3-varying portion of predictors via random slope variation/covariation across level-3 units
$R_3^{2(m_3)} = \frac{\sum_{q=1}^{3} \mathbf{m}'_{q*3}T_3\mathbf{m}_{q*3}}{\text{Eqn. B25}}$	Proportion of level-3 outcome variance explained by level-3 cluster-specific outcome means via intercept variation at the mean of all predictors with random slopes

Notes. Terms in this table are defined in Appendix B. *Non-cluster-mean-centered models = models in which *at least one* lower-level (here, level-1 or level-2) predictor is *not* cluster-mean-centered. See Table 1 for corresponding table of measures for three-level cluster-mean-centered models (i.e., models in which *all* lower-level predictors are cluster-mean-centered). †For these measures, there is no analogous quantity for cluster-mean-centered models.

Table 6. R^2 measures for L-level non-cluster-mean-centered* models: Interpret as a set of single-source-of-explained-variance measures or sum measures with the same denominator to quantify variance explained by a combination of sources.

Measure	Definition (Interpretation)
Total MLM R^2 measures	
$R_t^{2(f_l)} = \frac{\boldsymbol{\gamma}'\boldsymbol{\Phi}_l\boldsymbol{\gamma}}{\text{Eqn. B14}}$	Proportion of total outcome variance explained by the level-l-varying portion of predictors via fixed components of slopes
$R_t^{2(v_{l*q})} = \frac{tr(\boldsymbol{\Sigma}_{l*q}\mathrm{T}_q)}{\text{Eqn. B14}}$	Proportion of total outcome variance explained by the level-l-varying portion of predictors via random slope variation across level-q units
$R_t^{2(m_l)} = \frac{\sum_{q=1}^{L}\mathbf{m}'_{q*l}\mathrm{T}_l\mathbf{m}_{q*l}}{\text{Eqn. B14}}$	Proportion of total outcome variance explained by level-l (where $l > 1$) cluster-specific outcome means via random intercept variation at the mean of all predictors with random slopes
Level-1 MLM R^2 measures	
$R_1^{2(f_1)} = \frac{\boldsymbol{\gamma}'\boldsymbol{\Phi}_1\boldsymbol{\gamma}}{\text{Eqn. B19}}$	Proportion of level-1 outcome variance explained by the level-1-varying portion of predictors via fixed components of slopes
$R_1^{2(v_{1*q})} = \frac{tr(\boldsymbol{\Sigma}_{1*q}\mathrm{T}_q)}{\text{Eqn. B19}}$	Proportion of level-1 outcome variance explained by the level-1-varying portion of predictors via random slope variation across level-q units
Level-M (intermediate level in which $1 < M < L$) MLM R^2 measures	
$R_M^{2(f_M)} = \frac{\boldsymbol{\gamma}'\boldsymbol{\Phi}_M\boldsymbol{\gamma}}{\text{Eqn. B20}}$	Proportion of level-M outcome variance explained by the level-M-varying portion of predictors via fixed components of slopes
$R_M^{2(v_{M*q})} = \frac{tr(\boldsymbol{\Sigma}_{M*q}\mathrm{T}_q)}{\text{Eqn. B20}}$	Proportion of level-M outcome variance explained by the level-M-varying portion of predictors via random slope variation across level-q units
$R_M^{2(m_M)} = \frac{\sum_{q=1}^{L}\mathbf{m}'_{q*M}\mathrm{T}_M\mathbf{m}_{q*M}}{\text{Eqn. B20}}$	Proportion of level-M outcome variance explained by level-M cluster-specific outcome means via random intercept variation at the mean of all predictors with random slopes
Level-L MLM R^2 measures	
$R_L^{2(f_L)} = \frac{\boldsymbol{\gamma}'\boldsymbol{\Phi}_L\boldsymbol{\gamma}}{\text{Eqn. B21}}$	Proportion of level-L outcome variance explained by the level-L-varying portion of predictors via fixed components of slopes
$R_L^{2(v_{L*q})} = \frac{tr(\boldsymbol{\Sigma}_{L*q}\mathrm{T}_q)}{\text{Eqn. B21}}$	Proportion of level-L outcome variance explained by the level-L-varying portion of predictors via random slope variation across level-q units
$R_L^{2(m_L)} = \frac{\sum_{q=1}^{L}\mathbf{m}'_{q*L}\mathrm{T}_L\mathbf{m}_{q*L}}{\text{Eqn. B21}}$	Proportion of level-L outcome variance explained by level-L cluster-specific outcome means via random intercept variation at the mean of all predictors with random slopes

Notes. Terms in this table are defined in Appendix B. *Non-cluster-mean-centered models = models in which *at least one* lower-level (i.e., below level-L) predictor is not cluster-mean-centered. See Table 4 for corresponding table of measures for L-level cluster-mean-centered models (i.e., models in which *all* lower-level predictors are cluster-mean-centered). Also note that L is the integer denoting the highest level, M is >1 and $<L$, l can be an integer from 1 to L, and q can be an integer from 2 to L.

units, it can explain level-3 outcome variance. Similarly, if a level-2 predictor has mean variation across level-3 units, it can explain level-3 outcome variance. Hence, the single-source measures for three-level *non*-cluster-mean-centered models separately quantify variance explained by the level-1-varying *portion* of predictors (e.g., through $\text{var}_{i|jk}(x_{ijk})$) vs. the level-2-varying *portion* of predictors (e.g., through $\text{var}_{j|k}(x_{ijk})$) vs. the level-3-varying *portion* of predictors (e.g., through $\text{var}_k(x_{ijk})$); for associated derivations, see Appendix B. A second interpretational difference to note is that the definition of source m differs between the measures for cluster-mean-centered vs. non-cluster-mean-centered models. In the former, variance attributable to source m reflects the impact of cluster-specific outcome means via random intercept variation, whereas in the latter, variance attributable to source m reflects the impact of cluster-specific outcome means via random intercept variation *at the mean of all predictors with random slopes*, that is, the intercept variation that would be observed if all predictors with random slopes were centered by their grand mean (see Rights & Sterba, 2021, for supporting derivations and further detail). However, for either cluster-mean-centered or non-cluster-mean-centered models, variance attributable to source m can also be interpreted as the variance attributable to cluster-specific outcomes means above and beyond that accounted for by predictors (Rights & Sterba, 2019).

Researchers may also wonder how the obtained R^2 *values* would differ across alternative centering options. Researchers may be familiar with the fact that, in random slope models, the intercept mean and intercept variances (at level-2 and at any higher levels) and associated covariances will have different interpretations and will take on different values when changing how the predictors are centered, as the meaning of 0 for the predictors (and hence the meaning of the intercept) changes across alternative centering options (e.g., Biesanz et al., 2004). Nonetheless, when raw predictors have variance at only one level (which can occur, for instance, in longitudinal settings where time is balanced and thus does not differ across clusters),

Table 7. R^2 difference measures (ΔR^2) to compute effect sizes for individual terms in cluster-mean-centered and non-cluster-mean-centered* three-level models.

	For cluster-mean-centered models, compute R^2s using Table 1 formulas and then:		For non-cluster-mean-centered models*, compute R^2s using Table 5 formulas and then:	
To obtain the unique proportion of variance accounted for by this term …	**Interpret this total difference measure:**	**Interpret this level-specific difference measure:**	**Interpret these total difference measures:**	**Interpret these level-specific difference measures:**
Fixed component of level-1 predictor (including cross-level interaction product term involving level-1 and level-2/level-3 predictor †)	$\Delta R_t^{2(f_1)}$	$\Delta R_1^{2(f_1)}$	$\Delta R_t^{2(f_1)}$ $\Delta R_t^{2(f_2)}$ $\Delta R_t^{2(f_3)}$	$\Delta R_1^{2(f_1)}$ $\Delta R_2^{2(f_2)}$ $\Delta R_3^{2(f_3)}$
Fixed component of level-2 predictor (including cross-level interaction product term involving level-2 and level-3 predictor †)	$\Delta R_t^{2(f_2)}$	$\Delta R_2^{2(f_2)}$	$\Delta R_t^{2(f_2)}$ $\Delta R_t^{2(f_3)}$	$\Delta R_2^{2(f_2)}$ $\Delta R_3^{2(f_3)}$
Fixed component of level-3 predictor	$\Delta R_t^{2(f_3)}$	$\Delta R_3^{2(f_3)}$	$\Delta R_t^{2(f_3)}$	$\Delta R_3^{2(f_3)}$
Random component (across level-2 units) of level-1 predictor	$\Delta R_t^{2(v_{1*2})}$	$\Delta R_1^{2(v_{1*2})}$	$\Delta R_t^{2(v_{1*2})}$ $\Delta R_t^{2(v_{2*2})}$ $\Delta R_t^{2(v_{3*2})}$	$\Delta R_1^{2(v_{1*2})}$ $\Delta R_2^{2(v_{2*2})}$ $\Delta R_3^{2(v_{3*2})}$
Random component (across level-3 units) of level-1 predictor	$\Delta R_t^{2(v_{1*3})}$	$\Delta R_1^{2(v_{1*3})}$	$\Delta R_t^{2(v_{1*3})}$ $\Delta R_t^{2(v_{2*3})}$ $\Delta R_t^{2(v_{3*3})}$	$\Delta R_1^{2(v_{1*3})}$ $\Delta R_2^{2(v_{2*3})}$ $\Delta R_3^{2(v_{3*3})}$
Random component (across level-3 units) of level-2 predictor	$\Delta R_t^{2(v_{2*3})}$	$\Delta R_2^{2(v_{2*3})}$	$\Delta R_t^{2(v_{2*3})}$ $\Delta R_t^{2(v_{3*3})}$	$\Delta R_2^{2(v_{2*3})}$ $\Delta R_3^{2(v_{3*3})}$

