STRUCTURAL EQUATION MODELING: A MULTIDISCIPLINARY JOURNAL
2022, VOL. 29, NO. 4, 489–506
https://doi.org/10.1080/10705511.2021.1962325



# R-squared Measures for Multilevel Mixture Models with Random Effects

Sonya K. Sterba and Jason D. Rights*

Vanderbilt University and University of British Columbia

**ABSTRACT**

Multilevel regression mixtures involving both discrete latent classes and continuous random effects are an increasingly popular approach for accommodating nested data structures. However, their application has outpaced the development of effect size measures to aid model interpretation. In response, we provide a general framework of R-squared measures for multilevel regression mixtures with random effects as well as either classes only at level-1 (L1MIX), or classes only at level-2 (L2MIX), or classes at both levels (L1L2MIX). This work extends and unites a previous suite of R-squared measures for multilevel mixtures with latent classes but no random effects (Rights & Sterba, 2018) and a suite of R-squared measures for multilevel models with random effects but no latent classes (Rights & Sterba, 2019). The general framework provided here includes total and class-specific measures that each allow the researcher to distinguish among distinct sources of explained variance in the fitted model. We provide software for implementing these measures and provide two illustrative empirical examples.

**KEYWORDS**

Multilevel mixture model; R-squared measure; regression mixture model; multilevel model

Multilevel regression mixture models (e.g., Muthén & Asparouhov, 2009; Vermunt, 2004, 2008) have been applied by social scientists for over a decade to diverse outcomes such as bullying (Cho & Lee, 2018), reading achievement (Van Horn et al., 2016), foster care usage (Yampolskaya et al. (2011), support for democracy (Konte, 2016), time spent on homework (Flunger et al., 2019), and binge drinking (Soloski & Durtschi, 2019). In a multilevel regression mixture model, the term “mixture” indicates that the model contains categorical latent (unobserved) classes, the term “multilevel” indicates that the model accommodates nested or clustered data structures (e.g., students nested within school or clients nested within clinician), and the term “regression” indicates that, within latent class, the observed dependent variable or outcome is specified to be a function of exogenous observed predictors (for details see Sterba, 2014), as in the familiar case of (non-mixture) multilevel regression models.

There are three general kinds of multilevel regression mixtures (see Muthén & Asparouhov, 2009 for a review). One kind, that we will refer to here as L1MIX, accounts for clustering of level-1 units using continuously distributed random effect(s), wherein all observations within the same cluster[1] share the same value on the random effect. In the LIMIX there are also allowed to be $K$ level-1 latent classes (where $k = 1 \dots K$) across which parameters (e.g., intercepts and slopes) can vary. In another kind of multilevel regression mixture that we refer to here as L2MIX, clustering of level-1 units is accounted for using mean differences of parameters (e.g., intercepts and slopes) across $H$ level-2 latent classes (where $h = 1 \dots H$) and optionally also accounted for using continuously distributed random effects. In the third kind of multilevel regression mixture, here called L1L2MIX, continuously distributed random effects are again optional and there are now latent classes at both levels. In other words, in the L1L2MIX there are $K$ level-1 classes nested within each of $H$ level-2 classes. Though continuous random effects are required to account for dependency of level-1 units within cluster only for the L1MIX model (but are optional for the L2MIX and L1L2MIX), random effects are nonetheless regularly included in applications of all three kinds of multilevel regression mixtures for substantive reasons. For instance, researchers fitting L2MIX and L1L2MIX often find it substantively realistic to include random effects to account for quantitative differences within potentially qualitatively different latent classes (e.g., Muthén, 2001; Muthén, 2007; Sterba & Bauer, 2010).

In applications of conventional (i.e., non-mixture) multilevel models, once a best-fitting model is selected, researchers commonly report R-squared measures to describe and quantify the effect size associated with specific term(s) in the model (LaHuis et al., 2014). Indeed ten different preexisting R-squared measures for conventional multilevel models (Aguinis & Culpepper, 2015; Hox, 2010; Johnson, 2014; Kreft & de Leeuw, 1998; Raudenbush & Bryk, 2002; Snijders & Bosker, 2012; Vonesh & Chinchilli, 1997; Xu, 2003) were recently analytically related as special cases under a single umbrella framework (Rights & Sterba, 2019).

For multilevel regression *mixtures*, on the other hand, once a best-fitting model is selected, traditionally researchers have focused on reporting the results of significance tests of individual within-class parameters and providing accompanying qualitative

**CONTACT** Sonya K. Sterba Sonya.Sterba@vanderbilt.edu Department of Psychology & Human Development, Quantitative Methods Program, Vanderbilt University; 230 Appleton Place, Peabody #552, Nashville, TN, 37203.

*The authors contributed equally to this manuscript.

[1]In this paper we use the term “cluster” to refer to an *observed* hierarchical nesting unit at level-2. When level-1 units (e.g., students) share the same level-2 cluster (e.g., school), this typically induces dependency in their outcome scores. In this case, the cluster is the school. In contrast, in this paper we use the term level-1 or level-2 “latent class” to refer to an *unobserved* group of level-1 or level-2 units.

Supplemental data for this article can be accessed on the publisher’s website class labels. However, in response to widespread calls for also reporting effect sizes when presenting results from such models (e.g., American Psychological Association, 2009; Applebaum et al., 2018; Harlow et al., 1997; Kelley & Preacher, 2012; Panter & Sterba, 2011), a suite of R-squared measures were recently developed for multilevel regression mixtures; these measures can be used to communicate to what extent specific terms in the model are important in terms of accounting for outcome variance (Rights & Sterba, 2018). Though a useful first step toward ensuring effect sizes are more readily available for researchers, Rights and Sterba's (2018) suite of R-squared measures for multilevel regression mixtures is limited in the following ways:

(1) The previous suite of R-squared measures for multilevel regression mixtures is not applicable in the presence of continuous random effects. That is, the suite of measures from Rights and Sterba (2018) are applicable only to the restricted versions—often termed "nonparametric" versions—of multilevel regression mixtures that account for cluster dependencies solely using between-class differences in fixed effects, rather than using any continuous random effects. These "nonparametric" versions of multilevel regression mixtures would assume local independence of units within each class *h* for the L2MIX (Nagin, 2005; Sterba, 2013; Vermunt & Van Dijk, 2001) and assume local independence of units within each *kh*-class-combination for the L1L2MIX (Rights & Sterba, 2016; Vermunt & Magidson, 2005).

(2) The previous suite of R-squared measures for multilevel regression mixtures did not distinguish the contribution of level-1 (within-cluster) versus level-2 (between-cluster) predictors toward explaining variance. Yet in many substantive contexts (e.g., educational research on students nested within schools), it would be useful to separately assess and compare the variance accounted for by say, school-level predictors (e.g., principal experience or school-average delinquency) versus student-level predictors (e.g., how much a student's own delinquency deviates from their school-average or how much a student's time spent on homework deviates from their school average). Likewise it would be useful to be able to unconflate the between- versus within-school contributions of the student-level predictors (Cronbach, 1976; Curran & Bauer, 2011; Enders & Tofighi, 2007; Hedeker & Gibbons, 2006; Hofmann & Gavin, 1998; Snijders & Bosker, 2012).

In light of these existing limitations, the contributions of the current paper are as follows. To overcome limitation (1) mentioned above, here we provide a framework for constructing R-squared measures that is applicable to each kind of multilevel regression mixture *with continuous random effects* (i.e., L1MIX, L2MIX, and L1L2MIX).[2] We later explain which subset of R-squared measures are available in the simplified special cases in which (a) there are no continuous random effects (this subset includes but is not limited to those measures previously provided in Rights & Sterba, 2018), and (b) there are no categorical latent classes (i.e., those measures previously provided in Rights & Sterba, 2019). Our framework of measures is to be used as a descriptive assessment tool to quantify effect sizes for the researcher's final chosen multilevel mixture model. That is, the researcher should have already defined their set of predictors of substantive interest (typically based on substantive theory) and should have already identified the best-fitting number of classes in their multilevel regression mixture (e.g., using information criteria and interpretability) leading to a final model for which our R-squared measures can then be computed to aid in interpretation.[3]

To overcome limitation (2) mentioned above, here we provide a novel decomposition of model-implied outcome variance that is used to create R-squared measures in our framework that separately distinguish the contribution toward explained variance of purely within-cluster predictors versus purely between-cluster predictors. To facilitate distinguishing variance explained that is within- versus between-cluster, here we assume researchers have cluster-mean-centered level-1 predictors, following current recommendations (e.g., Hedeker & Gibbons, 2006; Kreft & de Leeuw, 1998; Preacher et al., 2010; Raudenbush & Bryk, 2002; Snijders & Bosker, 2012).

The next section begins with an overview of the framework for constructing R-squared measures for multilevel regression mixtures with random effects. We then review the data model for multilevel regression mixtures with classes at only one level (L1MIX and L2MIX). Subsequently we derive, define, and interpret the suite of R-squared measures for these L1MIX and L2MIX models and explain how these measures are constructed from novel decompositions of model-implied total and class-specific outcome variance. Next we demonstrate the application of the suite of R-squared measures with an empirical example involving classes at only one level, for concrete illustration. We then present the data model, suite of R-squared measures, and empirical illustration for the more complex multilevel regression mixture with classes at both levels (L1L2MIX). We conclude with a description of software we freely provide for computing all of the R-squared measures developed here.

## Overview of framework for constructing R-squared measures

We can generically define an R-squared measure in the population as

$$R^2 = \frac{explained\ variance}{outcome\ variance} \tag{1}$$

R-squared measures in our framework use model-implied variances from the fitted model for the numerator and denominator (following Johnson, 2014; Nakagawa & Schielzth, 2013;

[2] By way of clarification, note that the R-squared effect size measures developed in this paper are for assessing the correspondence between predicted and observed outcome scores, not for explaining class membership. Other entropy-based R-squareds are already available for quantifying how well class memberships are predicted from observed responses (e.g., Lukočienė et al., 2010; Wedel & Kamakura, 1998).

[3] If the predictors need to be determined in an atheoretical, data-driven manner, an iterative search algorithm could be employed that performs an automated search simultaneously across both alternative predictor variables and alternative numbers of classes (e.g., Khalili & Chen, 2012; Raftery & Dean, 2006). In this event, after such an algorithm is employed to identify a final model, our R-squared measures could still be used to aid in the description and interpretation of that final model.

Rights & Sterba, 2018, 2019, 2021, Accepted; Snijders & Bosker, 2012). The R-squared measures in our framework can be differentiated by what they consider *outcome variance*, which goes in the denominator of Eqn. (1), and what they consider *explained variance*, which goes in the numerator of Eqn. (1), as follows.

For the *denominator* of Equation (1), choices are: *total* outcome variance or *class-specific* outcome variance—specifically, level-2-class specific outcome variance, level-1 class-specific outcome variance, or level-1/level-2-class-combination-specific outcome variance. *Total measures*, which use model-implied *total* outcome variance in the measure's denominator, can be computed for any multilevel mixture model considered here (i.e., L1MIX, L2MIX, or L1L2MIX) and allow explaining overall outcome variance pooling across latent classes. In contrast, *level-2 class-specific measures* (which can be computed for L2MIX or L1L2MIX models) use model-implied *level-2* class-specific outcome variance in the denominator and allow explaining outcome variance within a level-2 class. Similarly, *level-1 class-specific measures* (which can be computed for the L1MIX model) use model-implied level-1 class-specific variance in the denominator and allow explaining outcome variance within a level-1 class. Finally, *level-1/level-2 class-combination-specific measures* (which can be computed for the L1L2MIX model) use model-implied class-combination outcome variance in the denominator and allow explaining outcome variance for each distinct level-1/level-2 ($k$, $h$) class combination.

The latter three kinds of *class-specific* R-squared measures are particularly useful when the researcher is making a "direct interpretation" of latent classes (in which latent classes are viewed as corresponding to actual unobserved population subgroups and are given class-specific substantive labels; McLachlan & Peel, 2000; Titterington et al., 1985). Class-specific R-squared measures allow determining the proportion of variance explained for each unobserved subpopulation; moreover, comparing and contrasting class-specific measures across classes may aid in the description of substantively meaningful subpopulation differences. *Total* R-squared measures, on the other hand, may intuitively seem more relevant under an "indirect interpretation" of latent classes (in which latent classes are used together as a set to semi-parametrically approximate unknown and potentially complex underlying distributions of effects, but each latent class is not given a unique substantive interpretation; e.g., Nagin, 2005; Sterba et al., 2012; Vermunt & Van Dijk, 2001). However, we have elsewhere explained and demonstrated that it is most informative to compute and inspect *both* total and class-specific R-squared measures – even when researchers desire a direct interpretation of latent classes (see Rights & Sterba, 2018). This is because substantively different patterns of class-specific R-squareds can occur in conjunction with the same total R-squared value, and because even if a model explains a large proportion of outcome variance within a particular latent class, that model may nonetheless explain a small proportion of the total outcome variance. Hence, juxtaposing class-specific and total measures yields a more complete understanding of explained variance.

