AMERICAN PSYCHOLOGICAL ASSOCIATION

Psychological Methods

ISSN: 1082-989X

https://doi.org/10.1037/met0000477

A Structural Equation Modeling Approach for Modeling Variability as a Latent Variable

Yi Feng and Gregory R. Hancock
Department of Human Development and Quantitative Methodology, University of Maryland

Abstract

Drawing upon recent developments in structural equation modeling, the current study presents an analytical framework for addressing research questions in which, rather than focusing on means, it is intraindividual (or intragroup) variability that is of direct research interest. Beyond merely serving as an alternative to existing multilevel modeling approaches, this framework allows for extensions to accommodate a variety of complex research scenarios by parameterizing variability as a latent variable that can in turn be embedded within a broader covariance and mean structure involving other observed and/or latent variables. The estimation procedures and parameter interpretation for the latent random variability models are discussed. The versatility of the proposed methods is demonstrated through four empirical examples. The Mplus, BUGS, and Stan model syntax for the illustrative examples are supplied to facilitate the application of the methods.

Translational Abstract

In many research and applied settings across the social, behavioral, and health sciences, it is variability, rather than averages, that is of key interest. Examples include consistency/stability of an individual over multiple measurements (intraindividual variability), and cohesiveness among members within a group or team (intragroup variability). Drawing upon recent developments in structural equation modeling, the current study presents an analytical framework for addressing research questions that focus on intraindividual, or intragroup, variability. Beyond merely serving as an alternative to existing multilevel modeling approaches, this framework allows for extensions to accommodate a variety of complex research scenarios by parameterizing variability as a latent variable, which can be studied as the outcome, predictor, and/or mediators simultaneously in relation to other observed and/or latent variables. This study delineates the latent random variability models and offers a discussion of model estimation as well as parameter interpretation. To demonstrate the versatility of the proposed methods, the latent random variability models are fit to empirical data and parameter estimates are obtained via Bayesian estimation. The Mplus, BUGS, and Stan model syntax for the illustrative examples are supplied for applied researchers' reference.

Keywords: variance, heterogeneity, intraindividual differences, multilevel models, variability

Supplemental materials: https://doi.org/10.1037/met0000477.supp

In many research and applied settings across the social, behavioral, and health sciences, it is *variability*, rather than averages, that is of key interest. Examples include consistency/stability of an individual over multiple measurements (*intraindividual variability*), and cohesiveness among members within a group or team (*intragroup variability*). Focusing on the former, for instance, an infant often has daily fluctuations in crying behaviors and motor movement, an adult usually experiences day-to-day variations in bedtime and emotional status, and a senior tends to have varying reaction times across trials on cognitive tasks. Instead of treating such within-person fluctuations as random errors, modern developmental science theorizes that intraindividual variability contains valuable information about human development (Hamaker, 2012; van Dijk & van Geert, 2015). As intraindividual consistency, and how it changes over time, have important implications across the life span, intraindividual variability has thus become an important research topic across a variety of domains, such as motor skill development, socioemotional development, cognitive and language development, and cognitive aging (e.g., Adolph et al., 2015; Fagot et al., 2018; Hultsch & MacDonald, 2004; Kupers et al., 2019). Methodologically parallel, intragroup variability is also of wide interest across several disciplines. For example, organizational researchers are intrigued by how group members develop increasingly similar perceptions over time (e.g., Lang et al., 2018), while workforce diversity more broadly (e.g., task-related attitudinal diversity) and intrateam cohesion are believed to have impacts on group performance and productivity (e.g., Cox & Blake, 1991;

Yi Feng https://orcid.org/0000-0002-3487-908X

The research ideas and some preliminary findings have been presented at the 2019 annual meeting of Society of Multivariate Experimental Psychology in 2019, and at the 12th International Multilevel Conference. The data used for Illustrative Examples 2–4 are all publicly available.

Correspondence concerning this article should be addressed to Yi Feng, Department of Human Development and Quantitative Methodology, University of Maryland, 1228 Benjamin Building, 3942 Campus Drive, College Park, MD 20742-1115, United States. Email: yifeng94@umd.edu

Easley, 2001; Jehn et al., 1999; Petersen et al., 2004; Van Knippenberg et al., 2004). Even in biological science, understanding intraspecies variation in genotype and phenotype helps to explain diverse collective behaviors of group-living animals (e.g., Kaufhold & Van Leeuwen, 2019; Knebel et al., 2019).

Given that many essential research questions involve variability (equivalently, *heterogeneity*, *diversity*, *consistency*, *stability*, *cohesion*), researchers from a wide variety of disciplines have an increasing need for analytical tools that can directly model such intraindividual (and intragroup) variability, as well as help to examine factors that might impact it, and in turn, how it could affect other distal outcomes. Additionally, research may also require methods to examine differences in variability (intraindividual or intragroup) across contexts or time (e.g., pretest and posttest) and/or across different types of individuals/groups (e.g., males or females, novices or experts). Even more challenging still, researchers studying individuals or groups need to be able to study longitudinal trajectories of their variability, parallel processes across multiple outcomes, and mixtures of different underlying populations. These, and many other increasingly complex scenarios (see Table 1), constitute a collective call for a broad analytical framework that can model the intraindividual or intragroup variability as the focal random variable, such that model parameters can be estimated and meaningfully interpreted to make inferences about that variability in the population. To be clear, statistical modeling approaches do currently exist that can accommodate randomly varying variability in specific research scenarios, but there is not yet a well-articulated, versatile, and comprehensive framework dedicated to random variability modeling that can be flexibly utilized across disciplines to address the wide range of research questions involving random variability (e.g., Table 1). Building upon the increasingly flexible structural equation modeling (SEM) paradigm, the current study aims to introduce a unified analytical framework for when intraindividual (or intragroup) variability is of research interest, modeling variability as a latent random variable that can be embedded within a complex covariance and mean structure involving other observed and/or latent variables. The proposed system will have broad implications for addressing research needs in the social, behavioral, educational, organizational, health sciences, and beyond.

In what follows, we will first review the existing analytical approaches for modeling variability, in terms of theoretical background as well as important limitations. Next, we will present the conceptual framework for the proposed alternative modeling approaches based on modern SEM, followed by illustrative examples using empirical data. To conclude, we will discuss the broad implications of the proposed analytical methods and how this modeling approach can be further extended to accommodate a wide variety of complicated research scenarios in future research.