Notes: Here we are assuming that the ΔR^2 is computed by subtracting the R^2 of a reduced model that excludes the term of interest from the R^2 from a full model that includes the term of interest (see Rights & Sterba, 2020, for more details).
*In cluster-mean-centered models (i.e., models in which *all* lower-level [here, level-1 and level-2] predictors are cluster-mean-centered), there is only one total and one level-specific ΔR^2 relevant for each individual term. For non-cluster-mean-centered models (i.e., models in which *at least one* lower-level predictor is *not* cluster-mean-centered), more ΔR^2 can be relevant for a given individual term because individual predictors can potentially explain variance at multiple levels.
†When adding cross-level interactions, these can lead to a decrease in $R_t^{2(v)}$, as the cross-level interaction can help explain across-cluster slope variability (Hoffman, 2015; Rights & Sterba, 2020).

the values of the R^2s in Table 4 computed from a cluster-mean-centered model will be equal to the values of the corresponding R^2s in Table 6 computed from a non-cluster-mean-centered model in which predictors are centered by any constant (see analytic proof in online Appendix C; additionally, for further detail and empirical demonstrations, see Rights & Sterba, 2021). Indeed, under this situation, cluster-mean-centered MLMs vs. centering-by-a-constant MLMs are likelihood-equivalent and yield the same model-implied total and level-specific outcome variances, as well as the same proportion of total and level-specific variance attributable to each individual source (f_1, f_2, v_1, and m). Even when raw predictors *do* have variance at multiple levels, so long as their level-specific effects do not differ, equivalent R^2 results in the population will be obtained when either computing the Table 4 measures from a fitted cluster-mean-centered model (assuming all cluster means of predictors are also included) or when computing Table 6 measures from a fitted *non*-cluster-mean-centered model using centering-at-a-constant. On the other hand, when raw predictors *do* have variance at multiple levels *and* their level-specific effects differ, results of Table 6 measures computed from a fitted *non*-cluster-mean-centered model will be distorted as compared to results of Table 4 measures computed from a (properly disaggregated) cluster-mean-centered model (for demonstrations and further detail see Rights, 2021).

Software implementation

To facilitate implementation of R^2 measures for three-level MLMs, we developed an R function, *r2mlm3*, that reads in three-level MLM parameter estimates and raw data and automatically outputs measures from Table 1 for cluster-mean-centered three-level MLMs (i.e., models in which all lower-level—here, level-1 or level-2—predictors are cluster-mean-centered), or measures from Table 5 for *non*-cluster-mean-centered three-level MLMs (i.e., models in which at least one lower-level predictor is not cluster-mean-centered). Additionally, this *r2mlm3* function produces barcharts that provide a visual representation of each measure, e.g., as shown in Figure 1. Note that this function can also accommodate two-level models,[13] wherein each measure quantifying variance explained via level-3 sources would, by definition, be 0. See online Appendix D for the R function syntax and descriptions of each input. This function has additionally be added to the *r2mlm* package (Shaw, Rights, Sterba, & Flake, 2020), which can be installed from CRAN, and also contains functions from Rights and Sterba (2019) and Rights and Sterba (2020). Also see online Appendix E for a step-by-step walkthrough for how one can obtain our empirical example dataset, read it into R, cluster-mean-center variables (if desired) in R, fit our empirical example model using R (with the *lmer* function), and use our *r2mlm3* R function to read in *lmer* output and obtain all R^2 measures. online Appendix E also displays the *lmer* output from model fitting and the R^2 measures and plot output from *r2mlm3*. A future direction involves extending this R function to handle L-level MLMs; until then, researchers with models having more than three levels can always compute measures manually using the formulas we provide in Tables 4 and 6.

Discussion

Although researchers increasingly need to model hierarchical data structures beyond two levels, options for R^2 measures of effect size for MLMs beyond two levels have been limited. Hence, here we generalized a framework of total and level-specific MLM R^2s (Rights & Sterba, 2019) to three or more levels for both cluster-mean-centered MLMs (Table 4) and non-cluster-mean-centered MLMs (Table 6), and clarified the impact of centering strategy on R^2 interpretation and computation. Measures in our framework are obtained from fitting a single MLM that can include random slopes. By employing a more general decomposition of model-implied variance, our framework provides a more comprehensive and flexible set of measures—subsuming previous three-level measures as special cases (see Table 2 and Online Appendix A) and yielding substantively relevant results not afforded by previous measures (see Tables 1 and 4).

[13]Furthermore, this function accommodates two-level models more generally than did software provided in Rights & Sterba (2019), as now all total *and* level-specific measures are provided for both cluster-mean-centered and non-cluster-mean centered models, whereas two-level software provided in Rights & Sterba (2019) included total and level-specific measures for cluster-mean-centered models but only total measures for non-cluster-mean-centered models.

We also newly derived and illustrated analytic relations between total and level-specific R^2 for given source(s) of explained variance (see Eqn. 26-28 and Online Appendix B). Last, we provided software to compute and graphically portray the framework of measures (see, e.g., *r2mlm* R package or *r2mlm3* R function).

Recommendations for practice

In practice, we encourage researchers to report the set of single-source measures (e.g., Table 1 for three levels or Table 4 for L levels), visualize them in juxtaposition using a barchart (e.g., Figure 1), and interpret each as an effect size. This allows assessing the relative importance of each individual source of explained variance and the relative importance of each level of the hierarchical data structure. As an optional, supplemental summary, combination-source measures can be computed by combining total measures of substantive focus and/or by combining level-l specific measures of substantive focus (where $l < L$).

Utility of framework for model comparison

In the current paper, we focused on computing R^2 measures for one fitted model. This framework can additionally be implemented when comparing/building models and computing R^2 differences (ΔR^2). However, there had been persistent confusion about *which* ΔR^2 to compute and interpret when comparing MLMs to assess the unique impact of a particular term(s) (see historical review and illustrations in Rights & Sterba, 2020). To address these misunderstandings, for two-level MLMs, Rights and Sterba (2020) defined which *target* single-source ΔR^2 to use to detect the unique contribution of each kind of term that could be added to the reduced MLM to form the full MLM. Extending these definitions for three-level MLMs, Table 7 provides the set of single-source target ΔR^2 appropriate to quantify the effect size for individual terms in both three-level cluster-mean-centered MLMs—in which a single added term can explain variance at only one level—and non-cluster-mean-centered MLMs—in which a single added term can explain variance at multiple levels (for cautions on using *combination-source* ΔR^2 measures, see Rights & Sterba, 2020). These single-source target ΔR^2 can be implemented either with a *hierarchical model-building approach*—starting with a baseline model, adding terms sequentially, and computing ΔR^2 at each step—or a *simultaneous approach*—always comparing each reduced model to the most complex model when computing ΔR^2 (Rights & Sterba, 2020).

Generally speaking, a primary reason to compute ΔR^2 using model comparisons is to isolate the unique contribution of a term when there are multiple terms with its same source of explained variance in the fitted model. Visually, this corresponds to subdividing a particular shaded segment in Figure 1 barchart to identify term-specific contributions. Several examples follow. As a first example, suppose a researcher were interested in an expanded version of our empirical model that also included as control variables teacher-means of each level-1 predictor (kindergarten math score, sex, and SES). All five teacher-level variables taken together account for a sizable portion of outcome variance ($\hat{R}_t^{2(f_2)}$=.18). However, suppose the researcher wanted to use a *simultaneous approach* to isolate variance uniquely attributable to each of the two new and three original teacher-level predictors via their fixed components. Computing *unique* contributions of each teacher-level predictor with the target measure $\Delta R_t^{2(f_2)}$ (see Table 7) using a simultaneous approach indicates that, consistent with our original R^2 results, the teacher-level predictors of substantive interest uniquely explain little outcome variance (for math preparation, $\Delta\hat{R}_t^{2(f_2)} < .01$, and for math knowledge, $\Delta\hat{R}_t^{2(f_2)} = .01$), whereas the control variables provide dominant contributions (in particular $\Delta\hat{R}_t^{2(f_2)} = .16$ for teacher-mean kindergarten math score; other teacher-level variables each individually yielding $\Delta\hat{R}_t^{2(f_2)} < .01$).

As another illustration, model comparison could be used to evaluate the unique contribution of a cross-level interaction to our empirical example model. For instance, suppose we wanted to add a cross-level interaction of teacher math knowledge × student kindergarten math score to our original empirical example MLM and evaluate its impact using a hierarchical model-building approach. This product term itself is a level-1 predictor (as it varies exclusively within-cluster given the cluster-mean-centering of original variables described above; Rights & Sterba, 2020). Thus, as shown in Table 7, we could quantify the total variance explained by the product term using the target measure $\Delta R_t^{2(f_1)}$ (and quantify the level-specific variance explained using the target measure $\Delta R_1^{2(f_1)}$). In this particular empirical example, the estimated fixed component of the product term was nearly zero, was non-significant, and did not lead to changes in either target ΔR^2 measure.[14]

[14]If the addition of this product term did in fact lead to an increase in target measures $\Delta R_t^{2(f_1)}$ and $\Delta R_1^{2(f_1)}$, we would also expect to see a *decrease* in the variance attributable to random slope variability (i.e., $R_t^{2(v_{1*2})}$ and $R_1^{2(v_{1*2})}$), as the random slope variance in the reduced model would be instead accounted for by the fixed component of the cross-level product term in the full model (Hoffman, 2015; Raudenbush & Bryk, 2002; Rights & Sterba, 2020).

Model comparisons can also be useful for illuminating the unique contribution of each random slope in MLMs that contain multiple random slopes. In three-level contexts, adding a random slope of a level-1 predictor varying across level-2 units (when this term is indeed present in the generating population) would lead to an expected increase in variance attributable to source v_{1*2} (detected with target measures $\Delta R_t^{2(v_{1*2})}$ and $\Delta R_1^{2(v_{1*2})}$ from Table 7), which is necessarily accompanied by a decrease in variance attributable to level-1 residuals. Similarly, correctly adding a random slope of a level-1 predictor varying across level-3 units would lead to an expected increase in variance attributable to source v_{1*3} (detected with target measures $\Delta R_t^{2(v_{1*3})}$ and $\Delta R_1^{2(v_{1*3})}$), again necessarily accompanied by a decrease in variance attributable to level-1 residuals. Lastly, correctly adding a random slope of a level-2 predictor (varying across level-3 units) leads to an expected increase in variance attributable to source v_{2*3} (detected using target measures $\Delta R_t^{2(v_{2*3})}$ and $\Delta R_2^{2(v_{2*3})}$), which would be accompanied by a decrease in the variance attributable to source m_2.