Regarding the *numerator* of Equation (1), our novel decomposition of model-implied outcome variance allows distinguishing how variance is accounted for via different possible *sources*. These sources include: level-1 predictors via marginal (weighted across-class-average) fixed components of slopes, level-1 predictors via across-class variation, level-2 predictors via marginal (weighted across-class-average) fixed components of slopes, level-2 predictors via across-class variation, level-1 predictors via random slope variation, cluster-specific outcome means via random intercept variation, and class-specific outcome means via across-class variation (details provided subsequently). There is precedent in the conventional multilevel literature and multilevel mixture literature for considering each of these sources as potential contributors to explained variance (see Vonesh & Chinchilli, 1997; Rights & Sterba, 2018, 2019 for rationales). As such, the framework provided in the present paper enables researchers to visualize and understand how variance is accounted for via each source individually by constructing a set of *single-source* R-squared measures (wherein one single source of explained variance is used at a time in the numerator of Equation (1)). Furthermore, we enable researchers to visualize and understand how variance is accounted for via multiple substantively interesting sources taken together by providing the option of supplementarily constructing *combination-source* measures (wherein a combination of sources of explained variance are used in the numerator of Equation (1)).

Put simply, it is not our goal to provide researchers with a particular omnibus "one-size-fits-all" R-squared measure that can represent the importance of every kind of term in the model at all once. Elsewhere the pursuit of such a "one-size-fits-all" measure for multilevel models/multilevel mixtures has been recognized to be futile (see explanations in, e.g., Edwards et al., 2008; Rights & Sterba, 2018, 2019). Rather, our recommendation here is for researchers to instead inspect a set of single-source measures for their fitted model to get a comprehensive breakdown of how variance is being accounted for in their model. Supplemental, optional construction of any particular combination-source measure can be driven by particular substantive objectives, as we will illustrate. In particular, later we illustrate the application and interpretation of single-source and combination-source R-squared measures using two empirical examples of multilevel regression mixtures.

## Data model for a multilevel regression mixture with classes only at level 1 (L1MIX)

Here we begin with a review of the multilevel regression mixture model with classes only at level-1, L1MIX (e.g., Muthén & Asparouhov, 2009; Vermunt, 2008), in Equation (2):

$$
\begin{aligned}
y_{ij|c_{ij}=k} &= \mathbf{x}_{ij}^{w'}\gamma^{w(k)} + \mathbf{x}_{j}^{b'}\gamma^{b(k)} + \mathbf{w}_{ij}'\mathbf{u}_j + \varepsilon_{ij} \\
&\varepsilon_{ij} \sim N(0,\theta^k) \\
&\mathbf{u}_j \sim N(\mathbf{0},\mathbf{T}) \\
p(c_{ij}=k) &= \pi^k = \frac{\exp(\omega^k)}{\sum_{k=1}^{K}\exp(\omega^k)}
\end{aligned}
\tag{2}
$$

Let $i$ index level-1 unit (e.g., student) and let $j$ index cluster (e.g., school), where $i = 1 \ldots N_j$ and $j = 1 \ldots J$. Our level-1 (e.g., student-level) categorical latent variable is denoted $c_{ij}$ and can take on values $k = 1 \ldots K$, where $K$ is the total number of level-1 classes.

Here we are modeling outcome $y_{ij}$ conditional on membership in level-1 class $k$, denoted $y_{ij|c_{ij}=k}$. The $P_1 \times 1$ vector $\mathbf{x}^w_{ij}$ – with $P_1$ denoting the number of level-1 predictors – contains all cluster-mean-centered level-1 predictors (including cross-level interaction product terms, which vary exclusively within-cluster and hence explain purely within-cluster variance, as derived in Rights & Sterba, 2019). The $(P_2+1) \times 1$ vector $\mathbf{x}^b_j$ – with $P_2$ denoting the number of level-2 predictors – contains 1 (for the intercept) and all level-2 predictors (which could include cluster means of level-1 predictors). The $P_1 \times 1$ vector $\gamma^{w(k)}$ contains the level-1 fixed component of coefficients associated with terms in $\mathbf{x}^w_{ij}$, and the $(P_2+1) \times 1$ vector $\gamma^{b(k)}$ contains the level-2 fixed coefficients associated with terms in $\mathbf{x}^b_j$. Coefficients in $\gamma^{w(k)}$ and $\gamma^{b(k)}$ may be allowed to vary across the $K$ classes (Muthén & Asparouhov, 2009). The $(P_1+1) \times 1$ vector $\mathbf{w}_{ij}$ consists of 1 (for the intercept) and all level-1 predictors, whereas the $(P_1+1) \times 1$ vector $\mathbf{u}_j$ contains the corresponding level-2 residuals $u_{0j}$ and $u_{1j} \ldots u_{P_1 j}$. elements of $\mathbf{u}_j$ are multivariate normally distributed with covariance matrix $\mathbf{T}$ (Muthén & Asparouhov, 2009). To obtain a fixed intercept or a fixed slope, the corresponding level-2 residual in $\mathbf{u}_j$ would be set to 0, so the corresponding term in $\mathbf{w}'_{ij}\mathbf{u}_j$ would also be 0, as would elements in $\mathbf{T}$ corresponding to its variance and covariance(s). $\varepsilon_{ij}$ is a normally-distributed level-1 residual with potentially class-varying variance $\theta^k$. The probability of individual $i$ in cluster $j$ being a member of class $k$, $p(c_{ij}=k)$, is denoted $\pi^k$ and is obtained from a multinomial logistic parameterization such that $\omega^k$ is a multinomial intercept and $\omega^K=0$ for identification. Note that it would also be possible for these multinomial logistic coefficients to randomly vary across clusters[4] (and doing so would not change the formulas we provide for R-squared computation[5]). In sum, this L1MIX model captures heterogeneity of regression coefficients both qualitatively (across level-1 latent classes) and quantitatively (across clusters).

## Data model for a multilevel regression mixture with classes only at level-2 (L2MIX)

The multilevel regression mixture model with classes only at level-2, L2MIX (e.g., Muthén & Asparouhov, 2009; Vermunt, 2008), is shown in Equation (3).

$$\begin{aligned} y_{ij|d_j=h} &= \mathbf{x}^{w\prime}_{ij}\gamma^{w(h)} + \mathbf{x}^{b\prime}_{j}\gamma^{b(h)} + \mathbf{w}'_{ij}\mathbf{u}_j + \varepsilon_{ij} \\ &\varepsilon_{ij} \sim N(0, \theta^h) \\ &\mathbf{u}_j \sim N(\mathbf{0}, \mathbf{T}^h) \\ p(d_j = h) &= \pi^h = \frac{\exp(\varpi^h)}{\sum_{h=1}^{H}\exp(\varpi^k)} \end{aligned} \quad (3)$$

The L2MIX specification in Equation (3) differs in the following ways from the L1MIX model in Equation (2). First, we are now modeling the outcome $y_{ij}$ conditional on membership in level-2 class $h$, denoted $y_{ij|c_j=h}$, and our level-2 (i.e., cluster-level) categorical latent variable $d_j$ can take on values $h=1 \ldots H$. Furthermore, in L2MIX not only $\theta^h$ and the elements of $\gamma^{w(h)}$ and $\gamma^{b(h)}$, but also the elements of $\mathbf{T}^h$ may be allowed to vary across the $H$ classes (see Muthén & Asparouhov, 2009). Lastly, the membership probability for cluster $j$ in class $h$, $p(c_j=h)$, given as $\pi^h$, is again modeled with class-specific multinomial intercepts $\omega^1 \ldots \omega^H$ where, for identification, $\omega^H=0$. Taken together, this L2MIX model also captures heterogeneity of regression coefficients both discretely (across level-2 latent classes) and continuously (across clusters).

## Constructing R-squared measures for a multilevel regression mixture with classes only at one level (L1MIX or L2MIX models)

### Constructing and interpreting Total R-squared measures for the L1MIX or L2MIX model

Using symbols defined in Appendix Table A1, we detail in Appendix Table A2 (Equations (A1)-(A8)) how to use the decomposition of *total* outcome variance we derived[6] for the L1MIX or L2MIX models to construct total R-squared measures. Specifically, to construct a total R-squared measure for L1MIX or L2MIX models, the expression for the total outcome variance (Equation A1) goes in the denominator. Possible numerator terms (Equations (A2)-(A8)) consist of variance attributable to seven different possible *sources*:

- Source $f_1$=L1 predictors via marginal[7] fixed components of slopes
- Source $f_2$=L2 predictors via marginal fixed components of slopes
- Source $v^r_1$=L1 predictors via random slope variation/covariation
- Source $m^r$=cluster-specific outcome means via random intercept variation
- Source $v^h_1$=L1 predictors via across-class slope variation/covariation
- Source $v^h_2$=L2 predictors via across-class slope variation/covariation
- Source $m^h$=class-specific outcome means via across-class variation

Note that this list of sources is written for L2MIX; for L1MIX, simply replace $h$ with $k$. This list of sources maps onto how applied researchers often seek to characterize their mixture

[4]If one were to specify a multinomial coefficient (e.g., the multinomial intercept) as a random coefficient, this adds $K$-1 dimensions of integration; Muthén and Asparouhov (2009) and Vermunt (2004) suggest some dimensionality reduction tricks in specification or estimation for managing this.

[5]If one were to include random effects for the multinomial portion of the model, in computing our subsequently defined R-squared measures, note that the $\mathbf{T}$ matrix in the R-squared formulas should contain *only* the variances/covariances of the level-2 residuals that reflect random intercepts/slopes in the regression of $y$ on the predictors (i.e., the components in $\mathbf{u}_j$ defined in Equation 1).

[6]A detailed derivation for this decomposition of *total* outcome variance for the L1MIX or L2MIX is provided in the Online Supplement.

[7]Here "marginal" refers to an across-class weighted average (weighted by the class probability).

model results (e.g., Halliday-Boykins et al., 2004; Morin & Marsh, 2015; Sher et al., 2011; Sterba & Bauer, 2014), though in the past these characterizations have instead been done qualitatively and heuristically, using textual summaries, rather than quantitatively. The present R-squared framework allows researchers the opportunity to *quantify* the contributions of these seven features.

A *single-source* total R-squared measure is defined as having one single such source of explained variance in the numerator at a time. Hence, there are seven possible single-source total R-squared measures for the L1MIX or L2MIX. These measures are enumerated, defined, and interpreted in Manuscript Table 1 (Column 1). Notation in Table 1 identifies and labels each R-squared measure by its denominator as well as its numerator source. For example, *Total* $R^2\_f_1$ is the *proportion of total variance attributable to level-1 predictors via marginal fixed components of slopes*. As another example, *Total* $R^2\_v_2^h$ is the *proportion of total variance attributable to level-2 predictors via across-class slope variation/covariation*. Our recommendation is to construct all of these *single-source* total R-squared measures for a complete understanding of variance accounted for.

If, as an optional accompaniment to the single-source measures, a researcher wanted to also construct *combination-source* total R-squared measures to speak to particular research questions involving multiple sources, the researcher would simply sum two or more of the single-source measures. Here we give some examples of combination-source measures that might have substantive appeal in particular contexts. For instance, a researcher might be interested in the proportion of total variance attributed to level-1 predictors via *both* across-class and across-cluster variation (i.e., *Total* $R^2\_v_1^h$+ *Total* $R^2\_v_1^r$). Another researcher might be interested in the proportion of total variance attributable to outcome means via *both* across-class and across-cluster variation (i.e., *Total* $R^2\_m^h$ + *Total* $R^2\_m^r$). Other examples of combination-source R-squareds that could be meaningful in certain contexts are: the proportion of total variance attributed to *both* level-1 and level-2 predictors via marginal fixed components of slopes (i.e., *Total* $R^2\_f_1$ + *Total* $R^2\_f_2$) and the proportion of total variance attributable to *both* level-1 and level-2 predictors via across-class slope variation/covariation (i.e., *Total* $R^2\_v_1^h$+ *Total* $R^2\_v_2^h$). Although these examples each feature the combination of two sources, it is possible to combine more than two sources in a R-squared measure, as long as this would be substantively meaningful.

Critically, it is not advisable to *exclusively* compute a combination-source without *also* reporting and interpreting its constituent single-source measures; computing only a combination-source measure would not yield insight into whether, for instance, one constituent source in the combination was accounting for the vast majority of the outcome variance while the other constituent sources in the combination were accounting for little,[8] or whether all constituent sources were accounting for about equal proportions of outcome variance (see also Rights & Sterba, 2018, 2019).