Theoretical Background

Multistep Approach

Acknowledging that individuals may have varying levels of intraindividual variability, one of the commonly used methods in developmental science is to first compute a summary statistic across an individual's observed repeated measures, and then submit these individual-level summary statistics to conventional statistical approaches for further analysis. Common examples of the summary statistics that are utilized to capture the intraindividual variability include intraindividual standard deviation (iSD), intraindividual mean square of successive differences (iMSSD), intraindividual residual standard deviation (iSDr), coefficient of variation (CV), and range (e.g., Adolph et al., 2015; Fagot et al.,

Table 1
A Selective List of Random Variability Models and the Corresponding Example Research Questions

	Example research questions	
Type of model	Intraindividual variability	Intragroup variability
Unconditional model	- Do people vary in terms of bedtime consistency? - Do people vary in terms of mental stability?	- Do teams differ in terms of cohesiveness? - Do schools differ in terms of the achievement gap among their students?
Variability as an outcome	- Do person-level characteristics predict bedtime consistency? - Do person-level characteristics predict mental status stability?	- Do team-level characteristics predict team cohesiveness? - Do school-level covariates predict within-school heterogeneity in students' math ability?
Variability as a predictor	- Does bedtime consistency predict physical health and mental health? - Does mental status stability predict physical health? - Does mental status stability predict mental well-being?	- Does team cohesiveness predict team productivity? - Does within-school achievement heterogeneity have an impact on student's future academic performance?
Between-subject design	- Do males and females differ in bedtime consistency? - Do young adults and senior adults differ in mental status stability?	- Do expert teams differ from novice teams in terms of cohesiveness?
Within-subject design	- Does an intervention program improve people's bedtime consistency? - Does a treatment improve individual's consistency on cognitive tasks?	- Does a training program improve teams' cohesiveness? - Does an intervention program help close the achievement gap within schools?
Growth structure of variability	- How does people's bedtime consistency change over time? - Does the maturation of white matter predict the changes in intraindividual behavioral consistency?	- How does teams' cohesiveness change over the course of training? - How does the achievement heterogeneity within schools change over the course of intervention?

2018; Hultsch & MacDonald, 2004; Ram et al., 2015). This multistep process is useful for person-oriented analysis, which does not necessarily involve a population of interest. Such an approach, however, is limited as the summary statistic conflates meaningful intraindividual fluctuations with measurement error, which is not properly accommodated in subsequent analytical steps. Further, treating the summary statistics as if they were true variabilities ignores the innate uncertainties in estimating the statistics from sample data, which can result in higher false positive rates within follow-up statistical tests (Dzubur et al., 2020).

Conventional Multilevel Modeling Approach

When there exists a population of individuals (or groups) of key inferential interest and the research questions involve repeated measures within those individuals (or groups), researchers often choose to model such nested data structures using multilevel modeling (MLM; also known as *hierarchical linear modeling* or *mixed effects models*). This strategy partitions the variance into components at different levels (e.g., within-person vs. between-person, within-group vs. between-group) and thus inferences can be made at each level accordingly (see, e.g., Hox et al., 2018; Raudenbush & Bryk, 2002; Snijders & Bosker, 2011). The general form of a two-level MLM can be written as:

$$Y_{ij} = \mathbf{X}'_{ij}\beta + \mathbf{Z}'_{ij}\mathbf{u}_j + e_{ij}, \tag{1}$$

where Y_{ij} is the observed outcome variable for Level-1 unit i nested within Level-2 cluster j; $\mathbf{X}_{ij}$ is the design matrix for fixed effects $\boldsymbol{\beta}$ and $\mathbf{Z}_{ij}$ is the design matrix for random effects $\mathbf{u}_j$; e_{ij} is the Level-1 residual, which is assumed to have a constant variance across all the Level-2 units: $e_{ij} \sim N(0, \sigma^2)$. Therefore, with conventional MLM, the levels of the observed outcome are allowed to randomly vary across clusters, whereas the residual within-cluster variance of the observed outcome is fixed to be constant in all the clusters. In other words, it assumes all the individuals have the same level of stability over repeated measures, or all the groups have the same level of cohesiveness among group members after conditioning on the relevant covariates. Thus, while MLM provides a useful approach for modeling within-level variability and between-level variability simultaneously within the same model, it is less helpful when the clusters are heterogeneous regarding the within-cluster variability and more importantly, when the randomly varying within-cluster variability itself is of research interest.

MLM Extensions

Recognizing the theoretical importance of stability for intraindividual processes, several extensions of conventional MLM have been put forth to model individually varying residual variances. Examples include the double hierarchical generalized linear models (Lee & Nelder, 2006), the dynamic multilevel first-order autoregressive model (Jongerling et al., 2015; Wang et al., 2012), and the location scale model (LSM; Hedeker & Mermelstein, 2007, 2012; Hedeker et al., 2008; Leckie et al., 2014). The LSM framework, which has well-documented applications with ecological momentary assessment (EMA) data, not only allows random effects in modeling the means of the observed outcome variable (i.e., the location), but also explicitly includes random effects in modeling the *variability* of the observed outcome variable (i.e., the scale). By parameterizing the residual variance with a log-linear model, the within-level residual variance is allowed to vary across Level-2 units. Further, the random residual variance is also allowed to covary with other Level-2 random coefficients, such as the random intercepts. Covariates can also be included in the model to explain the variability of the Level-1 residual variance.

The general form of the LSM can be expressed using Equation 1.[1] Again, the important difference between this and conventional MLM lies in the nature of the distribution for e_{ij}. Rather than assuming the residual variance to be constant, as in MLM, the residual variance is assumed to follow a log-normal distribution as per the following function:

$$\sigma^2_{e_{ij}} = \exp(\mathbf{W}'_{ij}\tau + u_{1j}), \tag{2}$$

where $\mathbf{W}_{ij}$ denotes the covariates that predict the within-cluster variability, and u_{1j} is the random effect of the within-level variance after controlling for the covariates. Within this modeling framework, u_{1j} is allowed to covary with other Level-2 random effects contained in $\mathbf{u}_j$. For example, it is possible to assess whether there is any linear relation between individuals' long-run average on the observed outcome (or starting level) and intraindividual fluctuation over repeated measures. Researchers can further have Level-2 covariance components differ across Level-2 units, either by directly modeling the covariance with a separate function (Leckie et al., 2014) or by indirectly allowing the corresponding standard deviations to vary (e.g., Rast & Ferrer, 2018; Williams et al., 2021). Most commonly, the LSM is used for intensive longitudinal data, such as EMA data (e.g., Hedeker et al., 2008), although it has also been shown to be applicable for heterogeneous Level-1 variance in a more generalized two-level scenario (e.g., Leckie et al., 2014). The LSM is thus a useful analytical tool when the focal outcome variable of interest is observed and effectively without measurement error (e.g., body weight).

The LSM approach is not without limitations, however. In the social sciences and educational studies, for example, more often than not researchers are interested in latent constructs that are difficult to measure directly, and thus are indicated by multiple error-laden observed variables. Consider psychologists who care about underlying depression but must collect symptom data through a survey of individuals' external behaviors and internal feelings, or education researchers who are interested in true math ability but must use responses to a set of test items. Simply put, observed variables are almost never perfect measures of a latent construct, but rather contain measurement error, and ignoring this fact by treating an observed variable as if it perfectly reflects the construct of interest can yield misleading results. Unfortunately, with MLM and its extensions, analyses are typically limited to observed variables or a linear composite thereof (e.g., an arithmetic mean or sum score), and there really is no straightforward way to take into consideration the measurement error. The sum score approach, however, is in fact a highly constrained measurement model, whose use is not warranted without theoretical support and empirical validation (McNeish & Wolf, 2020). With regard to LSM more

[1] Readers may refer to Kapur et al. (2015) for the multivariate extension of LSM with intensive longitudinal assessments.

specifically, the literature has focused primarily on a single observed outcome with only very few extensions to multivariate case for intensive longitudinal data (e.g., Kapur et al., 2015). Further, this modeling approach only allows the intraindividual variance or intragroup variance to be modeled as an outcome. In some research scenarios, however, it may be of interest to investigate whether such within-cluster variability is a predictor of other distal outcomes, such as, for example, a gerontologist interested in whether consistency in seniors' short-term memory (STM) predicts their degree of self-sufficiency with regard to life-care skills. Recent developments in LSM have introduced a multistage approach, where the random effects are first imputed based on the model parameter estimates obtained at the first stage and then carried to the next-stage analyses as predictors (Dzubur et al., 2020). It is, however, challenging to directly model variability as a random variable while serving as an outcome, predictor, and/or mediator simultaneously in a single model. Additionally, LSM is not yet able to accommodate other more complicated scenarios that are not typically considered under the MLM framework. For instance, from a very practical perspective, when researchers wish to directly assess the differences in Level-2 covariance structure across different types of Level-2 clusters (e.g., female vs. male, young adults vs. senior adults) it can be difficult to ensure positive definiteness if the covariance components are each modeled as a separate function of Level-2 covariates[2].