Limitations and future directions

One avenue for future work involves investigating the finite sample performance of these R^2 measures under a variety of generating conditions mirroring applied practice in different disciplines in which MLMs are applied. These measures are computed as a function of (a) the model parameter estimates (which, for instance, can be obtained as maximum likelihood estimates or as Bayesian posterior means/medians) as well as (b) the estimated variances and covariances of the predictors (which can be obtained using the sample variances and covariances, regardless of the distribution of the predictors, as done in the *r2mlm3* R function and the *r2mlm* R package). Hence, estimation of both model parameters and predictor variances/covariances should be accurate to ensure accurate estimation of R^2 measures. It would be useful to investigate, for instance, the impact of the number of clusters and cluster size, at level-2 and level-3, as well as the impact of the number of predictors.

Though here we addressed computation of R^2 for models with any number of levels, there are still additional modeling options in MLM that can further complicate computation of R^2 measures. For instance, in cross-classified MLMs, students can be simultaneously nested within both schools and neighborhoods, but with neither schools nested within neighborhood nor vice versa. In such a model, variance can be attributable to student-level, school-level, and neighborhood-level sources. R^2 measures can be developed to quantify variance explained by these different sources, but the computation done in the current paper (e.g., Appendix A would need to be modified to account for the non-hierarchically nested nature of the clustering. Though an R^2 has been developed specifically for cross-classified MLM (Luo & Kwok, 2010), this measure (an expanded version of $R^2_{S\&B}$ but for two-level cross-classified models) shares some of the same limitations noted earlier for three-level measures (e.g., only assesses variance explained by predictors via fixed effects and does not separately consider contribution of predictors at each level, etc.). Another example of an MLM specification that would necessitate further work to extend our framework would be multiple membership models (e.g., Goldstein, 2011), wherein observations can be nested within multiple clusters simultaneously. A third example would be partially nested MLMs, where observations are clustered for some study arms, but not other study arms (e.g., Sterba, 2017). A final example would be generalized linear mixed effects models (GLMMs). Our framework could be adapted for GLMMs (with, e.g., binary or count outcomes) using procedures similar to Nakagawa & Schielzeth (2013) and Johnson (2014) in which the residual variance in the model-implied variance expression (see Equation (9)) is modified to accommodate the particular error distribution and link function used.

Conclusions

In providing this extended framework of R^2 measures to accommodate any number of levels, it is our hope that R^2 effect sizes will more routinely be reported when fitting MLMs beyond two levels. A suite of interpretable measures can now easily be computed and graphically visualized with formulae and freely available software provided here, and researchers can reference our framework to help decide which measures are most important to consider given substantive research questions. These measures facilitate considering practical significance, rather than just statistical significance, in complex designs with multiple levels and many potential sources of explained variance to consider simultaneously.

Article information

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

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

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

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

Acknowledgments: The authors would like to thank Kristopher Preacher, Sun-Joo Cho, and Andrew Tomarken for their helpful comments. The ideas and opinions expressed herein are those of the authors alone, and endorsement by the authors' institutions or the Natural Sciences and Engineering Research Council of Canada is not intended and should not be inferred.

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*(2), 155–176. https://doi.org/10.1177/1094428114563618

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

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. *R Package Version*, *1*, 1–23.

Bickel, R. (2007). *Multilevel analysis for applied research. It's just regression*. Guilford.

Biesanz, J. C., Deeb-Sossa, N., Papadakis, A. A., Bollen, K. A., & Curran, J. C. (2004). The role of coding time in estimating and interpreting growth curve models. *Psychological Methods*, *9*(1), 30–52. https://doi.org/10.1037/1082-989X.9.1.30

Borko, H., Eisenhart, M., Brown, C. A., Underhill, R. G., Jones, D., & Agard, P. C. (1992). Learning to teach hard mathematics: Do novice teachers and their instructors give up too easily? *Journal for Research in Mathematics Education*, *23*(3), 194–222. https://doi.org/10.2307/749118

Brincks, A. M., Enders, C. K., Llabre, M. M., Bulotsky-Shearer, R. J., Prado, G., & Feaster, D. J. (2017).

Centering predictor variables in three-level contextual models. *Multivariate Behavioral Research*, *52*(2), 149–163.
Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods*. Sage.
Carpenter, T. P., Fennema, E., Peterson, P. L., Chiang, C.-P., & Loef, M. (1989). Using knowledge of children's mathematics thinking in classroom teaching: An experimental study. *American Educational Research Journal*, *26*(4), 499–531. https://doi.org/10.3102/00028312026004499
Chen, Z., Zhu, J., & Zhou, M. (2015). How does a servant leader fuel the service fire? A multilevel model of servant leadership, individual self identity, group competition climate, and customer service performance. *The Journal of Applied Psychology*, *100*(2), 511–561. https://doi.org/10.1037/a0038036
Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Erlbaum.
Curran, P. J., & Bauer, D. J. (2011). The disaggregation of within-person and between-person effects in longitudinal models of change. *Annual Review of Psychology*, *62*, 583–619. https://doi.org/10.1146/annurev.psych.093008.100356
Curran, P. J., Lee, T., Howard, A. L., Lane, S., & MacCallum, R. A. (2012). Disaggregating within-person and between-person effects in multilevel and structural equation growth models. In J. R. Harring & G. R. Hancock (Eds.), *Advances in longitudinal methods in the social and behavioral sciences.* (pp. 217–253). Information Age.
Curran, P. J., McGinley, J. S., Serrano, D., & Burfeind, C. (2012). A multivariate growth curve model for three-level data. In H. Cooper (Ed.). *APA Handbook of Research Methods in Psychology* (Vol. 3, pp. 335–358). American Psychological Association.
Dollard, M. F., Opie, T., Lenthall, S., Wakerman, J., Knight, S., Dunn, S., Rickard, G., & MacLeod, M. (2012). Psychosocial safety climate as an antecedent of work characteristics and psychological strain: A multilevel model. *Work & Stress*, *26*(4), 385–404. https://doi.org/10.1080/02678373.2012.734154
Edwards, L. J., Muller, K. E., Wolfinger, R. D., Qaqish, B. F., & Schabenberger, O. (2008). An R^2 statistic for fixed effects in the linear mixed model. *Statistics in Medicine*, *27*(29), 6137–6157.
Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods*, *12*(2), 121–138. https://doi.org/10.1037/1082-989X.12.2.121
Gelman, A., & Pardoe, I. (2006). Bayesian measures of explained variance and pooling in multilevel (hierarchical) models. *Technometrics*, *48*(2), 241–251. https://doi.org/10.1198/004017005000000517
Goldstein, H. (2011). *Multilevel statistical models.* (4th ed.). Wiley.
Gong, Y., Kim, T. Y., Lee, D. R., & Zhu, J. (2013). A multilevel model of team goal orientation, information exchange, and creativity. *Academy of Management Journal*, *56*(3), 827–851. https://doi.org/10.5465/amj.2011.0177
Gurka, M. J., Edwards, L. J., & Muller, K. E. (2011). Avoiding bias in mixed model inference for fixed effects. *Statistics in Medicine*, *30*(22), 2696–2707.
Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal*, *42*(2), 371–406. https://doi.org/10.3102/00028312042002371
Hoffman, L. (2015). *Longitudinal analysis: Modeling within-person fluctuation and change*. Routledge.
Hofmann, D. A., & Gavin, M. B. (1998). Centering decisions in hierarchical linear models: Implications for research in organizations. *Journal of Management*, *24*(5), 623–641. https://doi.org/10.1177/014920639802400504
Hox, J. J. (2010). *Multilevel analysis: Techniques and applications*. 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*(6), 1086–1105. https://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.
Ke, Z., & Wang, L. (2015). Detecting individual differences in change: Methods and comparisons. *Structural Equation Modeling: A Multidisciplinary Journal*, *22*(3), 382–400. https://doi.org/10.1080/10705511.2014.936096
Kramer, M. (2005). *R^2 statistics for mixed models* [Paper presentation]. 2005 Proceedings of the Conference on Applied Statistics in Agriculture, (pp. 148–160). Kansas State University: Manhattan, KS.
LaHuis, D. M., Hartman, M. J., Hakoyama, S., & Clark, P. C. (2014). Explained variance measures for multilevel models. *Organizational Research Methods*, *17*(4), 433–451. https://doi.org/10.1177/1094428114541701
Leinhardt, G., & Smith, D. A. (1985). Expertise in mathematics instruction: Subject matter knowledge. *Journal of Educational Psychology*, *77*(3), 247–271. https://doi.org/10.1037/0022-0663.77.3.247
Liu, D., Liao, H., & Loi, R. (2012). The dark side of leadership: A three-level investigation of the cascading effect of abusive supervision on employee creativity. *Academy of Management Journal*, *55*(5), 1187–1212. https://doi.org/10.5465/amj.2010.0400
Lorah, J., (2018). Effect size measures for multilevel models: Definition, interpretation, and TIMSS example. *Large-Scale Assessments in Education*, 1. 1–11.
Luo, W., & Kwok, O. M. (2010). Proportional reduction of prediction error in cross-classified random effects models. *Sociological Methods & Research*, *39*(2), 188–205. https://doi.org/10.1177/0049124110384062
Ma, L. (1999). *Knowing and teaching elementary mathematics: Teachers' understanding of fundamental mathematics in China and the United States*. Erlbaum.
Maier, M. F., Vitiello, V. E., & Greenfield, D. B. (2012). A multilevel model of child-and classroom-level psychosocial factors that support language and literacy resilience of children in Head Start. *Early Childhood Research Quarterly*, *27*(1), 104–114. https://doi.org/10.1016/j.ecresq.2011.06.002
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.
Orelien, J. G., & Edwards, L. J. (2008). Fixed-effect variable selection in linear mixed models using R^2 statistics. *Computational Statistics & Data Analysis*, *52*(4), 1896–1907. https://doi.org/10.1016/j.csda.2007.06.006

Peugh, J. L., & Heck, R. H. (2017). Conducting three-level longitudinal analyses. *The Journal of Early Adolescence*, *37*(1), 7–58. https://doi.org/10.1177/0272431616642329

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

Rabe-Hesketh, S., & Skrondal, A. (2008). *Multilevel and Longitudinal Modeling Using Stata. College Station*. TX: Stata Press.