**Table 1.** Summary of single-source measures in the R-squared framework for L1MIX or L2MIX models. (Shown for L2MIX; for L1MIX replace *h* with *k*).

| Total measures | Computation* | Interpretation | Class-specific measures | Computation* | Interpretation |
|---|---|---|---|---|---|
| *Total* $R^2\_f_1$ | Eqn (A2)/(A1) | Proportion of total variance attributable to source $f_1$ | *Class-specific* $R^2\_f_1$ | Eqn (A10)/(A9) | Proportion of class-specific variance attributable to source $f_1$ |
| *Total* $R^2\_f_2$ | Eqn (A3)/(A1) | Proportion of total variance attributable to source $f_2$ | *Class-specific* $R^2\_f_2$ | Eqn (A11)/(A9) | Proportion of class-specific variance attributable to source $f_2$ |
| *Total* $R^2\_v_1^r$ | Eqn (A4)/(A1) | Proportion of total variance attributable to source $v_1^r$ | *Class-specific* $R^2\_v_1^r$ | Eqn (A12)/(A9) | Proportion of class-specific variance attributable to source $v_1^r$ |
| *Total* $R^2\_m^r$ | Eqn (A5)/(A1) | Proportion of total variance attributable to source $m^r$ | *Class-specific* $R^2\_m^r$ | Eqn (A13)/A(9) | Proportion of class-specific variance attributable to source $m^r$ |
| *Total* $R^2\_v_1^h$ | Eqn (A6)/(A1) | Proportion of total variance attributable to source $v_1^h$ | | | |
| *Total* $R^2\_v_2^h$ | Eqn (A7)/(A1) | Proportion of total variance attributable to source $v_2^h$ | | | |
| *Total* $R^2\_m^h$ | Eqn (A8)/(A1) | Proportion of total variance attributable to source $m^h$ | | | |

**Summary of Source Definitions**:
Source $f_1$=L1 predictors via marginal† fixed components of slopes
Source $f_2$=L2 predictors via marginal† fixed components of slopes
Source $v_1^r$=L1 predictors via random slope variation/covariation
Source $m^r$=cluster-specific outcome means via random intercept variation
Source $v_1^h$=L1 predictors via across-class slope variation/covariation
Source $v_2^h$=L2 predictors via across-class slope variation/covariation
Source $m^h$=class-specific outcome means via across-class variation

*Equations (A1)-(A8) are provided in Appendix Table A2 and Equations (A9)-(A13) are provided in Appendix Table A3.
†marginal = across-class weighted average. For class-specific measure, marginal fixed component of slope is just the fixed component of slope for that class

[8]Our later empirical example 1 can be used to illustrate this phenomena; in that example, the combination-source measure *Total* $R^2\_f_1$ + *Total* $R^2\_f_2$ = .16, but this is very unevenly split in that $f_1$ accounts for .159 and $f_2$ accounts for .001. If only the combination-source measure were presented, without the constituent single-source measures, the imbalance of their contributions would not be recognized.

### Constructing and interpreting class-specific R-squared measures for the L1MIX and L2MIX models

Our Appendix Table A3 (Equations (A9)-(A13)) indicates how to use the decomposition of *class-specific* outcome variance we derived[9] for the L1MIX or L2MIX models to construct *class-specific* R-squared measures. In particular, to construct a class-specific R-squared measure for L1MIX or L2MIX, the expression for the class-specific outcome variance (Equation A9) goes in the denominator of the measure. For the numerator of the class-specific R-squared measure, there are four potential sources to which class-specific variance can be attributable (corresponding to Equations (A10)-(A13)). These sources ($f_1$,$f_2$,$v_1^r$,$m^r$) comprise four of the seven sources in the bulleted list above. The three sources from the bulleted list above that do *not* appear (i.e., are not potential sources of class-specific explained variance) each pertain to *across*-class variation, which is not relevant here because a class-specific measure isolates *within*-class variance for a particular class. We recommend obtaining all single-source class-specific measures by dividing Equation (A9) by each Equation (A10)-(A13) one source at a time. Manuscript Table 1 (Column 4) summarizes these single-source class-specific measures for the L1MIX or L2MIX. Obtaining class-specific R-squared measures allows researchers to compare these measures' magnitudes among classes and also allows researchers to compare these measures' magnitudes with the total R-squared having the same source of explained variance (in order to understand *for whom* that source is important). Obtaining *combination-source* class-specific measures (i.e., using multiple of Equations (A10)-(A13) in the numerator) is again optional and would need to be substantively motivated.

## Empirical example: Multilevel regression mixture with classes only at one level

Our first empirical example is based on a $H$=2 L2MIX application from Heck and Thomas (2015) wherein 13,189 workers are nested within 165 organizations. Here the goal is to predict workplace productivity from worker-level predictors cluster-mean-centered worker satisfaction ($worksat_{ij}$) and cluster-mean-centered worker experience ($exper_{ij}$), as well as organization-level predictors organization-mean worker satisfaction (*worksat.j*) and organization-mean proportion female (*female.j*). Equation (4) indicates how regression coefficients were allowed to vary across classes and across clusters. Specification of $\pi^h$, $\theta^h$, $\mathbf{T}^h$ was as described in Equation (3).

$$\begin{aligned} product_{ij|d_j=h} &= \gamma_{00}^{b(h)} + worksat_{ij}\gamma_{10}^{w(h)} + exper_{ij}\gamma_{20}^{w(h)} \\ &+ worksat_{\bullet j}\gamma_{30}^{b(h)} + female_{\bullet j}\gamma_{40}^{b(h)} + u_{0j} \\ &+ exper_{ij}u_{1j} + \varepsilon_{ij} \end{aligned}$$

$$\begin{aligned} &\varepsilon_{ij} \sim N(0, \theta^h) \\ \text{Where :}\ &\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim N\left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_{00}^h & \\ \tau_{01}^h & \tau_{11}^h \end{bmatrix} \right) \\ &p(d_j = h) = \pi^h = \frac{\exp(\varpi^h)}{\sum_{h=1}^{H} \exp(\varpi^h)} \end{aligned} \tag{4}$$

The $H$ =2 L2MIX was fit in Mplus 8.4 (Muthén & Muthén, 1998-2020). Results indicated that an estimated 42% of organizations would fall into class 1 and 58% into class 2; these classes were primarily differentiated by mean level of productivity rather than slope differences. Parameter estimates and SEs are reported for this model in the Online Supplement,[10] and accompanying single-source total and class-specific R-squared results (computed using our later-defined R function) are provided in Table 2, along with a river plot (computed using the *riverplot* R package of Weiner, 2015 as in Barstead, 2019) that depicts the total R-squared results in Figure 1. Inspecting parameter estimates and SEs in the Online Supplement indicates that, in both classes of organizations, workers with higher satisfaction and more experience relative to their organization mean have significantly higher productivity. Although both level-1 predictors had significant effects on productivity in each class, neither level-2 predictor had significant effects on productivity in either class. This pattern of effects can be further visualized and quantified using the river plot of *total* R-squared results in Figure 1, which depicts the fact that 16% of total variance in worker productivity is attributable to predictors via across-class-average fixed components of slopes – with a trivial amount of that being due to level-2 predictors (*total* $R^2\_f_2$ =.001)) and almost all of that being due to level-1 predictors (*total* $R^2\_f_1$ =.159). Although the estimated class-specific slope estimates were observed to differ somewhat across class, the R-squared results in Table 2 provide the useful perspective that the impact of predictors via across-class slope differences is of trivial importance in accounting for total variance in worker productivity (*total* $R^2\_v_1^h$=.001, *total* $R^2\_v_2^h$=.001).

Parameter estimates and SEs reported for this model in the Online Supplement also indicate that latent class 2 had a higher mean level of worker productivity whereas latent class 1 had

**Table 2.** R-squared results for the $H$ = 2 L2MIX empirical example model predicting productivity where workers are nested within organizations.

| Total R-squared measures | | |
|---|---|---|
| Total $R^2\_f_1$ | | 0.159 |
| Total $R^2\_f_2$ | | 0.001 |
| Total $R^2\_v_1^r$ | | 0.001 |
| Total $R^2\_m^r$ | | 0.323 |
| Total $R^2\_v_1^h$ | | 0.001 |
| Total $R^2\_v_2^h$ | | 0.001 |
| Total $R^2\_m^h$ | | 0.338 |
| **Level-2-class-specific R-squared measures** | | |
| | $h$ = 1 | $h$ = 2 |
| Level-2-class-specific $R^2\_f_1$ | 0.219 | 0.258 |
| Level-2-class-specific $R^2\_f_2$ | 0.001 | 0.000 |
| Level-2-class-specific $R^2\_v_1^r$ | 0.001 | 0.000 |
| Level-2-class-specific $R^2\_m^r$ | 0.525 | 0.461 |

[9]A detailed derivation for this decomposition of *class-specific* outcome variance for the L1MIX and L2MIX is provided in the Online Supplement.
[10]The Online Supplement is available at https://my.vanderbilt.edu/sonyasterba/

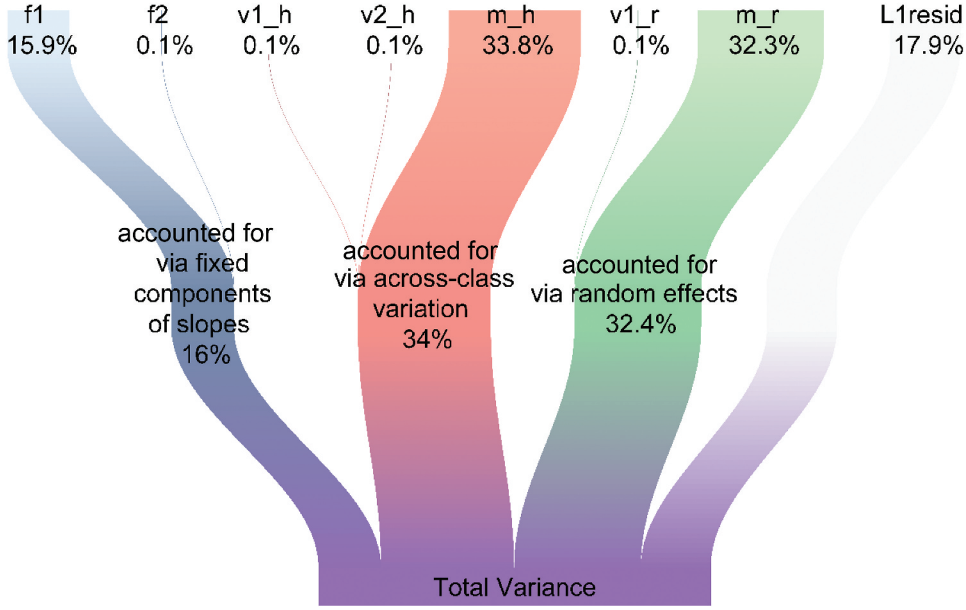


**Figure 1.** River plot depicting single-source total R-squared measures for the $H = 2$ L2MIX model predicting productivity, with workers nested within organizations.

a lower mean level of worker productivity. This pattern of results is visualized and quantified in the river plot of total R-squared results in Figure 1 in the fact that *total* $R^2\_m^h$ =.338, meaning that 33.8% of total variance in productivity is attributable to class-specific outcome means via across-class variation (i.e., $m^h$). Furthermore, the Online Supplement indicates that both classes showed random intercept but not random slope variation across organizations within-class in conditional mean productivity but not in slopes of level-1 predictor worker experience. This finding is more precisely quantified using the total R-squareds in the river plot in Figure 1 in that about one third of total variance in productivity is attributable to cluster-specific outcome means via random intercept variation (*total* $R^2\_m^r$ =.323), but virtually none is attributable to level-1 predictors via random slope variation (*total* $R^2\_v_1^r$=.001). Table 2 *class-specific* R-squared results further clarify that this amount of variability attributable to cluster-specific outcome means via random intercept variation is higher in class 1 (*class-specific* $R^2\_m^r$ =.525) than class 2 ($R^2\_m^r$ =.461).

Taken together, the Figure 1 river plot gives a parsimonious visual summary of the sources that are important in accounting for variance in worker productivity ($f_1, m^h, m^r$) and the sources that are not important for doing so ($f_2, v_1^h, v_2^h, v_1^r$). In sum, the R-squared results numerically quantify the facts that: (a) classes are distinguished mainly by across-class variation in means ($m^h$); (b) a sizable amount of across-cluster variation in intercepts remains within-class ($m^r$); (c) trivial amounts of the total outcome variance are attributable to level-2 predictors via any source ($f_2, v_2^h$); and (d) most of the variance that is attributable to level-1 predictors is via their marginal fixed components of slopes ($f_1$) rather than via their across-class variation ($v_1^h$) or across-cluster variation ($v_1^r$).

## Data model for multilevel regression mixture with classes at both level-1 & −2 (L1L2MIX)

Next, we turn to the specification of the multilevel regression mixture model with classes at *both* levels, L1L2MIX (e.g., Muthén & Asparouhov, 2009; Vermunt, 2008), in Equation (5).

$$\text{Where :}\quad \begin{aligned} y_{ij|c_{ij}=k,d_j=h} &= \mathbf{x}_{ij}^{w\prime}\gamma^{w(kh)} + \mathbf{x}_j^{b\prime}\gamma^{b(kh)} + \mathbf{w}_{ij}^{\prime}\mathbf{u}_j + \varepsilon_{ij} \\ \varepsilon_{ij} &\sim N(0, \theta^{kh}) \\ \mathbf{u}_j &\sim N(\mathbf{0}, \mathbf{T}^h) \end{aligned} \tag{5}$$

$$p(d_j = h) = \pi^h = \frac{\exp(\varpi^h)}{\sum_{h=1}^{H}\exp(\varpi^h)} \text{ and}$$

$$p(c_{ij} = k|d_j = h) = \pi^{k|h} = \frac{\exp(\omega^k + \delta^{kh})}{\sum_{k=1}^{K}\exp(\omega^k + \delta^{kh})}$$

The L1L2MIX model in Equation (5) combines elements of the L1MIX and L2MIX models from Equations (2) and (3). In particular, outcome $y_{ij}$ is now conditional on *class-combination kh* membership, that is, membership in level-1 class $k$ nested within level-2 class $h$, denoted $y_{ij|c_{ij}=k,d_j=h}$.