SEM-Based Approach

SEM is a multivariate analytic framework that has been widely used in the social, behavioral, and education sciences, and beyond. It is highly versatile, able to model hypothesized causal (structural) and noncausal links among not just observed variables, but can incorporate measurement models in order to allow the researchers to investigate relations among latent constructs that are free of measurement error.

Although historically SEM and MLM were developed as two distinct modeling frameworks, SEM and MLM are actually far more similar than different (e.g., Curran, 2003). The random effects in MLM can be interpreted as the unobserved latent confounders that cause the dependence among the lower-level observations belonging to the same higher-level unit. In SEM, common latent factors are usually introduced to account for the interdependence among different observed indicators (Rabe-Hesketh et al., 2012). Given the analytical similarity between SEM and MLM, research efforts have been made to expand the capability of SEM to accommodate multilevel data structure (e.g., Goldstein & McDonald, 1988; McDonald & Goldstein, 1989; Muthén, 1989, 1994; Rabe-Hesketh et al., 2004). Multilevel SEM (MSEM) has been developed as a synthesis of MLM and SEM, with performance comparable to traditional MLM but with the many versatility advantages of SEM (e.g., Bauer, 2003; Curran, 2003). Within MSEM, random effects can be conveniently modeled as latent variables, and as such they are able to be embedded within a more general latent variable framework and incorporated in a broader causal modeling structure. Further, it is easy to incorporate measurement models at both lower and higher levels. Following the within-between framework notations (e.g., Asparouhov & Muthén, 2007; Rabe-Hesketh et al., 2012), the general form of two-level MSEM can be summarized in the following equations:

$$Y_{pij} = Y^*_{\mathrm{B}pj} + Y^*_{\mathrm{W}pij} \tag{3}$$

$$\mathbf{Y}^*_{\mathrm{W}ij} = \boldsymbol{\Lambda}_{\mathrm{W}}\boldsymbol{\eta}_{\mathrm{W}ij} + \boldsymbol{\varepsilon}_{\mathrm{W}ij} \tag{4}$$

$$\boldsymbol{\eta}_{\mathrm{W}ij} = \mathbf{B}_{\mathrm{W}}\boldsymbol{\eta}_{\mathrm{W}ij} + \boldsymbol{\Gamma}_{\mathrm{W}}\mathbf{X}_{\mathrm{W}ij} + \boldsymbol{\xi}_{\mathrm{W}ij} \tag{5}$$

$$\mathbf{Y}^*_{\mathrm{B}j} = \mathbf{v}_{\mathrm{B}} + \boldsymbol{\Lambda}_{\mathrm{B}}\boldsymbol{\eta}_{\mathrm{B}j} + \boldsymbol{\varepsilon}_{\mathrm{B}j} \tag{6}$$

$$\boldsymbol{\eta}_{\mathrm{B}j} = \boldsymbol{\alpha}_{\mathrm{B}} + \mathbf{B}_{\mathrm{B}}\boldsymbol{\eta}_{\mathrm{B}j} + \boldsymbol{\Gamma}_{\mathrm{B}}\mathbf{X}_{\mathrm{B}j} + \boldsymbol{\xi}_{\mathrm{B}j}, \tag{7}$$

where Y_{pij} is the pth observed indicator for lower-level unit i belonging to higher-level unit j. The observed variable is modeled as the sum of latent variables at the between-cluster level ($Y^*_{p\mathrm{B}j}$) and the within-cluster level ($Y^*_{\mathrm{W}pij}$). The measurement model at the within level is expressed in Equation 4, where $\boldsymbol{\eta}_{\mathrm{W}ij}$ is a vector containing the latent constructs defined at the within level (e.g., an individual's depressive mood assessed each day, or the cultural beliefs of all members of a team member's cultural beliefs). The structural model at the within level is given by Equation 5, where $\mathbf{X}_{\mathrm{W}ij}$ denotes the observed covariates at the within level, and $\boldsymbol{\xi}_{\mathrm{W}ij}$ denotes the within-level random residuals for the latent constructs. Similarly, the measurement model and the structural model at the between level are given in Equations 6–7, respectively. In the MSEM literature, the residual vectors $\boldsymbol{\varepsilon}_{\mathrm{W}ij}$, $\boldsymbol{\xi}_{\mathrm{W}ij}$, $\boldsymbol{\varepsilon}_{\mathrm{B}j}$, and $\boldsymbol{\xi}_{\mathrm{B}j}$ are assumed to contain independent normally distributed random variables with means of zero and the corresponding variance-covariance matrices $\boldsymbol{\Theta}_{\mathrm{W}}$, $\boldsymbol{\Psi}_{\mathrm{W}}$, $\boldsymbol{\Theta}_{\mathrm{B}}$, and $\boldsymbol{\Psi}_{\mathrm{B}}$ (Asparouhov & Muthén, 2007). As suggested by the subscripts, the within-level variance-covariance structures are conventionally assumed to be invariant across clusters.

Recent work in SEM, however, has provided some promising directions for relaxing such restrictions of constant within-level variances. For example, as part of a multilevel confirmatory factor analysis illustration, Stapleton et al. (2016) utilized a so-called *phantom variable* (an unmeasured variable with no measured indicators) at the within level, which is assumed to have a standard normal distribution, thereby permitting the *scale* of within-cluster variability to be modeled as a random path coefficient at the between level. Another relevant recent development, designed for modeling the dynamics of stable processes with intensive longitudinal data, is dynamic structural equation modeling (DSEM; Asparouhov et al., 2018; McNeish & Hamaker, 2020), which combines time-series modeling with multilevel modeling along with SEM techniques. By incorporating the log-transformed residual variance as a person-level random parameter, DSEM allows individually-varying intraindividual variability over repeated measures

[2] Within the literature a separation strategy has been proposed to model the covariance matrix, which decomposes the covariance matrix into standard deviations and correlations that can be modelled separately (Barnard et al., 2000). It has been employed in the covariance modeling for the LSM, which usually incorporates covariates in the log-linear model for standard deviations (e.g., Rast & Ferrer, 2018; Williams et al., 2021). Although this approach ensures the positive definiteness of the covariance matrix, it assumes a constant correlation matrix, which may be less desirable in, for example, multigroup analysis.

that are collected within a relatively short observation window. The current implementation of DSEM is limited to two-level models using Bayesian estimation (Muthén & Muthén, 1998–2021), and its performance relies on large sample sizes with a relatively large number of repeated measurement occasions (e.g., Schultzberg & Muthén, 2018).