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

Recchia, A. (2010). R-squared measures for two-level hierarchical linear models using SAS. *Journal of Statistical Software*, *32* (Code Snippet 2), 1–9. https://doi.org/10.18637/jss.v032.c02

Rights, J. D. (2021). Aberrant distortion of variance components in multilevel models under conflation of level-specific effects.

Rights, J. D., Preacher, K. J., & Cole, D. A. (2020). The danger of conflating level-specific effects of control variables when primary interest lies in level-2 effects. *British Journal of Mathematical and Statistical Psychology*, *73*(S1), 194–211. https://doi.org/10.1111/bmsp.12194

Rights, J. D., & Sterba, S. K. (2016). The relationship between multilevel models and non-parametric multilevel mixture models: Discrete approximation of intraclass correlation, random coefficient distributions, and residual heteroscedasticity. *British Journal of Mathematical and Statistical Psychology*, *69*(3), 316–343. https://doi.org/10.1111/bmsp.12073

Rights, J. D., & Sterba, S. K. (2018). A framework of R-squared measures for single-level and multilevel regression mixture models. *Psychological Methods*, *23*(3), 434–457. https://doi.org/10.1037/met0000139

Rights, J. D., & Sterba, S. K. (2019). Quantifying explained variance in multilevel models: An integrative framework for defining R-squared measures. *Psychological Methods*, *24*(3), 309–338.

Rights, J. D., & Sterba, S. K. (2020). New recommendations on the use of R-squared differences in multilevel model comparisons. *Multivariate Behavioral Research*, *55*(4), 568–599. https://doi.org/10.1080/00273171.2019.1660605

Rights, J. D., & Sterba, S. K. (2021). Effect size measures for longitudinal growth analyses: Extending a framework of multilevel model R-squareds to accommodate heteroscedasticity, autocorrelation, nonlinearity, and alternative centering strategies. *New Directions for Child and Adolescent Development*, *2021*(175), 65–110.

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

Shaw, M., Rights, J. D., Sterba, S. K., & Flake, J. K. (2020). *r2mlm: R-Squared Measures for Multilevel Models*, https://doi.org/10.31234/osf.io/xc4sv

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. Oxford Univ. Press.

Snijders, T. A. B., & Bosker, R. J. (1994). Modeled variance in two-level models. *Sociological Methods & Research*, *22*(3), 342–363. https://doi.org/10.1177/0049124194022003004

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

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

Sterba, S. K. (2017). Partially nested designs in psychotherapy trials: A review of modeling developments. *Psychotherapy Research: Journal of the Society for Psychotherapy Research*, *27*(4), 425–436. https://doi.org/10.1080/10503307.2015.1114688

Stram, D., & Lee, J. W. (1994). Variance components testing in the longitudinal mixed effects model. *Biometrics*, *50*(4), 1171–1177.

Thompson, P., & Thompson, A. (1994). Talking about rates conceptually, Part I: A teacher's struggle. *Journal for Research in Mathematics Education*, *25*(3), 279–303. https://doi.org/10.2307/749339

Van den Noortgate, W., López-López, J. A., Marín-Martínez, F., & Sánchez-Meca, J. (2013). Three-level meta-analysis of dependent effect sizes. *Behavior Research Methods*, *45*(2), 576–594. https://doi.org/10.3758/s13428-012-0261-6

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

Wang, J., & Schaalje, G. B. (2009). Model selection for linear mixed models using predictive criteria. *Communications in Statistics - Simulation and Computation*, *38*(4), 788–801. https://doi.org/10.1080/03610910802645362

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

Yaremych, H., Preacher, K. J., & Hedeker, D. (2020). *Centering categorical predictors in multilevel models* [Paper presentation]. Poster Presented at the Association for Psychological Science Conference (Online).

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

Appendix A:

Derivation of model-implied outcome variance for cluster-mean-centered models with any number of levels

Here we compute the model-implied total and level-specific outcome variance of a cluster-mean-centered multilevel model (i.e., a model in which *all* lower-level [i.e., below level-L] predictors are cluster-mean-centered) with any number of levels. A generic L-level model with cluster-mean-centered predictors can be written as:

$$y_{ic} = \gamma_0 + \sum_{l=1}^{L} \mathbf{x}'_{lic}\boldsymbol{\gamma}_l + \sum_{l=1}^{L-1}\sum_{q=l+1}^{L} \mathbf{w}'_{l*qic}\mathbf{u}_{l*qc} + e_{ic} \tag{A1}$$

with γ_0 denoting the fixed intercept, l denoting level ($l = 1, \ldots, L$), i denoting observation within level-2 unit, and

c denoting the set of observation i's cluster memberships for all levels greater than 1 (e.g., level-2, -3, and -4 cluster membership for a four-level model). Each $\mathbf{x}'_{lic}$ denotes a vector of level-l predictors with $\boldsymbol{\gamma}_l$ denoting a vector of the level-l fixed components of slopes. Each $\mathbf{w}'_{l*qi\mathbf{c}}$ denotes a vector containing level-l predictors with random effects varying across level-q units ($1 < q \leq L$); $\mathbf{u}_{l*q\mathbf{c}}$ denotes a vector of corresponding residuals. For all combinations of l and q such that $l = 1$, the first element of $\mathbf{w}'_{l*qi\mathbf{c}}$ is 1 and the corresponding element of $\mathbf{u}_{l*q\mathbf{c}}$ is the intercept residual for level-q (to ensure each level of clustering has a single corresponding intercept residual). The last term, $e_{i\mathbf{c}}$, denotes the level-1 residual.

The variance of this expression is given as

$$\begin{aligned}\text{var}(y_{i\mathbf{c}}) &= \text{var}(\gamma_0 + \sum_{l=1}^{L}\mathbf{x}'_{li\mathbf{c}}\boldsymbol{\gamma}_l + \sum_{l=1}^{L-1}\sum_{q=l+1}^{L}\mathbf{w}'_{l*qi\mathbf{c}}\mathbf{u}_{l*q\mathbf{c}} + e_{i\mathbf{c}})\\ &= \text{var}(\sum_{l=1}^{L}\mathbf{x}'_{li\mathbf{c}}\boldsymbol{\gamma}_l) + \text{var}(\sum_{l=1}^{L-1}\sum_{q=l+1}^{L}\mathbf{w}'_{l*qi\mathbf{c}}\mathbf{u}_{l*q\mathbf{c}}) + \text{var}(e_{i\mathbf{c}})\\ &= \sum_{l=1}^{L}\text{var}(\mathbf{x}'_{li\mathbf{c}}\boldsymbol{\gamma}_l) + \sum_{l=1}^{L-1}\sum_{q=l+1}^{L}\text{var}(\mathbf{w}'_{l*qi\mathbf{c}}\mathbf{u}_{l*qi\mathbf{c}}) + \text{var}(e_{i\mathbf{c}})\end{aligned} \tag{A2}$$

These steps in Equation (A2) hold because the following pairs of terms are uncorrelated with each other: 1) any pair of predictors across different levels (e.g., a level-1 predictor and a level-2 predictor) given the centering described in the manuscript; 2) the fixed effects and random effect residuals; and 3) any pair of residuals across different levels (e.g., a level-2 intercept residual and a level-3 intercept residual).

The first part of Equation (A2) is:

$$\begin{aligned}\sum_{l=1}^{L}\text{var}(\mathbf{x}'_{li\mathbf{c}}\boldsymbol{\gamma}_l) &= \sum_{l=1}^{L}\boldsymbol{\gamma}'_l\text{var}(\mathbf{x}'_{li\mathbf{c}})\boldsymbol{\gamma}_l\\ &= \sum_{l=1}^{L}\boldsymbol{\gamma}'_l\boldsymbol{\Phi}_l\boldsymbol{\gamma}_l\end{aligned} \tag{A3}$$

For the second part of Equation (A2), we need to consider what it equals under two cases: $l = 1$ and $l > 1$ (because, for reasons explained above, $\mathbf{w}_{l*qi\mathbf{c}}$ contains 1 as the first element when $l = 1$ and contains only predictors when $l > 1$). Using the law of total variance, when $l = 1$:

$$\begin{aligned}\text{var}(\mathbf{w}'_{l*qi\mathbf{c}}\mathbf{u}_{l*q\mathbf{c}}) &= E\left[\text{var}(\mathbf{w}'_{l*qi\mathbf{c}}\mathbf{u}_{l*q\mathbf{c}}|\mathbf{u}_{l*q\mathbf{c}})\right]\\ &\quad+\text{var}(E\left[\mathbf{w}'_{l*qi\mathbf{c}}\mathbf{u}_{l*q\mathbf{c}}|\mathbf{u}_{l*q\mathbf{c}}\right])\\ &= E\left[\mathbf{u}'_{l*q\mathbf{c}}\boldsymbol{\Sigma}_{l*q}\mathbf{u}_{l*q\mathbf{c}}\right] + \text{var}(E\left[\mathbf{w}'_{l*qi\mathbf{c}}\right]\mathbf{u}_{l*q\mathbf{c}})\end{aligned} \tag{A4}$$