The level-2 (i.e., cluster-level) categorical latent variable $d_j$ can take on values $h$ =1 … $H$ for cluster $j$, and the level-1 categorical latent variable $c_{ij}$ can take on values $k$ =1 … $K$ for observation $i$ in cluster $j$. In turn this implies that level-1/level-2 class-combination memberships range from $k,h$ = 1,1 to $k,h$ = $K,H$. Now regression coefficients in $\gamma^{w(kh)}$and $\gamma^{b(kh)}$ and residual variance $\theta^{kh}$can be specific to class-combination $kh$ (e.g., Muthén & Asparouhov, 2009).[11]

The probability that cluster $j$ is a member of level-2 class $h$, $p(d_j = h)$, is denoted $\pi^h$ and is modeled by level-2 class-specific multinomial intercept, $\varpi^h$. The conditional probability that

[11]In the most general specification of a L1L2MIX, a given level-1 class $k$ need not be comparable across level-2 classes $h$ = 1 … $H$. However, researchers may choose to place equality constraints on, for instance, regression coefficient estimates and residual variance estimates across $h$ within $k$ to make each $k$ comparable across $h$. Then level-2 classes $h$ = 1 … $H$ in the L1L2MIX would differ only due to varying level-1 class proportions within $h$ (e.g., Lukočienė et al., 2010). The R-squared computations we employ accommodate the general or more constrained specifications.

observation $i$ in cluster $j$ is a member of level-1 class $k$, given their cluster belongs to level-2 class $h$, is denoted $\pi^{k|h}$ and is modeled by a multinomial intercept, $\omega^k$, and multinomial slope, $\delta^{kh}$. For identification $\varpi^H = \omega^K = \delta^{kH} = 0$ for all $k$, and $\delta^{Kh} = 0$ for all $h$.

## Constructing R-squared measures for a multilevel regression mixture with classes at *both* levels (i.e., L1L2MIX models)

### *Constructing and interpreting Total R-squared measures for the L1L2MIX model*

Our decomposition of *total* outcome variance for the L1L2MIX model[12] (provided in Appendix Table A5 using Equations [A14]-[A24] and using notation defined in Appendix Table A4) can be used to construct total R-squared measures for L1L2MIX. In particular, in the denominator of a total R-squared measure for L1L2MIX, we place the total outcome variance (Equation (A14)). Regarding the numerator of the measure, in the L1L2MIX total variance can be attributable to each of 10 possible *sources* defined in Equations (A15)-(A24) and also listed here:

- Source $f_1$=L1 predictors via marginal[13] fixed components of slopes
- Source $f_2$=L2 predictors via marginal fixed components of slopes
- Source $v_1^r$=L1 predictors via random slope variation/covariation
- Source $m^r$=cluster-specific outcome means via random intercept variation
- Source $v_1^k$=L1 predictors via across-L1-class slope variation/covariation within L2 class
- Source $v_2^k$=L2 predictors via across-L1-class slope variation/covariation within L2 class
- Source $m^k$=class-specific outcome means via across-L1-class variation within L2 class
- Source $v_1^h$=L1 predictors via across-L2-class slope variation/covariation
- Source $v_2^h$=L2 predictors via across-L2-class slope variation/covariation
- Source $m^h$=class-specific outcome means via across-L2-class variation

To construct each possible single-source total R-squared measure for the L1L2MIX, these Equations (A15)-(A24) are used, one at a time, in the numerator of the measure. See Manuscript Table 3 (Column 1) for a summary list of definitions and interpretations for all single-source total R-squared measures for the L1L2MIX. In Table 3 notation, each R-squared measure is identified and labeled by its denominator and also its numerator source. For instance, *Total* $R^2\_v_1^r$ measures the *proportion of variance attributable to level-1 predictors via random slope variation/covariation.*

In contrast to the ten potential numerator sources to which total variance can be attributable in the L1L2MIX model, recall that there were only seven such potential numerator sources for either the L1MIX or the L2MIX. There are a greater number of potential numerator sources in L1L2MIX because the existence of level-1 *and* level-2 classes in the L1L2MIX allows separately distinguishing whether predictors and outcome means are varying across L1 class, or across L2 class, or both. If a researcher wanted to jointly consider contributions via variation across L1 class vs. L2 class, the researcher could supplementarily create a combination-source measure, such as *Total* $R^2\_v_1^k$+ *Total* $R^2\_v_1^h$ or such as *Total* $R^2\_v_2^k$ +*Total* $R^2\_v_2^h$ or such as *Total* $R^2\_m^k$ + *Total* $R^2\_m^h$. As an example of a combination-source measure involving *more than two* sources, a researcher might be broadly interested in the proportion of total variance attributed to all predictors via any kind of class/cluster variation; this could be assessed using *Total* $R^2\_v_1^k$+ *Total* $R^2\_v_2^k$+ *Total* $R^2\_v_1^h$+ *Total* $R^2\_v_2^h$ .

### *Constructing and interpreting level-2 class-specific R-squared measures and kh class-combination-specific R-squared measures for the L1L2MIX model*

Whereas the multilevel mixtures with classes at only one level (L1MIX or L2MIX) each afforded only one kind of class-specific R-squared, the L1L2MIX affords two kinds of class-specific R-squareds. Specifically, we can examine how variance is accounted for within a level-2 class $h$ (via a level-2 class-specific R-squared) and can examine how variance is accounted for within the $kh$ class combination (via a $kh$ class-combination R-squared). Our decomposition of level-2 class-specific outcome variance for the L1L2MIX model is given Appendix Table A6 (Equations (A25)-(A32)) and is derived in the Online Supplement. A single-source level-2 class-$h$-specific R-squared uses Equation (A25) in the denominator and uses one or more of Equation (A26)-(A32), in the numerator. That is, there are seven potential sources to which class-$h$-specific variance can be attributable; these sources mirror the seven sources listed earlier in Appendix Table A2, with the only modification being that they are now specific to level-2 class $h$. Manuscript Table 3 (Column 4) provides a summary list defining these single-source level-2 class-specific measures for the L1L2MIX.

Our decomposition of $kh$ class-combination-specific outcome variance for the L1L2MIX is provided in Appendix Table A7 (Equations [A33]-[A37], with an accompanying derivation in the Online Supplement). The four single-source $kh$ class-combination R-squared measures use Equation (A33) in the denominator and one or more of Equations (A34)-(A37) in the numerator. The latter four numerator sources mirror the four sources listed earlier in Appendix Table A3, with the only exception being that they are now specific to class-combination $kh$. See the manuscript Table 3 (Column 7) for a list summarizing interpretations for the four single-source $kh$ class-combination-specific measures for the L1L2MIX. We next turn to an empirical example to illustrate how

[12]See Online Supplement for a detailed derivation of this decomposition of total variance for L1L2MIX.
[13]When constructing a total measure, marginal refers to an across-class weighted average. Later, when we construct a level-2-specific measure, marginal refers to an across-$k$ weighted average for class $h$. Note also that when we later construct a $kh$-class-combination measure, the marginal fixed component of the slope is simply the fixed component of the slope for class-combination $kh$.

**Table 3.** Summary of single-source measures in the R-squared framework for the L1L2MIX model.

| Total measures | Computation* | Interpretation | Level2-class-*h* specific measures | Computation* | Interpretation | *kh*-class-combination specific measures | Computation* | Interpretation |
|---|---|---|---|---|---|---|---|---|
| *Total* $R^2\_f_1$ | Eqn (A15)/ (A14) | Prop. of total variance attributable to source $f_1$ | *L2-class* $R^2\_f_1$ | Eqn (A26)/ (A25) | Prop. of class-*h* variance attributable to source $f_1$ | *kh-class-comb.* $R^2\_f_1$ | Eqn(A34)/ (A33) | Prop. of *kh*-class-comb. var. attributable to source $f_1$ |
| *Total* $R^2\_f_2$ | Eqn (A16)/ (A14) | Prop. of total variance attributable to source $f_2$ | *L2-class* $R^2\_f_2$ | Eqn (A27)/ (A25) | Prop. of class-*h* variance attributable to source $f_2$ | *kh-class-comb.* $R^2\_f_2$ | Eqn(A35/ (A33) | Prop. of *kh*-class-comb. var. attributable to source $f_2$ |
| *Total* $R^2\_v_1^r$ | Eqn (A17)/ (A14) | Prop. of total variance attributable to source $v_1{}^r$ | *L2-class* $R^2\_\ v_1{}^r$ | Eqn (A28)/ (A25) | Prop. of class-*h* variance attributable to source $v_1{}^r$ | *kh-class-comb.* $R^2\_\ v_1{}^r$ | Eqn(A36/ (A33) | Prop. of *kh*-class comb. var. attributable to source $v_1{}^r$ |
| *Total* $R^2\_m^r$ | Eqn (A18)/ (A14) | Prop. of total variance attributable to source $m^r$ | *L2-class* $R^2\_m^r$ | Eqn (A29)/ (A25) | Prop. of class-*h* variance attributable to source $m^r$ | *kh-class-comb.* $R^2\_m^r$ | Eqn(A37)/ (A33) | Prop. of *kh*-class-comb. var. attributable to source $m^r$ |
| *Total* $R^2\_v_1^k$ | Eqn (A19)/ (A14) | Prop. of total variance attributable to source $v_1{}^k$ | *L2-class* $R^2\_\ v_1{}^k$ | Eqn (A30)/ (A25) | Prop. of class-*h* variance attributable to source $v_1{}^k$ | | | |
| *Total* $R^2\_v_2^k$ | Eqn (A20)/ (A14) | Prop. of total variance attributable to source $v_2{}^k$ | *L2-class* $R^2\_\ v_2{}^k$ | Eqn (A31)/ (A25) | Prop. of class-*h* variance attributable to source $v_2{}^k$ | | | |
| *Total* $R^2\_m^k$ | Eqn (A21)/ (A14) | Prop. of total variance attributable to source $m^k$ | *L2-class* $R^2\_\ m^k$ | Eqn (A32)/ (A25) | Prop. of class-*h* variance attributable to source $m^k$ | | | |
| *Total* $R^2\_v_1^h$ | Eqn (A22)/ (A14) | Prop. of total variance attributable to source $v_1{}^h$ | | | | | | |
| *Total* $R^2\_v_2^h$ | Eqn (A23)/ (A14) | Prop. of total variance attributable to source $v_2{}^h$ | | | | | | |
| *Total* $R^2\_m^h$ | Eqn (A24)/ (A14) | Prop. of total variance attributable to source $m^h$ | | | | | | |

**Summary of Source Definitions**

Source $f_1$ =L1 predictors via marginal† fixed components of slopes
Source $f_2$ =L2 predictors via marginal† fixed components of slopes
Source $v_1{}^r$ =L1 predictors via random slope variation/covariation
Source $m^r$ =cluster-specific outcome means via random intercept variation
Source $v_1{}^k$ =L1 predictors via across-L1-class slope variation/covariation within L2 class
Source $v_2{}^k$ =L2 predictors via across-L1-class slope variation/covariation within L2 class
Source $m^k$ =class-specific outcome means via across-L1-class variation within L2 class
Source $v_1{}^h$ =L1 predictors via across-L2-class slope variation/covariation
Source $v_2{}^h$ =L2 predictors via across-L2-class slope variation/covariation
Source $m^h$ =class-specific outcome means via across-L2-class variation

*Eqn (A14)-(A24) given in Appx Table A5. Eqn (A25)-(A32) given in Appx Table A6. Eqn (A33)-(A37) given in Appx Table A7.† marginal = across-class weighted avg. For a level-2-class-specific measure, marginal = across-*k*-class weighted avg. for class *h*. For *kh*-class-combination-specific measure, the marginal fixed component of slope is just the fixed component of the slope for that *kh*-combination to interpret as an integrated set the total, level-2 class-*h*-specific, and *kh* class-combination specific R-squared measures for the L1L2MIX.

## Empirical example: Multilevel regression mixture model with classes at both level-1 and −2

This empirical example is based on a L1L2MIX application from Muthén and Asparouhov (2009, their Equations 26 & 29 on p. 652), which predicted mathematics achievement with a $K=3$, $H=2$ L1L2MIX model fit to the NELS 1988 eighth grade data (18,596 students nested within 767 schools in the public use data used here). Traditionally, interpretation of this L1L2MIX model would follow from inspecting point estimates and SEs of individual model parameters. However, because LIL2MIX models typically have many parameters (characterizing two different kinds of classes together with random effects), it becomes difficult to exclusively rely on this traditional strategy when trying to obtain an integrative substantive interpretation of model results and inform qualitative class labeling. Hence, here we aid L1L2MIX interpretability by reporting R-squared effect sizes to indicate what modeled sources are most and least important in terms of accounting for total and class-specific outcome variation.