It is worth noting that Nestler's recent work on modeling intraindividual variability with intensive longitudinal data (Nestler, 2020, 2021) has successfully extended the LSM framework while taking advantage of factor analysis techniques. By imposing a latent factor structure on the individually-varying residual variances for multiple observed measures, Nestler's work further exemplifies the flexibility of SEM framework, which has a great potential to offer in conjunction with MLM-based LSM approach. One important contribution of this work is that a latent variability factor is introduced, which takes the reliability of variability indicators into consideration. It is, however, a different modeling approach than the one introduced in the current study. As detailed in the next section, the proposed modeling framework is essentially built upon on MSEM with broad applications to both intraindividual and intragroup variability, where the within-level latent constructs and between-level latent constructs are substantively meaningful and of theoretical interest (see Stapleton et al., 2016). The random variability is then introduced as a latent variable based on this theoretically meaningful within-level latent construct that is free of measurement error, rather than being indicated by its own variability indicators. Essentially, the different modeling approaches and parameterization strategies reflect different underlying causal structure that is believed to drive the data generation process. They also have different implications for the covariance structure among observed variables. For instance, with the models presented in Nestler (2020), the within-individual residuals of the observed measures are independent across different measures, after the between-individual random effects are taken into account (as no meaningful within-level latent constructs are assumed to drive the data generation). In contrast, the structural models proposed in the current study imply that the within-level residuals across different observed measures would covary even after the between-level random effects are controlled for, which can be explained by theoretically meaningful within-level latent construct (s). The MSEM-based models can be very useful in many cases, for instance, when multiple observed measures are collected at each measurement occasion to indicate a latent construct (e.g., daily positive affect) or when multiple observed measures are administered to group members to measure an individual-level latent construct (e.g., personal perceived support). With that said, we believe the choice between different modeling approaches should ultimately be considered within a specific research context, depending on substantive knowledge and theoretical beliefs.

In sum, then, although recent developments in SEM provide several promising and viable options to accommodate heterogeneous clusters with varying within-cluster variability, there is not yet a well-articulated and comprehensive framework within SEM that is specifically dedicated to modeling random variability (both intraindividual and intragroup). The current study thus aims to unify the varying available approaches and introduce a comprehensive random variability modeling framework that capitalizes on the versatility and power of SEM, one that can be applied across a wide range of basic or complex research scenarios with measured or latent variability as the focal construct. In the following section, we introduce the details of the theoretical framework for the proposed modeling approaches.

Conceptual Framework

In this section, we aim to introduce the conceptual framework of the SEM-based approach for random variability under two different scenarios: observed outcomes and latent outcomes. The structural equations and corresponding graphical representations for each scenario will be first introduced in detail, followed by a discussion of model estimation options and parameter interpretations for when heterogeneous within-level variability is of research interest.

Random Variability for Observed Outcomes

We first present the structural equation models that can be applied to study variability for observed outcome variables. For simplicity and without loss of generality, we begin with the unconditional two-level random variability structural model that involves only one continuous observed outcome variable, which resembles the setting of an unconditional LSM. Building upon this basic model structure, more complex models will be introduced next to accommodate latent outcomes in the section that follows.

Consider a scenario where there are multiple repeated measures nested within individuals (or individuals nested within groups), while the focal research question concerns intraindividual (or intragroup) variability and its relation with the cluster-level average. We discuss the following two models: (a) Model A that models the random variance of the Level-1 random effects using the log-transformation approach (Figure 1a); and (b) Model B that models the random variability using the phantom variable approach (Figure 1b). As illustrated in the figures, both Model A and Model B have a multilevel structure. In Model A, the Level-1 equation is

$$Y_{ij} = \gamma_{0j} + e_{ij}, \tag{8}$$

where γ_{0j} is the cluster mean of the observed outcome in cluster j. The lower level random effects are assumed to follow a normal distribution given variance σ_j^2: $e_{ij} \mid \sigma_j^2 \sim N(0, \sigma_j^2)$, therefore we have Y_{ij} follow a normal distribution, given mean γ_{0j} and variance σ_j^2: $Y_{ij} \mid \gamma_{0j}, \sigma_j^2 \sim N(\gamma_{0j}, \sigma_j^2)$. Notice the Level-1 residual variance σ_j^2 is group-specific, whose natural log-transformation is modeled as a random latent variable at the higher level. The Level-2 equations for Model A are thus written as

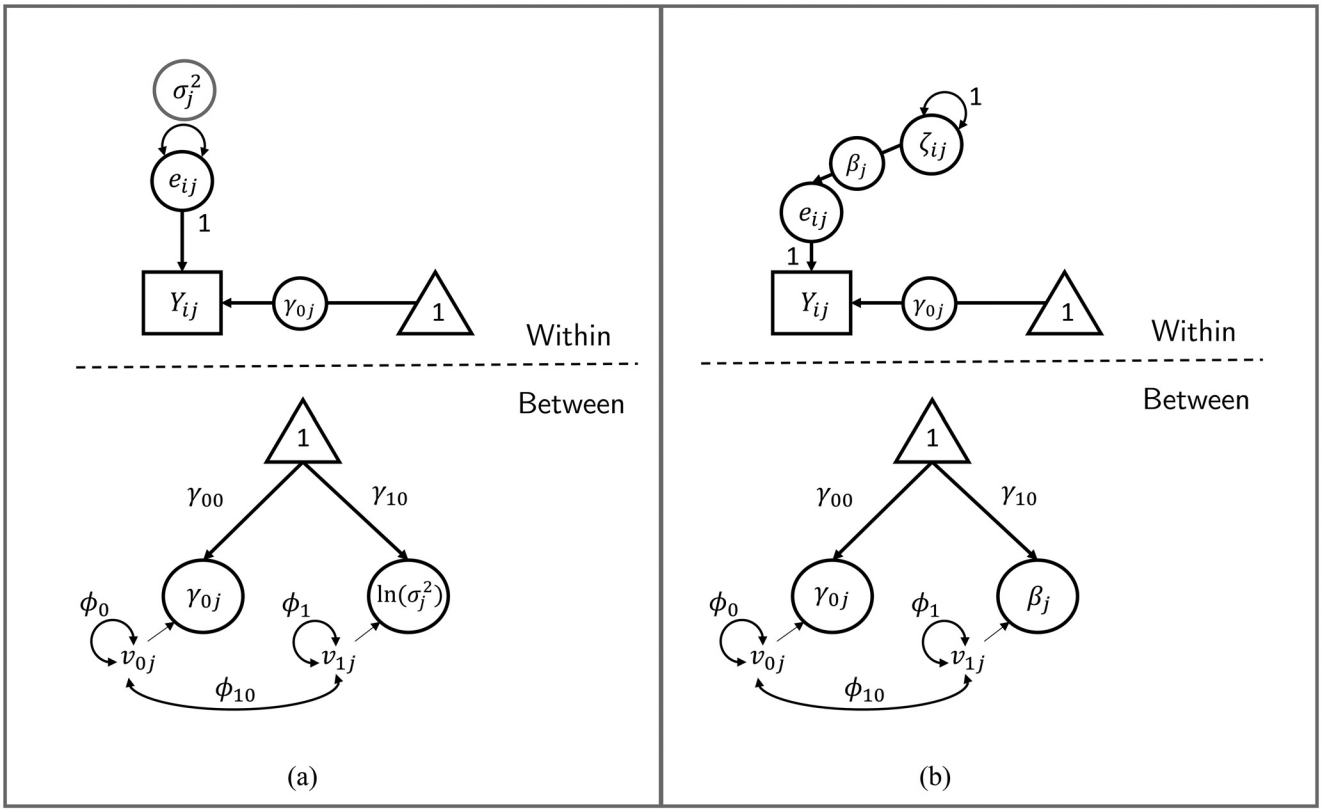
$$\boldsymbol{\eta}_j^{(B)} = \boldsymbol{\gamma}^{(B)} + \mathbf{v}_j^{(B)}, \tag{9}$$