For cluster-mean-centered models, all predictors with random slopes have means of 0, and hence $\text{var}(E[\mathbf{w}'_{l*qi\mathbf{c}}]\mathbf{u}_{l*q\mathbf{c}})$ simplifies to $\text{var}(u_{q\mathbf{c}})$, with $u_{q\mathbf{c}}$ denoting the intercept random effect residual for level-q. Thus,

$$\begin{aligned}\text{var}(\mathbf{w}'_{l*qi\mathbf{c}}\mathbf{u}_{l*q\mathbf{c}}) &= E\left[\mathbf{u}'_{l*q\mathbf{c}}\boldsymbol{\Sigma}_{l*q}\mathbf{u}_{l*q\mathbf{c}}\right] + \text{var}(u_{q\mathbf{c}})\\ &= E\left[tr(\mathbf{u}'_{l*q\mathbf{c}}\boldsymbol{\Sigma}_{l*q}\mathbf{u}_{l*q\mathbf{c}})\right] + \tau_q\\ &= E\left[tr(\boldsymbol{\Sigma}_{l*q}\mathbf{u}_{l*q\mathbf{c}}\mathbf{u}'_{l*q\mathbf{c}})\right] + \tau_q\\ &= tr(\boldsymbol{\Sigma}_{l*q}E\left[\mathbf{u}_{l*q\mathbf{c}}\mathbf{u}'_{l*q\mathbf{c}}\right]) + \tau_q\\ &= tr(\boldsymbol{\Sigma}_{l*q}\mathrm{T}_{l*q}) + \tau_q\end{aligned} \tag{A5}$$

Again using the law of total variance, when $l > 1$ (and thus $\mathbf{w}'_{l*qi\mathbf{c}}$ contains only predictors):

$$\begin{aligned}\text{var}(\mathbf{w}'_{l*qi\mathbf{c}}\mathbf{u}_{l*q\mathbf{c}}) &= E\left[\text{var}(\mathbf{w}'_{l*qi\mathbf{c}}\mathbf{u}_{l*q\mathbf{c}}|\mathbf{u}_{l*q\mathbf{c}})\right] + \text{var}(E\left[\mathbf{w}'_{l*qi\mathbf{c}}\mathbf{u}_{l*q\mathbf{c}}|\mathbf{u}_{l*q\mathbf{c}}\right])\\ &= E\left[\mathbf{u}'_{l*q\mathbf{c}}\boldsymbol{\Sigma}_{l*q}\mathbf{u}_{l*q\mathbf{c}}\right] + \text{var}(E\left[\mathbf{w}'_{l*qi\mathbf{c}}\right]\mathbf{u}_{l*q\mathbf{c}})\\ &= E\left[\mathbf{u}'_{l*q\mathbf{c}}\boldsymbol{\Sigma}_{l*q}\mathbf{u}_{l*q\mathbf{c}}\right]\\ &= E\left[tr(\mathbf{u}'_{l*q\mathbf{c}}\boldsymbol{\Sigma}_{l*q}\mathbf{u}_{l*q\mathbf{c}})\right]\\ &= E\left[tr(\boldsymbol{\Sigma}_{l*q}\mathbf{u}_{l*q\mathbf{c}}\mathbf{u}'_{l*q\mathbf{c}})\right]\\ &= tr(\boldsymbol{\Sigma}_{l*q}E\left[\mathbf{u}_{l*q\mathbf{c}}\mathbf{u}'_{l*q\mathbf{c}}\right])\\ &= tr(\boldsymbol{\Sigma}_{l*q}\mathrm{T}_{l*q})\end{aligned} \tag{A6}$$

When computing $\sum_{l=1}^{L-1}\sum_{q=l+1}^{L}\text{var}(\mathbf{w}'_{l*qi\mathbf{c}}\mathbf{u}_{l*q\mathbf{c}})$, we sum the $tr(\boldsymbol{\Sigma}_{l*q}\mathrm{T}_{l*q})$ across all combinations of l and q (such that $l < L$, as we assume predictors at the highest level do not have random slopes, and such that $q > l$) because the $tr(\boldsymbol{\Sigma}_{l*q}\mathrm{T}_{l*q})$ term is included for all such combinations of l and q (see Equations A5 and A6). However, with L levels there are only L – 1 combinations of $l = 1$ and $q > l$, starting with $q = 2$ up to $q = L$. Thus, when computing $\sum_{l=1}^{L-1}\sum_{q=l+1}^{L}\text{var}(\mathbf{w}'_{l*qi\mathbf{c}}\mathbf{u}_{l*q\mathbf{c}})$ we additionally sum τ_q (included in Equation A5 but not A6) from $q = 2$ to $q = L$ (or, equivalently, sum all τ_l from $l = 2$ to $l = L$), i.e.,

$$\begin{aligned}\sum_{l=1}^{L-1}\sum_{q=l+1}^{L}\text{var}(\mathbf{w}'_{l*qi\mathbf{c}}\mathbf{u}_{l*q\mathbf{c}}) &= \Big(\sum_{l=1}^{L-1}\sum_{q=l+1}^{L}tr(\boldsymbol{\Sigma}_{l*q}\mathrm{T}_{l*q})\Big)\\ &\quad+ (\sum_{l=2}^{L}\tau_l)\end{aligned} \tag{A7}$$

The third part in Equation (A2) is

$$\text{var}(e_{i\mathbf{c}}) = \sigma^2 \tag{A8}$$

Thus the model-implied outcome variance for the MLM data model with any number of levels is:

$$\begin{aligned}\text{var}(y_{i\mathbf{c}}) &= (\sum_{l=1}^{L}\boldsymbol{\gamma}'_l\boldsymbol{\Phi}_l\boldsymbol{\gamma}_l) + \Big(\sum_{l=1}^{L-1}\sum_{q=l+1}^{L}tr(\boldsymbol{\Sigma}_{l*q}\mathrm{T}_{l*q})\Big)\\ &\quad+ (\sum_{l=2}^{L}\tau_l) + \sigma^2\end{aligned} \tag{A9}$$

This model-implied total outcome variance, which serves as the denominator for the total R^2 measures, can be broken down into level-specific parts to form the denominators for level-specific R^2 measures. With **c** denoting the set of observation i's cluster memberships for all levels greater than 1 (e.g., level-2, -3, and -4 cluster membership for a four-level model), $\text{var}_{i|\mathbf{c}}(.)$ then represents the strictly level-1 variance, and hence we can compute the level-1 outcome variance as

$$\begin{aligned}\text{var}_{i|\mathbf{c}}(y_{i\mathbf{c}}) &= \text{var}_{i|\mathbf{c}}(\gamma_0 + \sum_{l=1}^{L}\mathbf{x}'_{li\mathbf{c}}\boldsymbol{\gamma}_l + \sum_{l=1}^{L-1}\sum_{q=l+1}^{L}\mathbf{w}'_{l*qi\mathbf{c}}\mathbf{u}_{l*q\mathbf{c}} + e_{i\mathbf{c}})\\ &= \text{var}_{i|\mathbf{c}}(\sum_{l=1}^{L}\mathbf{x}'_{li\mathbf{c}}\boldsymbol{\gamma}_l) + \text{var}_{i|\mathbf{c}}(\sum_{l=1}^{L-1}\sum_{q=l+1}^{L}\mathbf{w}'_{l*qi\mathbf{c}}\mathbf{u}_{l*q\mathbf{c}}) + \text{var}_{i|\mathbf{c}}(e_{i\mathbf{c}})\\ &= \text{var}_{i|\mathbf{c}}(\mathbf{x}'_{1i\mathbf{c}}\boldsymbol{\gamma}_1) + \text{var}_{i|\mathbf{c}}(\sum_{q=2}^{L}\mathbf{w}'_{1*qi\mathbf{c}}\mathbf{u}_{1*q\mathbf{c}}) + \text{var}_{i|\mathbf{c}}(e_{i\mathbf{c}})\\ &= \text{var}_{i|\mathbf{c}}(\mathbf{x}'_{1i\mathbf{c}}\boldsymbol{\gamma}_1) + \sum_{q=2}^{L}\text{var}_{i|\mathbf{c}}(\mathbf{w}'_{1*qi\mathbf{c}}\mathbf{u}_{1*q\mathbf{c}}) + \text{var}_{i|\mathbf{c}}(e_{i\mathbf{c}})\\ &= \boldsymbol{\gamma}'_1\boldsymbol{\Phi}_1\boldsymbol{\gamma}_1 + \sum_{q=2}^{L}tr(\boldsymbol{\Sigma}_{1*q}\mathrm{T}_{1*q}) + \sigma^2\end{aligned} \tag{A10}$$

The steps in this derivation are identical to those shown above in Equations A2-A9, with the exception that certain terms are dropped if they are constant across level-1 units within level-2 units (i.e., have no variance at level-1).