In the empirical example L1L2MIX model, level-1 predictors of math achievement are gender (*fem*) and socioeconomic status (*ses*), and level-2 predictors are percent of teachers with an advanced degree in the school (*peradv*), school-mean socioeconomic status (*meanses*), and school-type (*priv* representing private vs. public and *cath* representing catholic vs. public). Below, Equation (6) shows how regression coefficients were allowed to vary across $k$ and $h$.

$$\begin{aligned} math_{ij|c_{ij}=k,d_j=h} &= \gamma_{00}^{b(kh)} + fem_{ij}\gamma_{10}^{w(kh)} + ses_{ij}\gamma_{20}^{w(kh)} \\ &+ peradv_j\gamma_{01}^{b(kh)} + priv_j\gamma_{02}^{b(kh)} + cath_j\gamma_{03}^{b(kh)} \\ &+ meanses_j\gamma_{04}^{b(kh)} + u_{0j} + fem_{ij}u_{1j} + ses_{ij}u_{2j} + \varepsilon_{ij} \\ &\varepsilon_{ij}\sim N(0,\theta^k) \\ &\mathbf{u}_j\sim N(\mathbf{0},\mathbf{T}) \\ p(d_j=h) &= \pi^h = \frac{\exp(\varpi^h)}{\sum_{h=1}^{H}\exp(\varpi^h)} \text{ and} \\ p(c_{ij}=k|d_j=h) &= \pi^{k|h} = \frac{\exp\left(\omega^k+\delta^{kh}\right)}{\sum_{k=1}^{K}\exp\left(\omega^k+\delta^{kh}\right)} \end{aligned} \tag{6}$$

A $K=3$, $H=2$ L1L2MIX from Equation (6) was fit using M*plus* 8.4; results indicated that an estimated 15% of schools fall into school-level class $h=1$ whereas an estimated 85% of schools fall into school-level class $h=2$. The *kh* class-combination probabilities were as follows: $k=1,h=1$ (11%), $k=2,h=1$ (4%), $k=3,h=1$ (1%), $k=1,h=2$ (25%), $k=2,h=2$ (57%), $k=3,h=2$ (2%). Parameter estimates and SEs for this L1L2MIX are provided in the Online Supplement. Inspecting these shows that (unsurprising given the large sample size) nearly all intercept and slope coefficients in each *kh* class-combination are significant, as are each variance component. These intercept and slope coefficients look to differ somewhat across classes, but it is unclear from eyeballing these results how to gauge the impact and importance of these across-class differences. Fortunately, the R-squared results for this L1L2MIX empirical example can help quantify this.

Total, level-2-class-specific, and *kh*-class-combination-specific R-squared results for this L1L2MIX are provided in the manuscript's Table 4. Here, we use the Table 4 results to illustrate how it can be useful to juxtapose and compare total vs. level-2-specific vs. class-combination-specific R-squareds for the same source of explained variance (i.e., measures with different denominators but the same numerator). Such comparisons can illuminate a situation wherein the same source contribution is meaningful within a particular class but not overall. For instance, here the variance explained by level-2 predictors via fixed component of slopes was sizable for class-combination $k=3,h=1$ (*class-combination-specific* $R^2\_f_2=.598$), but was quite small overall (*Total* $R^2\_f_2=.008$). Such comparisons can also illuminate situations wherein the same source contribution is meaningful (or not) for explaining *all* types of outcome variance. For instance, here the contribution of cluster-specific outcome means via random intercept variability, $R^2\_m^r$, is meaningful for explaining all types of outcome variance (total, level-2-specific, and *kh*-class-combination-specific) in Table 4, whereas the contribution of predictors via across-cluster slope variability, $R^2\_v_1^r$ is not meaningful for explaining any type of outcome variance in Table 4.

For this empirical example we chose to create a river plot of the level-2-class-specific R-squared results, shown in Figure 2, to help visualize and highlight instances in which source contributions differed across level-2 class to, in turn, help qualitatively characterize the level-2 classes. Figure 2 illustrates that level-2 class $h=2$ is distinguished from $h=1$ by having a greater contribution of level-1 predictors via marginal fixed component of slopes (for $h=2$ the *level-2-class specific* $R^2\_f_1=.174$ but for $h=1$ it is $R^2\_f_1=.004$). Furthermore, because class-combination R-squareds were nearly zero for all but one $k$ within $h=2$, we can infer that the $R^2\_f_1$ result for $h=2$ is mainly driven by class-combination $k=2,h=2$ (whose $R^2\_f_1=.366$ in Table 4). Figure 2 also illustrates the fact that $h=2$ is distinguished from $h=1$ by having a greater contribution of predictors via across-$k$ slope variability (for $h=2$ the *level-2-class specific* $R^2\_v_1^k=.063$ and $R^2\_v_2^k=.141$, whereas for $h=1$ $R^2\_v_1^k=.003$ and $R^2\_v_2^k=.037$).

In sum, R-squared results for this empirical example indicated that: (a) overall, slope variability mainly occurred across $k$ within $h=2$ (rather than across clusters, or across $h$, or across $k$ within $h=1$); (b) across clusters, intercept variability occurred within all classes and class-combinations; (c) level-1 predictors explained more variance marginally than level-2 predictors; and (d) the former's influence was greatest in $h=2$, and this influence in turn was dominated by class combination $k=2,h=2$.

## Software implementation

We provide an R function *reRegMixR2* that computes all possible single-source measures for the researcher's fitted L1MIX, L2MIX, or L1L2MIX model and uses as input the fitted model's parameter estimates (that could be previously obtained from any mixture modeling software program). The framework of single-source R-squared measures discussed here (Tables 1 and Tables 3) is

**Table 4.** R-squared results for $K = 3$ $H = 2$ L1L2MIX empirical example model predicting math achievement, with students nested within schools.

| Total R-squared measures | |
|---|---|
| Total $R^2\_f_1$ | 0.124 |
| Total $R^2\_f_2$ | 0.008 |
| Total $R^2\_v_1^r$ | 0.004 |
| Total $R^2\_m^r$ | 0.331 |
| Total $R^2\_v_1^k$ | 0.040 |
| Total $R^2\_v_2^k$ | 0.031 |
| Total $R^2\_m^k$ | 0.223 |
| Total $R^2\_v_1^h$ | 0.034 |
| Total $R^2\_v_2^h$ | 0.005 |
| Total $R^2\_m^h$ | 0.139 |

| Level-2-class specific R-squared measures | h = 1 | h = 2 |
|---|---|---|
| Level-2-class-specific $R^2\_f_1$ | 0.004 | 0.174 |
| Level-2-class-specific $R^2\_f_2$ | 0.063 | 0.008 |
| Level-2-class-specific $R^2\_v_1^r$ | 0.007 | 0.004 |
| Level-2-class-specific $R^2\_m^r$ | 0.546 | 0.314 |
| Level-2-class-specific $R^2\_v_1^k$ | 0.003 | 0.063 |
| Level-2-class-specific $R^2\_v_2^k$ | 0.037 | 0.141 |
| Level-2-class-specific $R^2\_m^k$ | 0.205 | 0.163 |

| *kh*-class-combination specific R-squared measures | $k = 1, h = 1$ | $k = 2$, h = 1 | $k = 3$, h = 1 | $k = 1$, h = 2 | $k = 2$, h = 2 | $k = 3$, h = 2 |
|---|---|---|---|---|---|---|
| *kh*-class-combination-specific $R^2\_f_1$ | 0.008 | 0.112 | 0.009 | 0.030 | 0.366 | 0.001 |
| *kh*-class-combination-specific $R^2\_f_2$ | 0.111 | 0.054 | 0.598 | 0.162 | 0.023 | 0.596 |
| *kh*-class-combination-specific $R^2\_v_1^r$ | 0.010 | 0.006 | 0.005 | 0.009 | 0.005 | 0.005 |
| *kh*-class-combination-specific $R^2\_m^r$ | 0.804 | 0.528 | 0.388 | 0.737 | 0.386 | 0.398 |

not currently available as an output option in existing multilevel mixture software programs.[14] Our Online Supplement provides and describes the *reRegMixR2* function, and it is also being incorporated into the r2mlm R package (Shaw, Rights, Sterba, & Flake, Revised & Resubmitted).

## Discussion

Multilevel regression mixture models are an increasingly popular approach for accommodating nested data structures in psychology and education when a combination of discretely-distributed and continuously-distributed variability in intercepts and slopes are posited. The complexity of multilevel regression mixture models necessitates moving beyond typical reliance on significance testing of individual parameter estimates together with informal qualitative characterization/labeling of classes. The R-squared framework presented here is an additional tool to aid in describing and interpreting important features of multilevel regression mixtures with random effects (i.e., L1MIX, L2MIX or L1L2MIX) in a cohesive, integrated manner.

### Summary of contribution

One way in which the framework of R-squared measures developed here (which is summarized in manuscript Tables 1 and Tables 3) extends previous work is that it is newly applicable to multilevel regression mixtures with *random effects*. In contrast, Rights and Sterba (2018) R-squared measures were for single- or multi-level regression mixtures with locally-independent classes, and thus cannot be used when random effects are specified within class. As such, Rights and Sterba (2018) did not include sources $v_1^r$ and $m_r$ in their variance decompositions, and did not include measures representing sources $v_1^r$ and $m_r$ (i.e., they did not include *total* $R^2\_v_1^r$ and $R^2\_m^r$, *level-2-class-specific* $R^2\_v_1^r$ and $R^2\_m^r$, and *kh-class-combination-specific* $R^2\_v_1^r$ and $R^2\_m^r$).

Another way in which the framework developed here extends this previous work is by unconflating the contribution of between- versus within-cluster fixed components of level-1 predictor slopes. Previously, Rights and Sterba (2018) conflated the contribution of $f_1$ versus $f_2$ sources in their variance decompositions and measures (i.e., their 2018 suite of measures did not distinguish *total* $R^2\_f_1$ versus $R^2\_f_2$, *level-2 class-specific* $R^2\_f_1$ versus $R^2\_f_2$, or *kh-class-combination-specific* $R^2\_f_1$ versus $R^2\_f_2$, as was newly done here). The present framework is advantageous because it allows separately assessing and comparing the variance accounted for by sources $f_1$ versus $f_2$ – as recommended here – while still allowing the optional, supplemental creation of the earlier-used combination-source measures that combine contributions of sources $f_1$ and $f_2$ .

### How the present $R^2$ framework simplifies if there are either no classes or no random effects

Though general enough to accommodate models with both classes and random effects, the present framework is nonetheless applicable to simplified specifications that either

[14]See software review in Rights and Sterba (2018); for select multilevel regression mixtures, M*plus* and Latent GOLD (Vermunt & Magidson, 2013) provide a combination-source measure analogous to a *class-specific* $R^2\_f$, however, theirs is not able to be decomposed into $R^2\_f_1 + R^2\_f_2$ and is not accompanied by the other class-specific single-sources measures discussed here in Tables 1 and Tables 3. Latent GOLD also provides a total combination-source measure that is again not able to be decomposed into its constituent *f*, *v*, and *m* sources (which limits interpretability because these constituent sources *f*, *v*, and *m* substantively mean very different things).

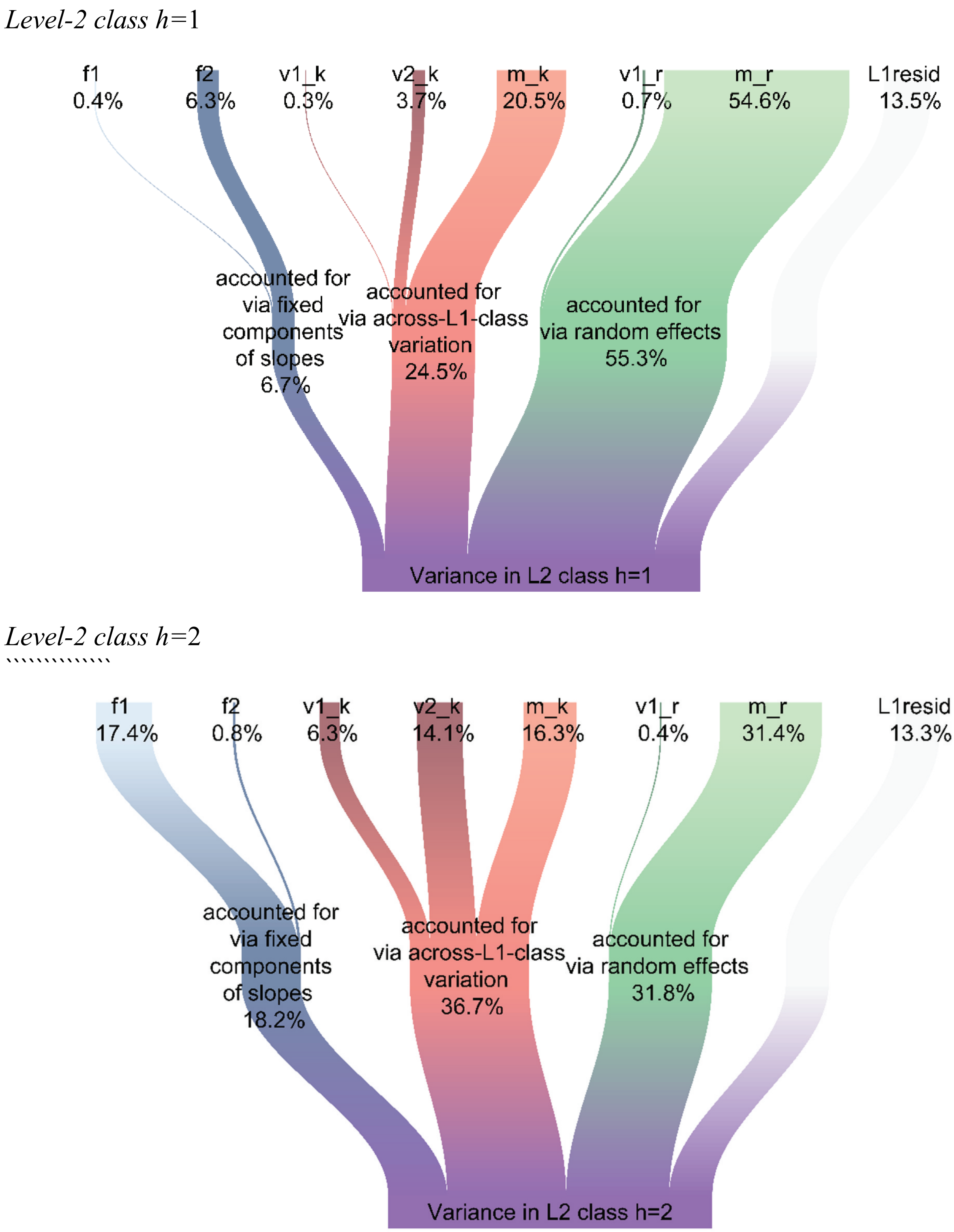


**Figure 2.** River plot depicting the single-source level-2 class-specific R-squared measures for the $K = 3, H = 2$ L1L2MIX model predicting mathematics achievement for students nested within schools .