where $\boldsymbol{\eta}_j^{(B)}$ is a vector of between-level latent variables, $\boldsymbol{\gamma}^{(B)}$ is the vector of fixed effects, and $\mathbf{v}_j^{(B)}$ contains the between level random effects:

$$\boldsymbol{\eta}_j^{(B)} = \begin{bmatrix} \gamma_{0j} \\ ln(\sigma_j^2) \end{bmatrix}, \boldsymbol{\gamma}^{(B)} = \begin{bmatrix} \gamma_{00} \\ \gamma_{10} \end{bmatrix}, \mathbf{v}_j^{(B)} = \begin{bmatrix} \nu_{0j} \\ \nu_{1j} \end{bmatrix}.$$

The average log residual variance across clusters is thus denoted by γ_{10}. The Level-2 random effects $\mathbf{v}_j^{(B)}$ are assumed to be multivariate normal:

Figure 1
Random Variability Structural Model With One Continuous Observed Outcome Variable Y



Note. (a) Model A: log-transformation approach for modeling the random within-level variance; (b) Model B: phantom variable approach for modeling the random within-level variance.

$$\mathbf{v}_j^{(\mathrm{B})} \sim \mathcal{N}(0, \mathbf{\Phi}_\mathrm{v}) \quad \mathbf{\Phi}_\mathrm{v} = \begin{bmatrix} \phi_0 & \\ \phi_{10} & \phi_1 \end{bmatrix}, \tag{10}$$

where ϕ_1 represents the between-level variance of the log residual variance, and ϕ_{10} denotes the covariance between the random intercepts and random log variances. Therefore, the model implies that the Level-2 latent variables (random effects) follow a multivariate normal distribution, given the mean vector $\boldsymbol{\gamma}^{(\mathrm{B})}$ and covariance matrix $\mathbf{\Phi}_\mathrm{v}$:

$$\boldsymbol{\eta}_j^{(\mathrm{B})} \mid \boldsymbol{\gamma}^{(\mathrm{B})}, \mathbf{\Phi}_\mathrm{v} \sim \mathcal{N}(\boldsymbol{\gamma}^{(\mathrm{B})}, \Phi_\mathrm{v}). \tag{11}$$

As an alternative, Model B (Figure 1b) can also be used to model the random variability by introducing a phantom variable that follows a standard normal distribution (Stapleton et al., 2016). The Level-1 equation for Model B can be written as

$$Y_{ij} = \gamma_{0j} + \beta_j \zeta_{ij}, \tag{12}$$

where ζ_{ij} is the phantom variable: $\zeta_{ij} \sim N(0, 1)$, and β_j is a group-specific *scaling factor* of the Level-1 residual. The outcome variable Y_{ij} thus follows a normal distribution given mean γ_{0j} and variance β_j^2: $Y_{ij} \mid \gamma_{0j}, \beta_j^2 \sim N(\gamma_{0j}, \beta_j^2)$. At Level-2, β_j can be conveniently modeled as a latent random variable. The Level-2 equation for Model B can also be expressed using Equation 9, where the between-level latent variables are $\boldsymbol{\eta}_j^{(\mathrm{B})} = (\gamma_{0j}, \beta_j)^\mathrm{T}$. Similar to the log variance approach (Model A), the Level-2 latent random variables follows a multivariate normal distribution as shown in Equation 11. The only difference is that in Model B, γ_{10} indicates the average scaling factor β_j across groups, ϕ_1 represents the between-cluster variance of the scaling factor, and ϕ_{10} denotes the covariance between random intercepts and random scaling factors.

Random Variability for Latent Outcomes

As discussed earlier, for many disciplines it is crucial to take measurement error into consideration, which also applies to the research scenario when random variability is of focal interest. One of the driving motivations for introducing the SEM-based framework for random variability is to take advantage of its powerful capability for latent variable modeling. A major goal of the current study is thus to illustrate how the observed outcome models for random variability can be extended to accommodate latent outcome variables.

For simplicity, we will again start with a basic two-level scenario, where there could be repeated measures nested within individuals (or multiple individuals nested within groups). For each measurement occasion within an individual (subject within a group), multiple continuous manifest indicators are collected to inform a single latent construct (e.g., depression, cultural identity).

Suppose the focal research question concerns to what extent individuals' depressive mood fluctuates over repeated measures and how individuals differ in their mood stability (i.e., the average level and variability of *intraindividual consistency*), or to what extent people on the same team share similar cultural beliefs and how different teams differ in the extent of shared cultural beliefs among the team members (i.e., the average level and variability of *team cohesiveness*). We therefore need to model the variance of the Level-1 latent construct as a random coefficient that varies across the Level-2 clusters. To this end, we can either use Model C with a log-transformation approach (Figure 2a) or Model D with a phantom variable approach (Figure 2b).

Similar to Model A and Model B for observed outcome variables, both Model C and Model D have a multilevel structure:

$$Y_{kij} = \eta_{k\mathrm{B},j} + \eta_{k\mathrm{W},ij}, \tag{13}$$

where $\eta_{k\mathrm{B},j}$ indicates the cluster average of indicator k for *group j*, and $\eta_{k\mathrm{W},ij}$ indicates the individual deviation from the cluster average of indicator k for person *i*. Assume that the observed indicators are intended to measure a single latent construct, we can have the following measurement models at Level-2 and Level-1, respectively:

$$\begin{aligned} \eta_{k\mathrm{B},j} &= \gamma_{k00} + \lambda_{k\mathrm{B}}\xi_{\mathrm{B},j} + \varepsilon_{k\mathrm{B},j} \\ \eta_{k\mathrm{W},ij} &= \lambda_{k\mathrm{W}}\xi_{\mathrm{W},ij} + \varepsilon_{k\mathrm{W},ij}, \end{aligned} \tag{14}$$

where $\xi_{\mathrm{B},j}$ is the Level-2 latent construct and $\xi_{\mathrm{W},ij}$ is the Level-1 latent construct. Alternatively, we could also have a saturated structure at the between level if the between-level latent construct is not of interest or not interpretable (see Stapleton et al., 2016 for more detailed discussion). In many cases, the factor loadings $\lambda_{k\mathrm{B}}$ and $\lambda_{k\mathrm{W}}$ can be constrained to be equal (i.e., *cross-level measurement invariance*; Stapleton et al., 2016). If we substitute Equation 14 into Equation 13, the multilevel measurement model becomes