Next, letting j denote level-M cluster membership, with M denoting an intermediate level (i.e., $1 < M < L$), and letting $\mathbf{c}_{>M}$ denote the set of cluster j's higher-level cluster memberships (i.e., cluster membership at levels greater than M), we can compute the level-M outcome variance as

$$\begin{aligned}
\text{var}_{j|\mathbf{c}_{>M}}(y_{i\mathbf{c}}) &= \text{var}_{j|\mathbf{c}_{>M}}(\gamma_0 + \sum_{l=1}^{L}\mathbf{x}'_{li\mathbf{c}}\boldsymbol{\gamma}_l + \sum_{l=1}^{L-1}\sum_{q=l+1}^{L}\mathbf{w}'_{l*qi\mathbf{c}}\mathbf{u}_{l*q\mathbf{c}} + e_{i\mathbf{c}}) \\
&= \text{var}_{j|\mathbf{c}_{>M}}(\sum_{l=1}^{L}\mathbf{x}'_{li\mathbf{c}}\boldsymbol{\gamma}_l) + \text{var}_{j|\mathbf{c}_{>M}}(\sum_{l=1}^{L-1}\sum_{q=l+1}^{L}\mathbf{w}'_{l*qi\mathbf{c}}\mathbf{u}_{l*q\mathbf{c}}) + \text{var}_{j|\mathbf{c}_{>M}}(e_{i\mathbf{c}}) \\
&= \text{var}_{j|\mathbf{c}_{>M}}(\mathbf{x}'_{Mi\mathbf{c}}\boldsymbol{\gamma}_M) + \text{var}_{j|\mathbf{c}_{>M}}(\mathbf{w}'_{1*Mi\mathbf{c}}\mathbf{u}_{1*M\mathbf{c}} + \sum_{q=M+1}^{L}\mathbf{w}'_{M*qi\mathbf{c}}\mathbf{u}_{M*q\mathbf{c}}) \\
&= \text{var}_{j|\mathbf{c}_{>M}}(\mathbf{x}'_{Mi\mathbf{c}}\boldsymbol{\gamma}_M) + \text{var}_{j|\mathbf{c}_{>M}}(\mathbf{w}'_{1*Mi\mathbf{c}}\mathbf{u}_{1*M\mathbf{c}}) \\
&\quad + \sum_{q=M+1}^{L}\text{var}_{j|\mathbf{c}_{>M}}(\mathbf{w}'_{M*qi\mathbf{c}}\mathbf{u}_{M*q\mathbf{c}}) \\
&= \boldsymbol{\gamma}'_M\boldsymbol{\Phi}_M\boldsymbol{\gamma}_M + \tau_M + \sum_{q=M+1}^{L}tr(\boldsymbol{\Sigma}_{M*q}\mathrm{T}_{M*q})
\end{aligned} \tag{A11}$$

The steps again follow those in Equation A2-A9, with the exception that terms are dropped if they have no variance at level-M.

Lastly, letting k denote level-L cluster membership, we can compute the level-L outcome variance as

$$\begin{aligned}
\text{var}_k(y_{i\mathbf{c}}) &= \text{var}_k(\gamma_0 + \sum_{l=1}^{L}\mathbf{x}'_{li\mathbf{c}}\boldsymbol{\gamma}_l + \sum_{l=1}^{L-1}\sum_{q=l+1}^{L}\mathbf{w}'_{l*qi\mathbf{c}}\mathbf{u}_{l*q\mathbf{c}} + e_{i\mathbf{c}}) \\
&= \text{var}_k(\sum_{l=1}^{L}\mathbf{x}'_{li\mathbf{c}}\boldsymbol{\gamma}_l) + \text{var}_k(\sum_{l=1}^{L-1}\sum_{q=l+1}^{L}\mathbf{w}'_{1*qi\mathbf{c}}\mathbf{u}_{1*q\mathbf{c}}) + \text{var}_k(e_{i\mathbf{c}}) \\
&= \text{var}_k(\mathbf{x}'_{Li\mathbf{c}}\boldsymbol{\gamma}_L) + \text{var}_k(\mathbf{w}'_{1*3i\mathbf{c}}\mathbf{u}_{1*3\mathbf{c}}) \\
&= \boldsymbol{\gamma}'_L\boldsymbol{\Phi}_L\boldsymbol{\gamma}_L + \tau_L
\end{aligned} \tag{A12}$$

These steps again follow those in Equation A2-A9, excluding terms with no variance at level-L.

Appendix B:

Derivation of model-implied outcome variance for non-cluster-mean-centered models with any number of levels

In this appendix, we derive the model-implied total and level-specific outcome variance for non-cluster-mean-centered models, that is, models in which *at least one* lower-level (i.e., below level-L) predictor is not cluster-mean-centered (e.g., models that utilize grand-mean-centered or uncentered lower-level predictors). We first express generically an L-level MLM that does not assume cluster-mean-centering as

$$\begin{aligned}
y_{i\mathbf{c}} &= \gamma_0 + \mathbf{x}'_{i\mathbf{c}}\boldsymbol{\gamma} + \sum_{l=2}^{L}\mathbf{w}'_{i\mathbf{c}*l}\mathbf{u}_{\mathbf{c}l} + e_{i\mathbf{c}} \\
e_{i\mathbf{c}} &\sim N(0, \sigma^2) \\
\mathbf{u}_{l\mathbf{c}} &\sim MVN(\mathbf{0}, \mathrm{T}_l)
\end{aligned} \tag{B1}$$

with γ_0 denoting the fixed component of the intercept, l denoting the level of a given vector of terms ($l = 2, \ldots, L$), i denoting observation within level-2 cluster, and $\mathbf{c}$ denoting the set of observation i's cluster memberships for all levels greater than 1 (e.g., level-2, -3, and -4 cluster membership for a four-level model). Here $\mathbf{x}_{i\mathbf{c}}$ denotes the vector of all predictors (which need not be cluster-mean-centered) with fixed components of slopes, $\boldsymbol{\gamma}$ the vector of all fixed components of slopes, $\mathbf{w}_{i\mathbf{c}*l}$ the vector of 1 (for the intercept) and all predictors with slopes that vary across level-l units, $\mathbf{u}_{\mathbf{c}l}$ the vector of all level-l random slope residuals (multivariate normally distributed with covariance matrix T_l), and $e_{i\mathbf{c}}$ the level-1 residual (normally distributed with variance σ^2).

To allow for a decomposition of variance into level-specific components, we will first reexpress this model by decomposing the vectors of predictors into level-specific portions (i.e., portions that have variance at only one level). For instance, for all of the predictors in the $\mathbf{x}_{i\mathbf{c}}$ vector, we can express their purely level-1 portion as their level-2 cluster-mean-centered versions, expressed in vector form as $\mathbf{x}_{i\mathbf{c}} - \mathbf{x}_{i\mathbf{c}\cdot 2}$, with $\mathbf{x}_{i\mathbf{c}\cdot 2}$ denoting the level-2 cluster means (note that, for predictors at level-2 or higher, their corresponding element in $\mathbf{x}_{i\mathbf{c}} - \mathbf{x}_{i\mathbf{c}\cdot 2}$ will be guaranteed to be 0, as they have no variance at level-1). At an intermediate level l (i.e., $1 < l < L$), we can express the purely level-l portion of the predictors in $\mathbf{x}_{i\mathbf{c}}$ as $\mathbf{x}_{i\mathbf{c}\cdot l} - \mathbf{x}_{i\mathbf{c}\cdot(l+1)}$, where $\mathbf{x}_{i\mathbf{c}\cdot l}$ denotes the level-l cluster means of the predictors and $\mathbf{x}_{i\mathbf{c}\cdot(l+1)}$ denotes the cluster means at the subsequent level. The purely level-L (i.e., the highest level) portion is simply the level-L cluster means, or $\mathbf{x}_{i\mathbf{c}\cdot L}$. We can hence fully decompose the $\mathbf{x}_{i\mathbf{c}}$ vector into L level-specific portions as such:

$$\mathbf{x}_{i\mathbf{c}} = \mathbf{x}_{i\mathbf{c}} - \mathbf{x}_{i\mathbf{c}\cdot 2} + \sum_{l=2}^{L-1}\left(\mathbf{x}_{i\mathbf{c}\cdot l} - \mathbf{x}_{i\mathbf{c}\cdot(l+1)}\right) + \mathbf{x}'_{i\mathbf{c}\cdot L} \tag{B2}$$

We can do the same computation for each of the L - 1 $\mathbf{w}_{i\mathbf{c}*l}$ vectors, noting that their level-1 portion is $\mathbf{w}_{i\mathbf{c}*l} - \mathbf{w}_{i\mathbf{c}\cdot 2*l}$ (where $\mathbf{w}_{i\mathbf{c}\cdot 2*l}$ denotes the level-2 cluster means), their intermediate level-q ($1 < q < L$) portion is $\mathbf{w}_{i\mathbf{c}\cdot k*l} - \mathbf{w}_{i\mathbf{c}\cdot(k+1)*l}$, and their level-$L$ portion is $\mathbf{w}_{i\mathbf{c}\cdot L*l}$. Hence we can fully decompose the $\mathbf{w}_{i\mathbf{c}*l}$ vector into L level-specific portions as

$$\begin{aligned}
\sum_{l=2}^{L}\mathbf{w}_{i\mathbf{c}*l} = \sum_{l=2}^{L}\Bigg(\mathbf{w}_{i\mathbf{c}*l} - \mathbf{w}_{i\mathbf{c}\cdot 2*l} + \sum_{q=2}^{L-1}\left(\mathbf{w}_{i\mathbf{c}\cdot q*l} - \mathbf{w}_{i\mathbf{c}\cdot(q+1)*l}\right) \\
+ \mathbf{w}_{i\mathbf{c}\cdot L*l}\Bigg)
\end{aligned} \tag{B3}$$

Replacing the vectors in the Equation B1 expression with their (equivalent) decomposed versions, we can write the level-L non-cluster-mean-centered model as

$$\begin{aligned}
y_{i\mathbf{c}} &= \gamma_0 + \mathbf{x}'_{i\mathbf{c}}\boldsymbol{\gamma} + \textstyle\sum_{l=2}^{L}\mathbf{w}'_{*li\mathbf{c}}\mathbf{u}_{l\mathbf{c}} + e_{i\mathbf{c}} \\
&= \gamma_0 + (\mathbf{x}'_{i\mathbf{c}} - \mathbf{x}'_{i\mathbf{c}\cdot 2} + \textstyle\sum_{l=2}^{L-1}(\mathbf{x}'_{i\mathbf{c}\cdot l} - \mathbf{x}'_{i\mathbf{c}\cdot(l+1)}) + \mathbf{x}'_{i\mathbf{c}\cdot L})\boldsymbol{\gamma} \\
&\quad + \textstyle\sum_{l=2}^{L}((\mathbf{w}'_{i\mathbf{c}*l} - \mathbf{w}'_{i\mathbf{c}\cdot 2*l} + \\
&\quad \textstyle\sum_{q=2}^{L-1}(\mathbf{w}'_{i\mathbf{c}\cdot q*l} - \mathbf{w}'_{i\mathbf{c}\cdot(q+1)*l}) + \mathbf{w}'_{i\mathbf{c}\cdot L*l})\mathbf{u}_{l\mathbf{c}}) + e_{i\mathbf{c}}
\end{aligned} \tag{B4}$$