only include random effects (i.e., there is only one class) or only include classes (i.e., there are no random effects), as follows. If there is only one class ($K$ =1 and $H$ =1), then (a) the L1MIX, L2MIX, and L1L2MIX would simplify to a conventional (i.e., non-mixture) multilevel model, (b) the potential sources of explained total variance would simplify to no longer include $v_1^k, v_2^k, m^k, v_1^h, v_2^h, m^h$ (therefore *total* $R^2\_v_1^k = R^2\_v_2^k = R^2\_m^k = R^2\_v_1^h = R^2\_v_2^h = R^2\_m^h = 0$), and (c) the total measures that can be constructed from remaining sources $f_1, f_2, v_1^r, m^r$ in Table 1 (Column 4) and Table 3 (Column 7) would match those total measures in the existing R-squared framework for *conventional* (i.e., non-mixture) multilevel models from Rights and Sterba (2019). If, on the other hand, there are only classes and no random effects, then (a) the L1MIX, L2MIX, and L1L2MIX would simplify to locally-independent single-level (for L1MIX) or multilevel (for L2MIX, L1L2MIX) regression mixture models, and (b) the potential sources of explained variance would no longer include $v_1^r$ and $m_r$ (and therefore *Total* $R^2\_v_1^r = R^2\_m^r = 0$ and *level-2-class-specific* $R^2\_v_1^r = R^2\_m^r = 0$ and *kh-class-combination-specific* $R^2\_v_1^r = R^2\_m^r = 0$).

### Limitations and future directions

In the current paper we focus on within-class models that are linear and have normally distributed within-class residuals. Such normal mixtures are the most common kind of mixture applied in practice (see review in Sterba et al., 2012). Future research could consider extensions of this framework of R-squared measures to accommodate alternative within-class level-1 residual distributions for generalized multilevel mixtures. The current framework also pertains to two levels of hierarchical nesting, and pertains to quantifying explained variance in a single outcome variable at a time. Future work could consider extensions to higher levels of nesting, to nonhierarchical nesting or partial nesting, and to multivariate outcomes.

Each R-squared measure in the framework presented here can, in theory, be estimated with a different degree of precision, which would be useful to report. In particular, confidence intervals can be used to convey the precision of effect size measures, such as R-squared measures (e.g., Kelley, 2007). However, confidence interval construction for R-squared measures—even for conventional (i.e., non-mixture) multilevel models—has been largely unstudied and has been identified as a gap in need of future research (LaHuis et al., 2014; Nakagawa & Schielzth, 2013; Rights & Sterba, 2019). Although bootstrapping has been suggested for constructing confidence intervals for R-squared measures in conventional multilevel models, there are multiple ways to bootstrap nested data (Goldstein, 2011) and an evaluation of the relative merit of each approach necessitates further methodological study. For instance, alternative possibilities include resampling level-1 units without regard to cluster membership, resampling clusters while keeping units within clusters intact, resampling level-1 units within each cluster for the original set of clusters, resampling clusters and then resampling units within each cluster, or resampling level-1 and level-2 residuals. Utilizing bootstrapping for confidence interval construction in the context of mixture models (e.g., Cole & Bauer, 2016) holds promise, but even in the context of single-level mixture models with no random effects it poses inherent challenges from poorly separated classes (due to the label switching problem manifesting across bootstrap resamples; Taushanov & Berchtold, 2019) and from outlying cases (whose influence can be amplified in certain bootstrap resamples; Jaki et al., 2018). Hence, the development and evaluation of bootstrap approaches for constructing confidence intervals for R-squared measures in the context of multilevel mixtures with random effects is an area recommended for future methodological work.

### Recommendations and conclusions

For applied researchers interested in using R-squared measures to help describe and interpret their fitted multilevel mixture, we recommend constructing the set of single-source total and class-specific R-squareds in the present framework as illustrated empirically in Tables 2 and Tables 4. For the most complete understanding, we suggest considering these measures in juxtaposition with each other, rather than pinpointing one single-source measure or combination-source measure to look at in isolation. In conclusion, we hope this paper, and its accompanying software, serves to increase access to and awareness of R-squared effect size measures relevant to popular multilevel mixtures with both classes and random effects.

## Funding

Jason Rights was supported by a Canadian Natural Science and Engineering Research Council (NSERC) Discovery Grant [RGPIN-2020-06132] and Supplement [DGECR-2020-00362].

## References

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

American Psychological Association. (2009). *Publication manual of the American Psychological Association* (6th ed.).

Applebaum, M., Cooper, H., Kline, R., Mayo-Wilson, E., Nezu, A. M., & Rao, S. (2018). Journal article reporting standards for quantitative research in psychology: The APA Publications and Communications Board task force report. *American Psychologist*, *73*, 3–25. https://doi.org/10.1037/amp0000191

Barstead, M. (2019, January 19). *Visualizing variance in multilevel models using the riverplot package [Blog post]*. Retrieved from https://www.deadreckoning.consulting/blog/2019/01/19/2019-01-19-visualizing-variance-in-multilevel-models-using-the-riverplot-package/

Cho, S., & Lee, J. R. (2018). Impacts of low self-control and delinquent peer associations on bullying growth trajectories among Korean youth: A latent growth mixture modeling approach. *Journal of Interpersonal Violence*, *36*, 4139–4169. https://doi.org/10.1177/0886260518786495

Cole, V. T., & Bauer, D. J. (2016). A note on the use of mixture models for individual prediction. *Structural Equation Modeling*, *23*, 615–631. https://doi.org/10.1080/10705511.2016.1168266

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

Curran, P. J., & Bauer, D. J. (2011). The disaggregation of within-person and between-person effects in longitudinal models of change. *Annual Review of Psychology*, *62*, 583–619. https://doi.org/10.1146/annurev.psych.093008.100356

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*, 6137–6157. https://doi.org/10.1002/sim.3429

Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods*, *12*, 121–138. https://doi.org/10.1037/1082-989X.12.2.121

Flunger, B., Trautwein, U., Nagengast, B., Ludke, O., Niggli, A., & Schnyder, I. (2019). Using multilevel mixture models in educational research: An illustration with homework research. *The Journal of Experimental Education*, *89*, 209–236. https://doi.org/10.1080/00220973.2019.1652137

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

Halliday-Boykins, C. A., Henggeler, S. W., Rowland, M. D., & DeLucia, C. (2004). Heterogeneity in youth symptom trajectories following psychiatric crisis: Predictors and placement outcomes. *Journal of Consulting and Clinical Psychology*, *72*, 993–1003. https://doi.org/10.1037/0022-006X.72.6.993

Harlow, L. L., Muliak, S. A., & Steiger, J. H. 1997. *What if there were no significance tests?* Eds.. Erlbaum.

Heck, R. H., & Thomas, S. L. (2015). *An introduction to Multilevel Modeling Techniques: MLM and SEM approaches using Mplus (Third Edition)*. Routledge.

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

Hofmann, D. A., & Gavin, M. B. (1998). Centering decisions in hierarchical linear models: Implications for research in organizations. *Journal of Management*, *24*, 623–641. https://doi.org/10.1177/014920639802400504

Hox, J. J. (2010). *Multilevel analysis: Techniques and applications*. Routledge.

Jaki, T., Su, T.-L., Kim, M., & Van Horn, M. L. (2018). An evaluation of the bootstrap for model validation in mixture models. *Statistical Simulation and Computation*, *47*, 1028–1038. https://doi.org/10.1080/03610918.2017.1303726

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

Kelley, K. (2007). Confidence intervals for standardized effect sizes: Theory, application, and implementation. *Journal of Statistical Software, 20*, 1–24. https://doi.org/10.18637/jss.v020.i08

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

Khalili, A., & Chen, J. (2012). Variable selection in finite mixture of regression models. *Journal of the American Statistical Association*, *102*, 1025–1038. https://doi.org/10.1198/016214507000000590

Konte, M. (2016). The effects of remittances on support for democracy in Africa: Are remittances a curse or a blessing? *Journal of Comparative Economics*, *44*, 1002–1022. https://doi.org/10.1016/j.jce.2016.02.004

Kreft, I. G., & de Leeuw, J. (1998). *Introducing multilevel modeling*. Sage.

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

Lukočienė, O., Varriale, R., & Vermunt, J. K. (2010). The simultaneous decision(s) about the number of lower- and higher-level classes in multilevel latent class analysis. *Sociological Methodology*, *40*, 247–283. https://doi.org/10.1111/j.1467-9531.2010.01231.x

McLachlan, G., & Peel, D. (2000). *Finite mixture models*. Wiley.

Morin, A. J. S., & Marsh, H. W. (2015). Person-centered analyses: An illustration based on university teachers' multidimensional profiles of effectiveness. *Structural Equation Modeling*, *22*, 39–59. https://doi.org/10.1080/10705511.2014.919825

Muthén, B. (2001). Latent variable mixture modeling. In G. A. Marcoulides & R. E. Schumacker (Eds.), *New developments and techniques in structural equation modeling* (pp. 1–33). Erlbaum.

Muthén, B. O. (2007). Latent variable hybrids: Overview of new and old models. In G. R. Handcock & K. M. Samuelsen Eds., *Advances in latent variable mixture models*. Muthén, B., & Asparouhov, T (2009). (1–24). Information Age Publishing.

Muthén, B. O., & Asparouhov, T. (2009). Multilevel regression mixture analysis. *Journal of the Royal Statistical Society: Series A*, *172*, 639–657. https://doi.org/10.1111/j.1467-985X.2009.00589.x

Muthén, L. K., & Muthén, B. O. (1998-2020). *Mplus User's Guide: Version 8.4*. Muthén & Muthén.

Nagin, D. (2005). *Group-based modeling of development*. Harvard University Press.

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. https://doi.org/10.1111/j.2041-210x.2012.00261.x

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

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

Raftery, A. E., & Dean, N. (2006). Variable selection for model-based clustering. *Journal of the American Statistical Association*, 101, 168–178. https://doi.org/10.1198/016214506000000113

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

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*, 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*, 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*, 309–338. https://doi.org/10.1037/met0000184

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 (Special Issue on Developmental Methods)*, *2021*, 65–110. https://doi.org/10.1002/cad.20387

Rights, J. D., & Sterba, S. K. (Accepted). R-squared measures for multilevel models with three or more levels. *Multivariate Behavioral Research*.

Shaw, M., Rights, J. D., Sterba, S. K., & Flake, J. (Revised & Resubmitted). r2mlm: R-Squared measures for multilevel models. *Behavioral Research Methods*.

Sher, K. J., Jackson, K. M., & Steinley, D. (2011). Alcohol use trajectories and the ubiquitous cat's cradle: Cause for concern? *Journal of Abnormal Psychology*, *120*, 322–335. https://doi.org/10.1037/a0021813

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

Soloski, K. L., & Durtschi, J. A. (2019). Identifying different ways people change: A latent basis growth mixture model example identifying nonlinear trajectories of binge drinking. *Journal of Marital and Family Therapy*, *46*, 638–660. https://doi.org/10.1111/jmft.12382

Sterba, S. K. (2013). Understanding linkages among mixture models. *Multivariate Behavioral Research*, *48*, 775–815. https://doi.org/10.1080/00273171.2013.827564

Sterba, S. K. (2014). Handling missing covariates in conditional mixture models under missing at random assumptions. *Multivariate Behavioral Research*, *49*, 614–632. https://doi.org/10.1080/00273171.2014.950719

Sterba, S. K., Baldasaro, R. E., & Bauer, D. J. (2012). Factors affecting the adequacy and preferability of semiparametric groups-based approximations of continuous growth trajectories. *Multivariate Behavioral Research*, *47*, 590–634. https://doi.org/10.1080/00273171.2012.692639

Sterba, S. K., & Bauer, D. J. (2010). Matching method with theory in person-oriented developmental psychopathology research. *Development and Psychopathology*, *22*, 239–254. https://doi.org/10.1017/S0954579410000015

Sterba, S. K., & Bauer, D. J. (2014). Predictions of individual change recovered with latent class or random coefficient growth models. *Structural Equation Modeling*, *21*, 342–360. https://doi.org/10.1080/10705511.2014.915189

Taushanov, Z., & Berchtold, A. (2019). Bootstrap validation of the estimated parameters in mixture models used for clustering. *Journal of the French Statistical Society*, *160*, 114–129.