$$Y_{kij} = \gamma_{k00} + \lambda_{k\mathrm{B}}\xi_{\mathrm{B},j} + \varepsilon_{k\mathrm{B},j} + \lambda_{k\mathrm{W}}\xi_{\mathrm{W},ij} + \varepsilon_{k\mathrm{W},ij}, \tag{15}$$

or equivalently in matrix form:

$$\mathbf{Y} = \mathbf{\Gamma} + \boldsymbol{\lambda}_{\mathrm{B}}\boldsymbol{\xi}_{\mathrm{B}} + \boldsymbol{\varepsilon}_{\mathrm{B}} + \boldsymbol{\lambda}_{\mathrm{W}}\boldsymbol{\xi}_{\mathrm{W}} + \boldsymbol{\varepsilon}_{\mathrm{W}}. \tag{16}$$

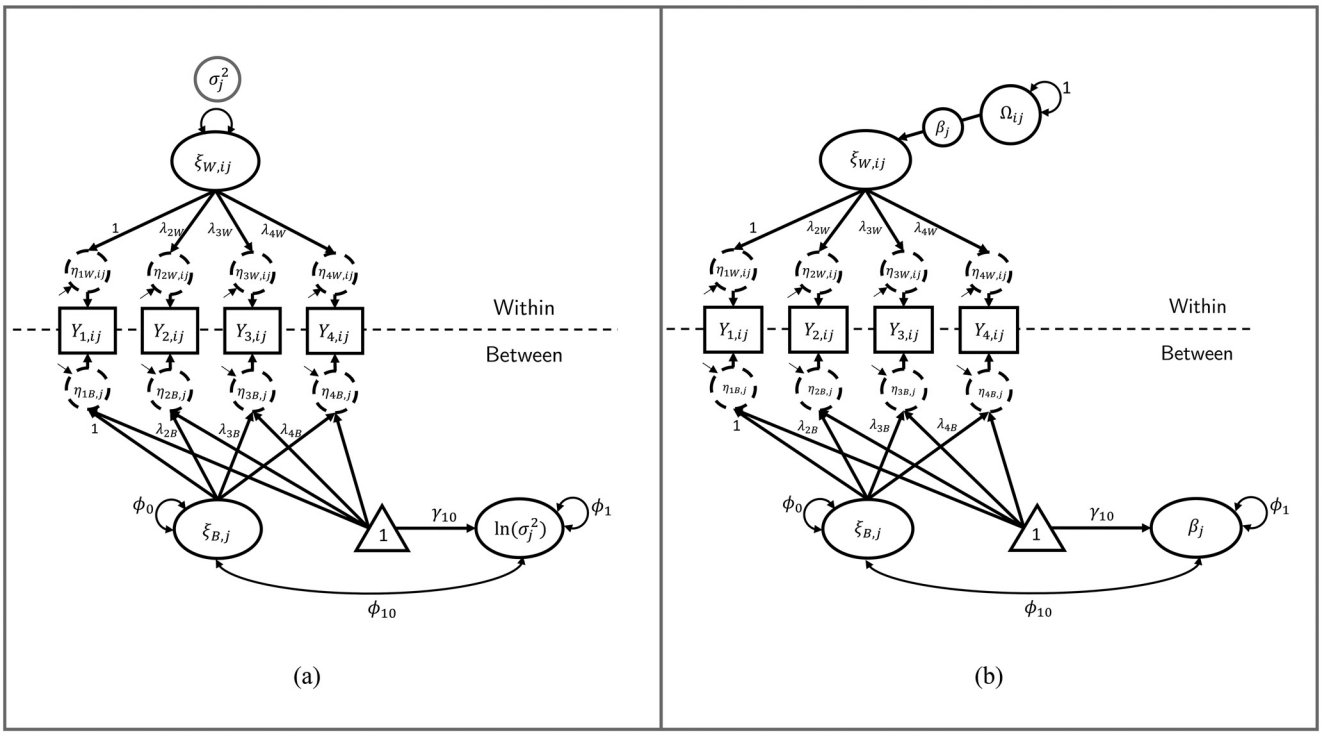
Similar to conventional MSEM, the within-level residuals $\boldsymbol{\varepsilon}_{\mathrm{W}}$ and between-level residuals $\boldsymbol{\varepsilon}_{\mathrm{B}}$ are assumed to be independently and normally distributed: $\boldsymbol{\varepsilon}_{\mathrm{W}} \sim N(\mathbf{0}, \boldsymbol{\Theta}_{\mathrm{W}})$, $\boldsymbol{\varepsilon}_{\mathrm{B}} \sim N(\mathbf{0}, \boldsymbol{\Theta}_{\mathrm{B}})$. In Model C (Figure 2a), the within-level latent construct $\xi_{\mathrm{W},ij}$ is assumed to be normally distributed given variance σ_j^2, $\xi_{\mathrm{W},ij} \mid \sigma_j^2 \sim N(0, \sigma_j^2)$. The variance of the Level-1 latent construct for *group j* is denoted as σ_j^2, indicating it is cluster-specific. Next, the natural logarithm (ln) of σ_j^2 is modeled as a random latent variable at between-level. The Level-2 structural equations in Model C can be written as

$$\boldsymbol{\xi}_j^{(\mathrm{B})} = \boldsymbol{\gamma}^{(\mathrm{B})} + \mathbf{v}_j^{(\mathrm{B})}, \tag{17}$$

where $\boldsymbol{\xi}_j^{(\mathrm{B})} = (\xi_{\mathrm{B},j}, \ln(\sigma_j^2))^{\mathrm{T}}$, $\boldsymbol{\gamma}^{(\mathrm{B})} = (0, \gamma_{10})^{\mathrm{T}}$, and $\mathbf{v}_j^{(\mathrm{B})} = (\nu_{0j}, \nu_{1j})^{\mathrm{T}}$, with γ_{10} denoting the average log-variance of the Level-1 latent construct, and ϕ_1 denoting the between-cluster variance of the log-variance. Similar to Model A and Model B, the Level-2 random

Figure 2

Random Variability SEM With One Within-Level Latent Outcome Variable ξ_{W} and One Between-Level Latent Outcome Variable ξ_{B}



Note. For each within-level unit, four observed measures (Y_1–Y_4) are collected as indicators of the latent construct. (a) log-variance approach for modeling the random within-level variance; (b) phantom variable approach for modeling the random within-level variance.

effects $\mathbf{v}_j^{(B)}$ are assumed to follow a multivariate normal distribution. Therefore, the Level-2 latent variables also follow a multivariate normal distribution

$$\boldsymbol{\xi}_j^{(B)} \mid \boldsymbol{\gamma}^{(B)}, \boldsymbol{\Phi}_v \sim \mathcal{N}(\boldsymbol{\gamma}^{(B)}, \boldsymbol{\Phi}_v). \tag{18}$$