We can then compute the model-implied total outcome variance as

$$\begin{aligned}
\mathrm{var}(y_{ic}) &= \mathrm{var}(\gamma_0 + (\mathbf{x}'_{ic} - \mathbf{x}'_{ic\cdot 2} + \textstyle\sum_{l=2}^{L-1}(\mathbf{x}'_{ic\cdot l} - \mathbf{x}'_{ic\cdot(l+1)}) + \mathbf{x}'_{ic\cdot L})\boldsymbol{\gamma} + \textstyle\sum_{l=2}^{L}((\mathbf{w}'_{ic*l} - \mathbf{w}'_{ic\cdot 2*l} + \textstyle\sum_{q=2}^{L-1}(\mathbf{w}'_{ic\cdot q*l} - \mathbf{w}'_{ic\cdot(q+1)*l}) + \mathbf{w}'_{ic\cdot L*l})\mathbf{u}_{lc}) + e_{ic}) \\
&= \mathrm{var}((\mathbf{x}'_{ic} - \mathbf{x}'_{ic\cdot 2} + \textstyle\sum_{l=2}^{L-1}(\mathbf{x}'_{ic\cdot l} - \mathbf{x}'_{ic\cdot(l+1)}) + \mathbf{x}'_{ic\cdot L})\boldsymbol{\gamma}) + \mathrm{var}(\textstyle\sum_{l=2}^{L}((\mathbf{w}'_{ic*l} - \mathbf{w}'_{ic\cdot 2*l} + \textstyle\sum_{q=2}^{L-1}(\mathbf{w}'_{ic\cdot q*l} - \mathbf{w}'_{ic\cdot(q+1)*l}) + \mathbf{w}'_{ic\cdot L*l})\mathbf{u}_{lc})) + \mathrm{var}(e_{ic}) \\
&= \underbrace{\mathrm{var}((\mathbf{x}'_{ic} - \mathbf{x}'_{ic\cdot 2})\boldsymbol{\gamma})}_{(a)} + \underbrace{\textstyle\sum_{l=1}^{L-1}\mathrm{var}((\mathbf{x}'_{ic\cdot l} - \mathbf{x}'_{ic\cdot(l+1)})\boldsymbol{\gamma})}_{(b)} + \underbrace{\mathrm{var}(\mathbf{x}'_{ic\cdot L}\boldsymbol{\gamma})}_{(c)} + \underbrace{\textstyle\sum_{l=2}^{L}\mathrm{var}(\mathbf{w}'_{ic*l} - \mathbf{w}'_{ic\cdot 2*l})\mathbf{u}_{lc}}_{(d)} \\
&\quad + \underbrace{\textstyle\sum_{l=2}^{L}\textstyle\sum_{q=2}^{L-1}\mathrm{var}((\mathbf{w}'_{ic\cdot q*l} - \mathbf{w}'_{ic\cdot(q+1)*l})\mathbf{u}_{lc})}_{(e)} + \underbrace{\textstyle\sum_{l=2}^{L}\mathrm{var}(\mathbf{w}'_{ic\cdot L*l}\mathbf{u}_{lc})}_{(f)} + \underbrace{\mathrm{var}(e_{ic})}_{(g)}
\end{aligned} \tag{B5}$$

The outcome variance is separable into parts (a)-(g) given that the following pairs are uncorrelated with each other: 1) fixed components and random components, 2) the level-specific portion of predictors at a given level and the level-specific portion of predictors at a different level, and 3) the level-1 residuals and all other terms.

Part (a) in Equation B5 is computed as

$$\mathrm{var}\Big((\mathbf{x}'_{ic} - \mathbf{x}'_{ic\cdot 2})\boldsymbol{\gamma}\Big) = \boldsymbol{\gamma}'\boldsymbol{\Phi}_1\boldsymbol{\gamma} \tag{B6}$$

where $\boldsymbol{\Phi}_1$ denotes the covariance matrix of all terms in $\mathbf{x}_{ic} - \mathbf{x}_{ic\cdot 2}$, i.e., the purely level-1 covariance matrix of all predictors with fixed components. Similarly, part (b) in Equation B5 is computed as

$$\sum_{l=1}^{L-1}\mathrm{var}\Big(\Big(\mathbf{x}'_{ic\cdot l} - \mathbf{x}'_{ic\cdot(l+1)}\Big)\boldsymbol{\gamma}\Big) = \sum_{l=1}^{L-1}\boldsymbol{\gamma}'\boldsymbol{\Phi}_l\boldsymbol{\gamma} \tag{B7}$$

where $\boldsymbol{\Phi}_l$ denotes the covariance matrix of all terms in $\mathbf{x}_{ic\cdot l} - \mathbf{x}_{ic\cdot(l+1)}$, and part (c) is computed as

$$\mathrm{var}(\mathbf{x}'_{ic\cdot L}\boldsymbol{\gamma}) = \boldsymbol{\gamma}'\boldsymbol{\Phi}_L\boldsymbol{\gamma} \tag{B8}$$

where $\boldsymbol{\Phi}_L$ denotes the covariance matrix of all terms in $\mathbf{x}_{ic\cdot L}$.

Part (d) in Equation B5, using the law of total variance as in Appendix A Equations A4-A6, is computed as

$$\begin{aligned}
\sum_{l=2}^{L}\mathrm{var}(\mathbf{w}'_{ic*l} - \mathbf{w}'_{ic\cdot 2*l})\mathbf{u}_{lc} &= \sum_{l=2}^{L}E\Big[\mathrm{var}\Big((\mathbf{w}'_{ic*l} - \mathbf{w}'_{ic\cdot 2*l})\mathbf{u}_{lc}|\mathbf{u}_{lc}\Big)\Big] \\
&\quad + \sum_{l=2}^{L}\mathrm{var}\Big(E[(\mathbf{w}'_{ic*l} - \mathbf{w}'_{ic\cdot 2*l})\mathbf{u}_{lc}|\mathbf{u}_{lc}]\Big) \\
&= \sum_{l=2}^{L}tr(\boldsymbol{\Sigma}_{1*l}\mathrm{T}_l) + \sum_{l=2}^{L}\mathrm{var}(E[\mathbf{w}'_{ic*l} - \mathbf{w}'_{ic\cdot 2*l}]\mathbf{u}_{lc}) \\
&= \sum_{l=2}^{L}tr(\boldsymbol{\Sigma}_{1*l}\mathrm{T}_l) + \sum_{l=2}^{L}\mathrm{var}(\mathbf{m}'_{1*l}\mathbf{u}_{lc}) \\
&= \sum_{l=2}^{L}tr(\boldsymbol{\Sigma}_{1*l}\mathrm{T}_l) + \sum_{l=2}^{L}\mathbf{m}'_{1*l}\mathrm{var}(\mathbf{u}_{lc})\mathbf{m}_{1*l} \\
&= \sum_{l=2}^{L}tr(\boldsymbol{\Sigma}_{1*l}\mathrm{T}_l) + \sum_{l=2}^{L}\mathbf{m}'_{1*l}\mathrm{T}_l\mathbf{m}_{1*l}
\end{aligned} \tag{B9}$$

Where $\boldsymbol{\Sigma}_{1*l}$ denotes the covariance matrix of terms in $\mathbf{w}_{ic*l} - \mathbf{w}_{ic\cdot 2*l}$ (i.e., the purely level-1 covariance of all predictors with random components) and $\mathbf{m}_{1*l}$ denotes the vectors of means of the elements of $\mathbf{w}_{ic*l} - \mathbf{w}_{ic\cdot 2*l}$. Using the same operations, part (e) is computed as

$$\begin{aligned}
&\sum_{l=2}^{L}\sum_{q=1}^{L-1}\mathrm{var}((\mathbf{w}'_{ic\cdot q*l} - \mathbf{w}'_{ic\cdot(q+1)*l})\mathbf{u}_{lc}) \\
&= \sum_{l=2}^{L}\sum_{q=1}^{L-1}E[\mathrm{var}((\mathbf{w}'_{ic\cdot q*l} - \mathbf{w}'_{ic\cdot(q+1)*l})\mathbf{u}_{lc}|\mathbf{u}_{lc})] \\
&\quad + \sum_{l=2}^{L}\sum_{q=1}^{L-1}\mathrm{var}(E[(\mathbf{w}'_{ic\cdot q*l} - \mathbf{w}'_{ic\cdot(q+1)*l})\mathbf{u}_{lc}|\mathbf{u}_{lc}]) \\
&= \sum_{l=2}^{L}\sum_{q=1}^{L-1}tr(\boldsymbol{\Sigma}_{q*l}\mathrm{T}_l) + \sum_{l=2}^{L}\sum_{q=1}^{L-1}\mathbf{m}'_{q*l}\mathrm{T}_l\mathbf{m}_{q*l}
\end{aligned} \tag{B10}$$

Where $\boldsymbol{\Sigma}_{q*l}$ and $\mathbf{m}_{q*l}$ denote, respectively, the covariance matrix and the mean vector of terms in $\mathbf{w}_{ic\cdot q*l} - \mathbf{w}_{ic\cdot(q+1)*l}$, and part (f) is computed as

$$\begin{aligned}
\sum_{l=2}^{L}\mathrm{var}(\mathbf{w}'_{ic\cdot L*l}\mathbf{u}_{lc}) &= \sum_{l=2}^{L}E\big[\mathrm{var}(\mathbf{w}'_{ic\cdot L*l}\mathbf{u}_{lc}|\mathbf{u}_{lc})\big] + \sum_{l=2}^{L}\mathrm{var}(E\big[\mathbf{w}'_{ic\cdot L*l}\mathbf{u}_{lc}|\mathbf{u}_{lc}\big]) \\
&= \sum_{l=2}^{L}tr(\boldsymbol{\Sigma}_{L*l}\mathrm{T}_l) + \sum_{l=2}^{L}\mathbf{m}'_{L*l}\mathrm{T}_l\mathbf{m}_{L*l}
\end{aligned} \tag{B11}$$