Titterington, D. M., Smith, A. F., & Makov, U. E. (1985). *Statistical analysis of finite mixture distributions*. Wiley.

Van Horn, M. L., Feng, Y., Kim, M., Lamont, A., Feaster, D., & Jaki, T. (2016). Using multilevel regression mixture models to identify level-1 heterogeneity in level-2 effects. *Structural Equation Modeling*, *23*, 259–269. https://doi.org/10.1080/10705511.2015.1035437

Vermunt, J. K. (2004). An EM algorithm for the estimation of parametric and nonparametric hierarchical nonlinear models. *Statistica Neerlandica*, *58*, 220–233. https://doi.org/10.1046/j.0039-0402.2003.00257.x

Vermunt, J. K., & Magidson, J. (2005). Hierarchical mixture models for nested data structures. In W. Weihs & G. Wolfgang (Eds.), *Classification—the ubiquitous challenge* (pp. 240–247). Springer.

Vermunt, J. K. (2008). Latent class and finite mixture models for multilevel data sets. *Statistical Methods in Medical Research*, *17*, 33–51. https://doi.org/10.1177/0962280207081238

Vermunt, J. K., & Magidson, J. (2013). *Technical guide for Latent GOLD 5.0: Basic, advanced, and syntax*. Statistical Innovations.

Vermunt, J. K., & Van Dijk, L. (2001). A nonparametric random-coefficients approach: The latent class regression model. *Multilevel Modelling Newsletter*, *13*, 6–13.

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

Wedel, M., & Kamakura, W. A. (1998). *Market segmentation: Concepts and methodological foundations*. Kluwer Academic Publishers.

Weiner, J. (2015). *riverplot: R package version 0.5*.

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

Yampolskaya, S., Armstrong, M. I., & King-Miller, T. (2011). Contextual and individual-level predictors of abused children's reentry into out-of-home care: A multilevel mixture survival analysis. *Child Abuse & Neglect*, *35*, 670–67. https://doi.org/10.1016/j.chiabu.2011.05.005

## Appendix

**Table A1.** Definition of symbols in the model-implied total outcome variance expression (Eqn. A1 in Appx. Table 2) for a multilevel regression mixture model with classes at only one level (i.e., L1MIX or L2MIX).*

| Symbol* | Definition | Computation* |
|---|---|---|
| $\boldsymbol{\gamma}^{w(\cdot)}$ | Marginalized within-cluster regression coefficients (weighted across-class avg) | $\sum_{h=1}^{H} \pi^h \boldsymbol{\gamma}^{w(h)}$ |
| $\boldsymbol{\Phi}^w$ | Covariance matrix of all elements of $\mathbf{x}_{ij}^w$ across all observations | $E\left[\mathbf{x}_{ij}^w \mathbf{x}_{ij}^{w\prime}\right] - E\left[\mathbf{x}_{ij}^w\right] E\left[\mathbf{x}_{ij}^{w\prime}\right]$ |
| $\mathbf{Z}^w$ | Covariance matrix of within-cluster regression coefficients across class | $\sum_{h=1}^{H} \pi^h \boldsymbol{\gamma}^{w(h)} \boldsymbol{\gamma}^{w(h)\prime} - \left(\sum_{h=1}^{H} \pi^h \boldsymbol{\gamma}^{w(h)}\right)\left(\sum_{h=1}^{H} \pi^h \boldsymbol{\gamma}^{w(h)\prime}\right)$ |
| $\boldsymbol{\gamma}^{b(\cdot)}$ | Marginalized within-cluster regression coefficients (weighted across-class average) | $\sum_{h=1}^{H} \pi^h \boldsymbol{\gamma}^{b(h)}$ |
| $\boldsymbol{\Phi}^b$ | Covariance matrix of all elements of $\mathbf{x}_{ij}^b$ across all observations | $E\left[\mathbf{x}_{ij}^b \mathbf{x}_{ij}^{b\prime}\right] - E\left[\mathbf{x}_{ij}^b\right] E\left[\mathbf{x}_{ij}^{b\prime}\right]$ |
| $\mathbf{Z}^b$ | Covariance matrix of between-cluster regression coefficients across class | $\sum_{h=1}^{H} \pi^h \boldsymbol{\gamma}^{b(h)} \boldsymbol{\gamma}^{b(h)\prime} - \left(\sum_{h=1}^{H} \pi^h \boldsymbol{\gamma}^{b(h)}\right)\left(\sum_{h=1}^{H} \pi^h \boldsymbol{\gamma}^{b(h)\prime}\right)$ |
| $\mathbf{m}^b$ | Vector of means of elements of $\mathbf{x}_{ij}^b$ | $E\left[\mathbf{x}_{ij}^b\right]$ |
| $\mathbf{T}^{\cdot}$ | Marginalized random effect covariance matrix (weighted across-class average). | $\sum_{h=1}^{H} \pi^h \mathbf{T}^h$ |
| $\boldsymbol{\Phi}^r$ | Covariance matrix of all elements of $\mathbf{w}_{ij}$ across all observations | $E\left[\mathbf{w}_{ij}\mathbf{w}_{ij}{}'\right] - E\left[\mathbf{w}_{ij}\right] E\left[\mathbf{w}_{ij}{}'\right]$ |
| $\tau_{00}^{\cdot}$ | Marginalized random intercept variance (weighted across-class average). | $\sum_{h=1}^{H} \pi^h \tau_{00}^h$ |
| $\theta^{\cdot}$ | Marginalized level-1 residual variance (weighted across-class avg) | $\sum_{h=1}^{H} \pi^h \theta^h$ |

*Equations and symbols are shown for L2MIX. For L1MIX, replace $h$, $H$, $\mathbf{T}^{\cdot}$, $\dot{\tau}_{00}$ , $\mathbf{Z}$ with $k$, $K$, $\mathbf{T}$, $\tau_{00}$, $\boldsymbol{\Omega}$, respectively. Dot superscript means "averaged across class."

**Table A2.** Constructing *total* $R^2$ measures for multilevel regression mixtures with classes at only one level (i.e., L1MIX or L2MIX model) using the variance decomposition derived in the Online Supplement and using the symbols defined in Appendix Table A1.

| Constituent parts to construct *total* $R^2$s | Definition‡ | Eqn. # | Equation* |
|---|---|---|---|
| Denominator of *total* $R^2$ | *Total* outcome variance | Eqn (A1) | $\text{var}(y_{ij\mid d_j=h}) = \text{var}(\mathbf{x}_{ij}^{w\prime}\boldsymbol{\gamma}^{w(h)} + \mathbf{x}_j^{b\prime}\boldsymbol{\gamma}^{b(h)} + \mathbf{w}'_{ij}\mathbf{u}_j + \varepsilon_{ij})$ $= \boldsymbol{\gamma}^{w(\bullet)\prime}\boldsymbol{\Phi}^w\boldsymbol{\gamma}^{w(\bullet)} + \boldsymbol{\gamma}^{b(\bullet)\prime}\boldsymbol{\Phi}^b\boldsymbol{\gamma}^{b(\bullet)} + tr(\mathbf{T}^\bullet\boldsymbol{\Phi}^r) + \tau_{00}^\bullet$ $+ tr(\mathbf{Z}^w\boldsymbol{\Phi}^w) + tr(\mathbf{Z}^b\boldsymbol{\Phi}^b) + \mathbf{m}^{b\prime}\mathbf{Z}^b\mathbf{m}^b + \theta^\bullet$ |
| Numerator of *total* $R^2$: Each term is included one at a time in the numerator of single-source measures | Variance attributable to source $f_1$ | Eqn (A2) | $\boldsymbol{\gamma}^{w(\bullet)\prime}\boldsymbol{\Phi}^w\boldsymbol{\gamma}^{w(\bullet)}$ |
| | Variance attributable to source $f_2$ | Eqn (A3) | $\boldsymbol{\gamma}^{b(\bullet)\prime}\boldsymbol{\Phi}^b\boldsymbol{\gamma}^{b(\bullet)}$ |
| | Variance attributable to source $v_1^r$ | Eqn (A4) | $tr(\mathbf{T}^\bullet\boldsymbol{\Phi}^r)$ |
| | Variance attributable to source $m^r$ | Eqn (A5) | $\tau_{00}^\bullet$ |
| | Variance attributable to source $v_1^h$ | Eqn (A6) | $tr(\mathbf{Z}^w\boldsymbol{\Phi}^w)$ |
| | Variance attributable to source $v_2^h$ | Eqn (A7) | $tr(\mathbf{Z}^b\boldsymbol{\Phi}^b)$ |
| | Variance attributable to source $m^h$ | Eqn (A8) | $\mathbf{m}^{b\prime}\mathbf{Z}^b\mathbf{m}^b$ |

‡=Sources ($f_1$, $f_2$, $v_1^h$, $v_2^h$, $m^h$, $v_1^r$ $m^r$) are defined in the manuscript text and also in the manuscript's Table 1. *Equations are shown for L2MIX; for L1MIX, replace $h$, $H$, $d_j$, $\mathbf{T}^\bullet$, $\tau_{00}^\bullet$, and $\mathbf{Z}$ with $k$, $K$, $c_{ij}$, $\mathbf{T}$, $\tau_{00}$, and, $\boldsymbol{\Omega}$, respectively.

**Table A3.** Constructing *class-specific* $R^2$ measures for multilevel regression mixtures with classes at only one level (i.e., L1MIX or L2MIX) using the variance decomposition derived in the Online Supplement and using the symbols defined in Appendix Table A1.

| Constituent parts to construct *class-specific* $R^2$s | Definition‡ | Eqn. # | Equation* |
|---|---|---|---|
| Denominator of *class-specific* $R^2$ | *Class-specific* outcome variance | Eqn (A9) | $\text{var}_{ij\mid h}(y_{ij\mid d_j=h}) = \text{var}_{ij\mid h}(\mathbf{x}_{ij}^{w\prime}\boldsymbol{\gamma}^{w(h)} + \mathbf{x}_j^{b\prime}\boldsymbol{\gamma}^{b(h)} + \mathbf{w}'_{ij}\mathbf{u}_j + \varepsilon_{ij})$ $= \boldsymbol{\gamma}^{w(h)\prime}\boldsymbol{\Phi}^w\boldsymbol{\gamma}^{w(h)} + \boldsymbol{\gamma}^{b(h)\prime}\boldsymbol{\Phi}^b\boldsymbol{\gamma}^{b(h)} + tr(\mathbf{T}^h\boldsymbol{\Phi}^r) + \tau_{00}^h + \theta^h$ |
| Numerator of *class-specific* $R^2$: Each term is included one at a time in the numerator of single-source measures | Variance attributable to source $f_1$ | Eqn (A10) | $\boldsymbol{\gamma}^{w(h)\prime}\boldsymbol{\Phi}^w\boldsymbol{\gamma}^{w(h)}$ |
| | Variance attributable to source $f_2$ | Eqn (A11) | $\boldsymbol{\gamma}^{b(h)\prime}\boldsymbol{\Phi}^b\boldsymbol{\gamma}^{b(h)}$ |
| | Variance attributable to source $v_1^r$ | Eqn (A12) | $tr(\mathbf{T}^h\boldsymbol{\Phi}^r)$ |
| | Variance attributable to source $m^r$ | Eqn (A13) | $\tau_{00}^h$ |

‡= Sources ($f_1$,$f_2$,$v_1^r$ $m^r$) are defined in the manuscript text and also in the manuscript's Table 1.
*Equations are shown for L2MIX; for L1MIX, replace $h$, $H$, $d_j$, $\mathbf{T}^h$, and $\tau_{00}^h$ with $k$, $K$, $c_{ij}$,, $\mathbf{T}$, and $\tau_{00}$ respectively.