As an alternative, Model D (Figure 2b) can be used to model the heterogeneous within-cluster variances, by introducing a phantom variable Ω that is assumed to follow a standard normal distribution. Using this approach, the Level-1 latent variable is modeled as

$$\xi_{W,ij} = \beta_j \Omega_{ij}, \tag{19}$$

and the multilevel measurement model can thus be written as

$$Y_{kij} = \gamma_{k00} + \lambda_{kB}\xi_{B,j} + \varepsilon_{kB,j} + \lambda_{kW}(\beta_j \Omega_{ij}) + \varepsilon_{kW,ij}, \tag{20}$$

where Ω_{ij} is the phantom variable that follows a standard normal distribution, $\Omega_{ij} \sim N(0, 1)$; β_j is the scaling factor for the within-cluster variance, such that $\xi_{W,ij} \mid \beta_j^2 \sim N(0, \beta_j^2)$. At Level-2, β_j can be conveniently modeled as a latent random variable. The Level-2 structural equation for Model D can also be expressed using Equation 17, except that the Level-2 latent variables contain the scaling factor: $\boldsymbol{\xi}_j^{(B)} = (\xi_{B,j}, \beta_j)^T$. More specifically, with this alternative parameterization, γ_{10} indicates the average scaling factor β_j for the latent outcome variable across clusters and ϕ_1 represents the between-cluster variance of the scaling factor for the latent outcome.

Model Estimation

In practice, the true values of model parameters remain unknown and thus need to be estimated; as such, the estimation of model parameters is an important topic to be addressed for variability modeling as well. In this section, we first review the existing estimation procedures for both the MLM-based and the SEM-based approaches that accommodate varying within-cluster variability. We next provide a brief discussion about the application of Bayesian estimation, specifically in the context of the proposed MSEM-based models with variability as a latent random variable.

In general, two estimation approaches have been employed for models with a heterogeneous variability component: maximum likelihood (ML) estimation and Bayesian estimation. The univariate LSM with one observed outcome variable has been traditionally estimated by marginal ML with Newton-Raphson algorithm using SAS PROC NLMIXED procedure (Hedeker et al., 2008) or the MIXREGLS program (Hedeker & Nordgren, 2013), when the random variability is only treated as an outcome. ML estimation is also used for the multivariate extension of LSM with latent variability, as detailed by Nestler (2020). A potential issue with the ML estimation approach for LSM is that it may be challenging to implement with smaller sample size (e.g., small number of clusters, few individuals, or measurement occasions), which can be circumvented by the Bayesian framework. Therefore, in practice, researchers have also employed the alternative Bayesian approach via Markov chain Monte Carlo (MCMC) simulation for estimating univariate (Rast et al., 2012) as well as multivariate longitudinal LSM (Kapur et al., 2015) using JAGS/WinBUGS. Recent developments in LSM for univariate observed outcomes further propose a two-stage approach that utilizes the empirical Bayes methods to obtain multiple draws of the random coefficients based on the Stage-1 ML parameter estimates, which can in turn be used in Stage-2 statistical analyses as predictors in the MixWILD program (Dzubur et al., 2020).

On the other hand, in the MSEM literature, although ML and Bayesian estimation have both been employed to estimate MSEM models (Asparouhov & Muthén, 2016; Rosseel, 2017), Bayesian estimation is much more commonly utilized and easier to implement given the novelty and complexity of such models, especially when measurement models are present. The latest version of the popular SEM software Mplus (Version 8.6; Muthén & Muthén, 1998–2021) also offers built-in DSEM/MSEM modules that can model random residual variance on its log transformed scale via Bayesian estimation, which greatly facilitates the implementation of random variability models in practice. With the current functionality of Mplus, it is also straightforward and computationally efficient to fit the LSM as a MSEM or DSEM via Bayesian estimation (e.g., McNeish, 2020; Nestler, 2021). Given the anticipated increasing model complexity as well as the corresponding computation burden for latent variable models in the context of modeling random variability, in the current study we focus on Bayesian estimation only. Below we provide a brief overview of how Bayesian estimation can be applied for MSEM-based random variability models. Readers who are interested can refer to Supplemental Material A for more technical details, while those who are less familiar with the mathematical terms may choose to skip this section without impacting their application of these methods in practice.

With Bayesian estimation, inference can be made about the parameters of focal interest using the posterior distribution given the observed data. For the models discussed above, we work with the augmented posterior density given the observed data $\mathbf{Y}$:

$$P(\boldsymbol{\Theta} \mid \mathbf{Y}) \propto P(\mathbf{Y}, \boldsymbol{\Theta}), \tag{21}$$

where $\boldsymbol{\Theta}$ is a vector that contains the model parameters including all the fixed effects, variance-covariance of the random effects, as well as the between-level latent random variables (hence unobserved; for more about the data augmentation approach, see Tanner & Wong, 1987, 2010). In the case of Model A and Model B, we have $\boldsymbol{\Theta} = (\boldsymbol{\theta}^T, \text{vec}(\boldsymbol{\eta}^{(B)})^T)^T$, where the model parameters vector $\boldsymbol{\theta}$ contains both fixed effects and variance-covariance of the random effects: $\boldsymbol{\theta} = (\boldsymbol{\gamma}^{(B)T}, \text{vech}(\boldsymbol{\Phi}_v)^T)^T$. The augmented posterior distribution under the normality assumption, as detailed in the Supplemental Material A, is expressed as

$$\begin{aligned} P(\boldsymbol{\theta}, \boldsymbol{\eta}^{(B)} \mid \mathbf{Y}) &\propto P(\mathbf{Y} \mid \boldsymbol{\eta}^{(B)}, \boldsymbol{\theta}) P(\boldsymbol{\eta}^{(B)} \mid \boldsymbol{\theta}) P(\boldsymbol{\theta}) \\ &= \left[\prod_{j=1}^{J} \prod_{i=1}^{n_j} \frac{1}{\sqrt{2\pi\sigma_j^2}} \exp\left(-\frac{(y_{ij} - \gamma_{0j})^2}{2\sigma_j^2} \right) \right] \\ &\times \left[\prod_{j=1}^{J} \frac{1}{2\pi^{m/2} \mid \boldsymbol{\Phi}_v \mid^{1/2}} \exp\left\{ -\frac{1}{2} (\boldsymbol{\eta}_j^{(B)} - \boldsymbol{\gamma}^{(B)})^T \boldsymbol{\Phi}_v^{-1} \left(\boldsymbol{\eta}_j^{(B)} - \boldsymbol{\gamma}^{(B)} \right) \right\} \right] \\ &\times P(\boldsymbol{\gamma}^{(B)}) P(\boldsymbol{\Phi}_v), \end{aligned} \tag{22}$$

where m is the number of Level-2 random effects.