Where $\boldsymbol{\Sigma}_{L*l}$ and $\mathbf{m}_{L*l}$ denote, respectively, the covariance matrix and the mean vector of terms in $\mathbf{w}_{ic\cdot L*l}$. The final part (g) in Equation B5 is simply

$$\mathrm{var}(e_{ic}) = \sigma^2 \tag{B12}$$

Putting Equations B6-B12 together, the total model-implied total outcome variance is

$$\begin{aligned}
\mathrm{var}(y_{ic}) &= \boldsymbol{\gamma}'\boldsymbol{\Phi}_1\boldsymbol{\gamma} + \sum_{l=2}^{L-1}\boldsymbol{\gamma}'\boldsymbol{\Phi}_l\boldsymbol{\gamma} + \boldsymbol{\gamma}'\boldsymbol{\Phi}_L\boldsymbol{\gamma} + \sum_{l=2}^{L}tr(\boldsymbol{\Sigma}_{1*l}\mathrm{T}_l) \\
&\quad + \sum_{l=2}^{L}\mathbf{m}'_{1*l}\mathrm{T}_l\mathbf{m}_{1*l} + \sum_{l=2}^{L}\sum_{q=2}^{L-1}tr(\boldsymbol{\Sigma}_{q*l}\mathrm{T}_l) \\
&\quad + \sum_{l=2}^{L}\sum_{q=2}^{L-1}\mathbf{m}'_{q*l}\mathrm{T}_l\mathbf{m}_{q*l} + \sum_{l=2}^{L}tr(\boldsymbol{\Sigma}_{L*l}\mathrm{T}_l) \\
&\quad + \sum_{l=2}^{L}\mathbf{m}'_{L*l}\mathrm{T}_l\mathbf{m}_{L*l} + \sigma^2 \\
&= \sum_{l=1}^{L}\boldsymbol{\gamma}'\boldsymbol{\Phi}_l\boldsymbol{\gamma} + \sum_{l=2}^{L}\sum_{q=1}^{L}tr(\boldsymbol{\Sigma}_{q*l}\mathrm{T}_l) \\
&\quad + \sum_{l=2}^{L}\sum_{q=1}^{L}\mathbf{m}'_{q*l}\mathrm{T}_l\mathbf{m}_{q*l} + \sigma^2
\end{aligned} \tag{B13}$$

To facilitate discussion of decomposing this expression into level-specific portions (by ensuring that each term with an l

subscript reflects variance specifically at level-l) we will (equivalently) express this variance as such

$$\mathrm{var}(y_{ic}) = \sum_{l=1}^{L} \boldsymbol{\gamma}'\boldsymbol{\Phi}_l\boldsymbol{\gamma} + \sum_{q=2}^{L}\sum_{l=1}^{L} tr(\boldsymbol{\Sigma}_{l*q}\mathrm{T}_q) + \sum_{l=2}^{L}\sum_{q=1}^{L} \mathbf{m}'_{q*l}\mathrm{T}_l\mathbf{m}_{q*l} + \sigma^2 \quad \text{(B14)}$$

Where $\boldsymbol{\Sigma}_{l*k}$ denotes the purely level-*1* covariance of all predictors with random components across level-k units.

Each set of terms in the variance decomposition in Equation B14 reflects variance attributable to a specific source. Namely:

- $\boldsymbol{\gamma}'\boldsymbol{\Phi}_l\boldsymbol{\gamma}$ represents the variance explained by the level-l varying portion of predictors via fixed components of slopes . . . . (B15)
- $tr(\boldsymbol{\Sigma}_{l*q}\mathrm{T}_q)$ represents the variance explained by the level-l varying portion of predictors via random slope variation across level-q units . . . . (B16)
- $\sum_{q=1}^{L} \mathbf{m}'_{q*l}\mathrm{T}_l\mathbf{m}_{q*l}$ represents the variance explained by the level-l outcome means via random intercept variation at the mean of all predictors with random slopes . . . . (B17)
- σ^2 represents the variance attributable to level-1 errors . . . . (B18)

Hence, from the model-implied outcome variance in Equation B14 and the individual variance components in Equations B15-B18, we can form the total R-squared measures provided in Table 6.

From this decomposition in Equation B14, we can similarly compute the outcome variance specific to each level. Specifically, the level-1 outcome variance is given as

$$\mathrm{var}(y_{ic}|l=1) = \boldsymbol{\gamma}'\boldsymbol{\Phi}_1\boldsymbol{\gamma} + \sum_{k=2}^{L} tr(\boldsymbol{\Sigma}_{1*k}\mathrm{T}_k) + \sigma^2 \quad \text{(B19)}$$

Here in Equation B19, we are simply including only the terms from the total outcome variance in Equation B14 that reflect purely level-1 variance. Similarly, at an intermediate level-M (i.e., $1 < M < L$), the level-specific outcome variance is

$$\mathrm{var}(y_{ic}|l=M) = \boldsymbol{\gamma}'\boldsymbol{\Phi}_M\boldsymbol{\gamma} + \sum_{k=2}^{L} tr(\boldsymbol{\Sigma}_{M*k}\mathrm{T}_k) + \sum_{k=1}^{L} \mathbf{m}'_{k*M}\mathrm{T}_M\mathbf{m}_{k*M} \quad \text{(B20)}$$

Lastly, at the highest level (L), the level-specific outcome variance is

$$\mathrm{var}(y_{ic}|l=L) = \boldsymbol{\gamma}'\boldsymbol{\Phi}_L\boldsymbol{\gamma} + \sum_{k=2}^{L} tr(\boldsymbol{\Sigma}_{L*k}\mathrm{T}_k) + \sum_{k=1}^{L} \mathbf{m}'_{k*L}\mathrm{T}_L\mathbf{m}_{k*L} \quad \text{(B21)}$$

From the level-specific decomposition in Equations B19-B21, and from the sources listed in Equations B15-B18, we can form each of the level-specific measures defined in Table 6.

As special cases of the above formulas, we can consider the model-implied total and level-specific variance from a three-level model, which can be used to form the three-level measures defined in Table 5. The total outcome variance for a three-level model is given as

$$\begin{aligned}\mathrm{var}(y_{ijk}) &= \boldsymbol{\gamma}'\boldsymbol{\Phi}_1\boldsymbol{\gamma} + \boldsymbol{\gamma}'\boldsymbol{\Phi}_2\boldsymbol{\gamma} + \boldsymbol{\gamma}'\boldsymbol{\Phi}_3\boldsymbol{\gamma} + tr(\boldsymbol{\Sigma}_{1*2}\mathrm{T}_2) + tr(\boldsymbol{\Sigma}_{1*3}\mathrm{T}_3) + tr(\boldsymbol{\Sigma}_{2*2}\mathrm{T}_2) \\ &+ tr(\boldsymbol{\Sigma}_{2*3}\mathrm{T}_3) + tr(\boldsymbol{\Sigma}_{3*2}\mathrm{T}_2) + tr(\boldsymbol{\Sigma}_{3*3}\mathrm{T}_3) + \mathbf{m}'_{1*2}\mathrm{T}_2\mathbf{m}_{1*2} + \mathbf{m}'_{2*2}\mathrm{T}_2\mathbf{m}_{2*2} \\ &+ \mathbf{m}'_{3*2}\mathrm{T}_2\mathbf{m}_{3*2} + \mathbf{m}'_{1*3}\mathrm{T}_3\mathbf{m}_{1*3} + \mathbf{m}'_{2*3}\mathrm{T}_3\mathbf{m}_{2*3} + \mathbf{m}'_{3*3}\mathrm{T}_3\mathbf{m}_{3*3} + \sigma^2\end{aligned} \quad \text{(B22)}$$

The level-1, level-2, and level-3 outcome variances in a three-level model are, respectively, given as

$$\mathrm{var}_{i|jk}(y_{ijk}) = \boldsymbol{\gamma}'\boldsymbol{\Phi}_1\boldsymbol{\gamma} + tr(\boldsymbol{\Sigma}_{1*2}\mathrm{T}_2) + tr(\boldsymbol{\Sigma}_{1*3}\mathrm{T}_3) + \sigma^2 \quad \text{(B23)}$$

$$\mathrm{var}_{j|k}(y_{ijk}) = \boldsymbol{\gamma}'\boldsymbol{\Phi}_2\boldsymbol{\gamma} + tr(\boldsymbol{\Sigma}_{2*2}\mathrm{T}_2) + tr(\boldsymbol{\Sigma}_{2*3}\mathrm{T}_3) + \mathbf{m}'_{1*2}\mathrm{T}_2\mathbf{m}_{1*2} + \mathbf{m}'_{2*2}\mathrm{T}_2\mathbf{m}_{2*2} + \mathbf{m}'_{3*2}\mathrm{T}_2\mathbf{m}_{3*2} \quad \text{(B24)}$$

$$\mathrm{var}_{k}(y_{ijk}) = \boldsymbol{\gamma}'\boldsymbol{\Phi}_3\boldsymbol{\gamma} + tr(\boldsymbol{\Sigma}_{3*2}\mathrm{T}_2) + tr(\boldsymbol{\Sigma}_{3*3}\mathrm{T}_3) + \mathbf{m}'_{1*3}\mathrm{T}_3\mathbf{m}_{1*3} + \mathbf{m}'_{2*3}\mathrm{T}_3\mathbf{m}_{2*3} + \mathbf{m}'_{3*3}\mathrm{T}_3\mathbf{m}_{3*3} \quad \text{(B25)}$$