**Table A4.** Definition of symbols in the population model-implied total outcome variance expression (Eqn A14 in Appx. Table A5) for the multilevel regression mixture with classes at both levels (L1L2MIX).

| Symbol | Definition | Computation |
|---|---|---|
| $\boldsymbol{\gamma}^{w(\bullet\bullet)}$ | Marginalized within-cluster regression coefficients (weighted across-class-combination average) | $\sum_{h=1}^{H}\sum_{k=1}^{K}\pi^{kh}\boldsymbol{\gamma}^{w(kh)}$ |
| $\boldsymbol{\Phi}^{w}$ | Covariance matrix of all elements of $\mathbf{x}_{ij}^{w}$ across all observations | $E\left[\mathbf{x}_{ij}^{w}\mathbf{x}_{ij}^{w\prime}\right] - E\left[\mathbf{x}_{ij}^{w}\right]E\left[\mathbf{x}_{ij}^{w\prime}\right]$ |
| $\boldsymbol{\Omega}^{w}$ | Covariance matrix of within-cluster regression coefficients across L1 class within L2 class | $\sum_{h=1}^{H}\sum_{k=1}^{K}\pi^{kh}(\boldsymbol{\gamma}^{w(kh)} - \sum_{k=1}^{K}\pi^{kh}\boldsymbol{\gamma}^{w(kh)})(\boldsymbol{\gamma}^{w(kh)} - \sum_{k=1}^{K}\pi^{kh}\boldsymbol{\gamma}^{w(kh)})' - \left(\sum_{h=1}^{H}\sum_{k=1}^{K}\pi^{kh}(\boldsymbol{\gamma}^{w(kh)} - \sum_{k=1}^{K}\pi^{kh}\boldsymbol{\gamma}^{w(kh)})\right)\left(\sum_{h=1}^{H}\sum_{k=1}^{K}\pi^{kh}(\boldsymbol{\gamma}^{w(kh)} - \sum_{k=1}^{K}\pi^{kh}\boldsymbol{\gamma}^{w(kh)})'\right)$ |
| $\mathbf{Z}^{w}$ | Covariance matrix of within-cluster regression coefficients across L2 class | $\sum_{h=1}^{H}\pi^{h}\left(\sum_{k=1}^{K}\pi^{kh}\boldsymbol{\gamma}^{w(kh)}\right)\left(\sum_{k=1}^{K}\pi^{kh}\boldsymbol{\gamma}^{w(kh)\prime}\right) - \left(\sum_{h=1}^{H}\pi^{h}\sum_{k=1}^{K}\pi^{kh}\boldsymbol{\gamma}^{w(kh)}\right)\left(\sum_{h=1}^{H}\pi^{h}\sum_{k=1}^{K}\pi^{kh}\boldsymbol{\gamma}^{w(kh)\prime}\right)$ |
| $\boldsymbol{\gamma}^{b(\bullet\bullet)}$ | Marginalized within-cluster regression coefficients (weighted across-class-combination average) | $\sum_{h=1}^{H}\sum_{k=1}^{K}\pi^{kh}\boldsymbol{\gamma}^{b(kh)}$ |
| $\boldsymbol{\Phi}^{b}$ | Covariance matrix of all elements of $\mathbf{x}_{ij}^{b}$ across all observations | $E\left[\mathbf{x}_{ij}^{b}\mathbf{x}_{ij}^{b\prime}\right] - E\left[\mathbf{x}_{ij}^{b}\right]E\left[\mathbf{x}_{ij}^{b\prime}\right]$ |
| $\boldsymbol{\Omega}^{b}$ | Covariance matrix of between-cluster regression coefficients across L1 class within L2 class | $\sum_{h=1}^{H}\sum_{k=1}^{K}\pi^{kh}(\boldsymbol{\gamma}^{b(kh)} - \sum_{k=1}^{K}\pi^{kh}\boldsymbol{\gamma}^{b(kh)})(\boldsymbol{\gamma}^{b(kh)} - \sum_{k=1}^{K}\pi^{kh}\boldsymbol{\gamma}^{b(kh)})' - \left(\sum_{h=1}^{H}\sum_{k=1}^{K}\pi^{kh}(\boldsymbol{\gamma}^{b(kh)} - \sum_{k=1}^{K}\pi^{kh}\boldsymbol{\gamma}^{b(kh)})\right)\left(\sum_{h=1}^{H}\sum_{k=1}^{K}\pi^{kh}(\boldsymbol{\gamma}^{b(kh)} - \sum_{k=1}^{K}\pi^{kh}\boldsymbol{\gamma}^{b(kh)})'\right)$ |
| $\mathbf{Z}^{b}$ | Covariance matrix of between-cluster regression coefficients across L2 class | $\sum_{h=1}^{H}\pi^{h}\left(\sum_{k=1}^{K}\pi^{kh}\boldsymbol{\gamma}^{w(kh)}\right)\left(\sum_{k=1}^{K}\pi^{kh}\boldsymbol{\gamma}^{b(kh)\prime}\right) - \left(\sum_{h=1}^{H}\pi^{h}\sum_{k=1}^{K}\pi^{kh}\boldsymbol{\gamma}^{b(kh)}\right)\left(\sum_{h=1}^{H}\pi^{h}\sum_{k=1}^{K}\pi^{kh}\boldsymbol{\gamma}^{b(kh)\prime}\right)$ |
| $\mathbf{m}^{b}$ | Vector of means of elements of $\mathbf{x}_{ij}^{b}$ | $E\left[\mathbf{x}_{ij}^{b}\right]$ |
| $\mathbf{T}^{\bullet}$ | Marginalized random effect covariance matrix (weighted across-L2 -class average) | $\sum_{h=1}^{H}\pi^{h}\mathbf{T}^{h}$ |
| $\boldsymbol{\Phi}^{r}$ | Covariance matrix of all elements of $\mathbf{w}_{ij}$ across all observations | $E\left[\mathbf{w}_{ij}\mathbf{w}_{ij}'\right] - E\left[\mathbf{w}_{ij}\right]E\left[\mathbf{w}_{ij}'\right]$ |
| $\tau_{00}^{\bullet}$ | Marginalized random intercept variance (weighted across-L2-class average) | $\sum_{h=1}^{H}\pi^{h}\tau_{00}^{h}$ |
| $\theta^{\bullet\bullet}$ | Marginalized L1 residual variance (weighted across-class-combination average) | $\sum_{h=1}^{H}\sum_{k=1}^{K}\pi^{kh}\theta^{kh}$ |

**Table A5.** Constructing *total* $R^2$ measures for a multilevel regression mixture with classes at both levels (L1L2MIX) using the variance decomposition derived in the Online Supplement and using the symbols defined in Appendix Table A4.

| Constituent parts to construct *total* $R^2$s | Definition‡ | Eqn. # | Equation |
|---|---|---|---|
| Denominator of *total* $R^2$ | *Total* outcome variance | Eqn (A14) | $\text{var}(y_{ij\mid c_{ij}=k,d_j=h}) = \text{var}(\mathbf{x}_{ij}^{w\prime}\boldsymbol{\gamma}^{w(kh)} + \mathbf{x}_{j}^{b\prime}\boldsymbol{\gamma}^{b(kh)} + \mathbf{w}_{ij}'\mathbf{u}_{j} + \varepsilon_{ij}) = \boldsymbol{\gamma}^{w(\bullet\bullet)\prime}\boldsymbol{\Phi}^{w}\boldsymbol{\gamma}^{w(\bullet\bullet)} + \boldsymbol{\gamma}^{b(\bullet\bullet)\prime}\boldsymbol{\Phi}^{b}\boldsymbol{\gamma}^{b(\bullet\bullet)} + \text{tr}(\mathbf{T}^{\bullet}\boldsymbol{\Phi}^{r}) + \tau_{00}^{\bullet\bullet} + tr(\boldsymbol{\Omega}^{w}\boldsymbol{\Phi}^{w}) + tr(\boldsymbol{\Omega}^{b}\boldsymbol{\Phi}^{b}) + \mathbf{m}^{b\prime}\boldsymbol{\Omega}^{b}\mathbf{m}^{b} + tr(\mathbf{Z}^{w}\boldsymbol{\Phi}^{w}) + tr(\mathbf{Z}^{b}\boldsymbol{\Phi}^{b}) + \mathbf{m}^{b\prime}\mathbf{Z}^{b}\mathbf{m}^{b} + \theta^{\bullet\bullet}$ |
| Numerator of *total* $R^2$: Each term is included one at a time in the numerator of single-source measures | Variance attributable to source $f_1$ | Eqn (A15) | $\boldsymbol{\gamma}^{w(\bullet\bullet)\prime}\boldsymbol{\Phi}^{w}\boldsymbol{\gamma}^{w(\bullet\bullet)}$ |
| | Variance attributable to source $f_2$ | Eqn (A16) | $\boldsymbol{\gamma}^{b(\bullet\bullet)\prime}\boldsymbol{\Phi}^{b}\boldsymbol{\gamma}^{b(\bullet\bullet)}$ |
| | Variance attributable to source $v_1^r$ | Eqn (A17) | $tr(\mathbf{T}^{\bullet}\boldsymbol{\Phi}^{r})$ |
| | Variance attributable to source $m^r$ | Eqn (A18) | $\tau_{00}^{\bullet}$ |
| | Variance attributable to source $v_1^k$ | Eqn (A19) | $tr(\boldsymbol{\Omega}^{w}\boldsymbol{\Phi}^{w})$ |
| | Variance attributable to source $v_2^k$ | Eqn (A20) | $tr(\boldsymbol{\Omega}^{b}\boldsymbol{\Phi}^{b})$ |
| | Variance attributable to source $m^k$ | Eqn (A21) | $\mathbf{m}^{b\prime}\boldsymbol{\Omega}^{b}\mathbf{m}^{b}$ |
| | Variance attributable to source $v_1^h$ | Eqn (A22) | $tr(\mathbf{Z}^{w}\boldsymbol{\Phi}^{w})$ |
| | Variance attributable to source $v_2^h$ | Eqn (A23) | $tr(\mathbf{Z}^{b}\boldsymbol{\Phi}^{b})$ |
| | Variance attributable to source $m^h$ | Eqn (A24) | $\mathbf{m}^{b\prime}\mathbf{Z}^{b}\mathbf{m}^{b}$ |

‡ = Sources ($f_1$,$f_2$,$v_1^k$,$v_2^k$,$v_1^h$,$v_2^h$,$m^k$,$m^h$,$v_1^r$,$m^r$) were defined in the manuscript text and in the manuscript's Table 3.

**Table A6.** Constructing *level-2-class-specific* $R^2$ measures for a multilevel regression mixture with classes at both levels (L1L2MIX) using the variance decomposition derived in the Online Supplement and using the symbols defined in Appendix Table A4.

| Constituent parts to construct *level-2-class-specific* $R^2$s | Definition‡ | Eqn. # | Equation |
|---|---|---|---|
| Denominator of *level-2-class-specific* $R^2$ | *level-2 class-specific* outcome variance | Eqn (A25) | $\text{var}_{k\|h}(y_{ij\|c_{ij}=k,d_j=h}) = \text{var}_{k\|h}(\mathbf{x}_{ij}^{w\prime}\gamma^{w(kh)} + \mathbf{x}_j^{b\prime}\gamma^{b(kh)} + \mathbf{w}_{ij}'\mathbf{u}_j + \varepsilon_{ij}) = \gamma^{w(\bullet h)\prime}\Phi^w\gamma^{w(\bullet h)} + \gamma^{b(\bullet h)\prime}\Phi^b\gamma^{b(\bullet h)} + tr(\mathbf{T}^h\Phi^r) + \tau_{00}^h + tr\left(\Omega^{w(h)}\Phi^w\right) + tr\left(\Omega^{b(h)}\Phi^b\right) + \mathbf{m}^{b\prime}\Omega^{b(h)}\mathbf{m}^b + \theta^h$ |
| Numerator of *level-2-class-specific* $R^2$: Each term is included one at a time in the numerator of single-source measures | Variance attributable to source: $f_1$ | Eqn (A26) | $\gamma^{w(\bullet h)\prime}\Phi^w\gamma^{w(\bullet h)}$ |
| | Variance attributable to source $f_2$ | Eqn (A27) | $\gamma^{b(\bullet h)\prime}\Phi^b\gamma^{b(\bullet h)}$ |
| | Variance attributable to source: $v_1^r$ | Eqn (A28) | $tr(\mathbf{T}^h\Phi^r)$ |
| | Variance attributable to source $m^r$ | Eqn (A29) | $\tau_{00}^h$ |
| | Variance attributable to source $v_1^k$ | Eqn (A30) | $tr(\Omega^{w(h)}\Phi^w)$ |
| | Variance attributable to source $v_2^k$ | Eqn (A31) | $tr(\Omega^{b(h)}\Phi^b)$ |
| | Variance attributable to source $m^k$ | Eqn (A32) | $\mathbf{m}^{b\prime}\Omega^{b(h)}\mathbf{m}^b$ |

‡ = Sources ($f_1, f_2, v_1^k, v_2^k, m^k, v_1^r, m^r$) were defined in the manuscript text and in the manuscript's Table 3.

**Table A7.** Constructing *kh-class-combination-specific* $R^2$ measures for a multilevel regression mixture model with classes at both levels (L1L2MIX) using the variance decomposition derived in the Online Supplement and using the symbols defined in Appendix Table A4.

| Constituent parts to construct $R^2$s | Definition‡ | Eqn. # | Equation |
|---|---|---|---|
| Denominator of *class-combination-specific* $R^2$ | *class-combination (kh) specific* outcome variance | Eqn (A33) | $\text{var}_{ij\|kh}\left(y_{ij\|c_{ij}=k,d_j=h}\right) = \text{var}_{ij\|kh}(\mathbf{x}_{ij}^{w\prime}\gamma^{w(kh)} + \mathbf{x}_j^{b\prime}\gamma^{b(kh)} + \mathbf{w}_{ij}'\mathbf{u}_j + \varepsilon_{ij}) = \gamma^{w(kh)\prime}\Phi^w\gamma^{w(kh)} + \gamma^{b(kh)\prime}\Phi^b\gamma^{b(kh)} + tr(\mathbf{T}^h\Phi^r) + \tau_{00}^h + \theta^{kh}$ |
| Numerator of *class-combination-specific* $R^2$: Each term is included one at a time in the numerator of single-source measures | Variance attributable to source $f_1$ | Eqn (A34) | $\gamma^{w(kh)\prime}\Phi^w\gamma^{w(kh)}$ |
| | Variance attributable to source $f_2$ | Eqn (A35) | $\gamma^{b(kh)\prime}\Phi^b\gamma^{b(kh)}$ |
| | Variance attributable to source: $v_1^r$ | Eqn (A36) | $tr(\mathbf{T}^h\Phi^r)$ |
| | Variance attributable to source $m^r$ | Eqn (A37) | $\tau_{00}^h$ |

‡ = Sources ($m^r$,$v_1^r$,$f_1$,$f_2$) were defined in the manuscript text and in the manuscript's Table 3