In the case of Model C and Model D, we have $\boldsymbol{\Theta} = (\boldsymbol{\theta}^T, \text{vec}(\boldsymbol{\xi}^{(B)})^T, \boldsymbol{\varepsilon}_B{}^T)^T$, where the model parameter vector $\boldsymbol{\theta} = (\boldsymbol{\Gamma}^T, \boldsymbol{\gamma}^{(B)T}, \boldsymbol{\lambda}_B{}^T, \boldsymbol{\lambda}_W{}^T, \text{vech}(\boldsymbol{\Phi}_v)^T, \text{vech}(\boldsymbol{\Theta}_B)^T, \text{vech}(\boldsymbol{\Theta}_W)^T)^T$. The augmented

posterior distribution, as detailed in Supplemental Material A, can be expressed as:

$$
\begin{aligned}
P(\boldsymbol{\theta}, \boldsymbol{\xi}^{(\mathrm{B})}, \boldsymbol{\varepsilon}_{\mathrm{B}} \mid \mathbf{Y}) &\propto P(\mathbf{Y} \mid \boldsymbol{\xi}^{(\mathrm{B})}, \boldsymbol{\varepsilon}_{\mathrm{B}}, \boldsymbol{\theta}) P(\boldsymbol{\xi}^{(\mathrm{B})} \mid \boldsymbol{\theta}) P(\boldsymbol{\varepsilon}_{\mathrm{B}} \mid \boldsymbol{\theta}) P(\boldsymbol{\theta}) \\
&= \left[\prod_{j=1}^{J} \prod_{i=1}^{n_j} \frac{1}{2\pi^{K/2} \mid \boldsymbol{\lambda}_{\mathrm{w}} \sigma_j^2 \boldsymbol{\lambda}_{\mathrm{w}}{}^T + \Theta_{\mathrm{w}} \mid^{1/2}} \right. \\
&\exp\left\{ -\frac{1}{2} (\mathbf{Y}_{ij} - (\boldsymbol{\gamma}_{00} + \boldsymbol{\lambda}_{\mathrm{B}} \xi_{\mathrm{B},j} + \boldsymbol{\varepsilon}_{\mathrm{B},j}))^T (\boldsymbol{\lambda}_{\mathrm{w}} \sigma_j^2 \boldsymbol{\lambda}_{\mathrm{w}}{}^T + \Theta_{\mathrm{w}})^{-1} \right. \\
&\left. \left. (\mathbf{Y}_{ij} - (\boldsymbol{\gamma}_{00} + \boldsymbol{\lambda}_{\mathrm{B}} \xi_{\mathrm{B},j} + \boldsymbol{\varepsilon}_{\mathrm{B},j})) \right\} \right] \\
&\times \left[\prod_{j=1}^{J} \frac{1}{2\pi^{m/2} \mid \Phi_{\mathrm{v}} \mid^{1/2}} \exp\left\{ -\frac{1}{2} (\xi_j^{(\mathrm{B})} - \boldsymbol{\gamma}^{(\mathrm{B})})^T \boldsymbol{\Phi}_{\mathrm{v}}^{-1} (\xi_j^{(\mathrm{B})} - \boldsymbol{\gamma}^{(\mathrm{B})}) \right\} \right] \\
&\times \left[\prod_{j=1}^{J} \frac{1}{2\pi^{K/2} \mid \boldsymbol{\Theta}_{\mathrm{B}} \mid^{1/2}} \exp\left\{ -\frac{1}{2} (\boldsymbol{\varepsilon}_{\mathrm{B},j} - \boldsymbol{\Gamma})^T \boldsymbol{\Theta}_{\mathrm{B}}^{-1} (\boldsymbol{\varepsilon}_{\mathrm{B},j} - \boldsymbol{\Gamma}) \right\} \right] \\
&\times P(\boldsymbol{\Gamma}) P(\boldsymbol{\gamma}^{(\mathrm{B})}) P(\boldsymbol{\lambda}_{\mathrm{B}}, \boldsymbol{\lambda}_{\mathrm{W}}) P(\boldsymbol{\Phi}_{\mathrm{v}}) P(\boldsymbol{\Theta}_{\mathrm{B}}) P(\boldsymbol{\Theta}_{\mathrm{W}}), \qquad (23)
\end{aligned}
$$

with K denoting the number of observed indicators for the measurement model.

With the augmented posterior distribution, the parameter of focal interest can be estimated by examining its marginal posterior distributions via MCMC. For instance, the posterior mean can be used as parameter point estimate and a posterior credible interval can be obtained as the interval estimate. Bayesian estimation via MCMC is flexible and straightforward to implement for the basic models as well as for more complex models. It can be implemented using software package such as `rjags` (Plummer, 2016) and Stan (Carpenter et al., 2017). The popular SEM software Mplus also offers convenient Bayesian estimation functionality (Muthén & Muthén, 1998–2021).

Interpretation and Inference

The unconditional models (Models A–D) can each be applied to research scenarios where the inferential interest is in how clusters vary in terms of the intracluster variability. Model A and Model B are useful when the research questions concern the within-cluster variability regarding one single observed outcome measure (or multiple observed measures that do not indicate a common latent construct), while Model C and Model D can both be applied to research scenarios that concern the within-cluster variability in an unobserved latent construct that is indicated by multiple observed measures (for example, depression). For research questions that mainly aim to address heterogeneous within-cluster variability, the model parameters that are of most inferential interest are the fixed effects $\boldsymbol{\gamma}^{(\mathrm{B})}$ and the variance-covariance of the between-level random effects $\boldsymbol{\Phi}_{\mathrm{v}}$. With the log-transformation approach, researchers can thus describe and test the average log within-cluster variance (γ_{10}), how much log within-cluster variance varies across clusters (ϕ_1), as well as how log within-cluster variance correlates with the cluster average level on the outcome (ϕ_{10}). With the phantom variable parameterization, the interpretations can be similarly done, except that it would involve the within-cluster scaling factor rather than the log variance.

As readers may have already noticed, with either parameterization (log-transformation or phantom variable approach), the *within-level residual variance* is never directly modeled as a random variable, and thus its distributional characteristics remain unknown. Given that sometimes researchers may be more interested in learning about the variance per se, which can make more intuitive sense than the log variance or scaling factor, next we suggest some guidelines that can be useful for making inferences about the within-cluster variability on a variance scale. More specifically, the following moment transformation approach is proposed.

With the log variance approach (Model A and Model C), the log of the residual variance is assumed to follow a normal distribution (Equations 11 and 18); the within-level residual variance thus follows a lognormal distribution, $\sigma_j^2 \sim \text{lognormal}(\gamma_{10}, \phi_1)$. Based on our knowledge of lognormal distributions, the expectation and variance of the within-level variance σ_j^2 can be expressed as functions of the mean (γ_{10}) and variance (ϕ_1) of the log variance, respectively:

$$\mathrm{E}(\sigma_j^2) = \exp(\gamma_{10} + \frac{1}{2}\phi_1), \qquad (24)$$

$$\mathrm{VAR}(\sigma_j^2) = \exp(2\gamma_{10} + \phi_1)[\exp(\phi_1) - 1]. \qquad (25)$$

For the phantom variable approach (Model B and Model D), it is assumed that β_j follows a normal distribution (Equations 11 and 18). The within-level residual variance β_j^2 thus follows a scaled noncentral χ^2 distribution. Given the common definition of variance for any random variable X

$$\mathrm{VAR}(X) = \mathrm{E}(X^2) - [\mathrm{E}(X)]^2, \qquad (26)$$

the expectation and variance for this scaled noncentral χ^2 distribution can thus be computed, respectively, using the following equations after rearranging Equation 26:

$$
\begin{aligned}
\mathrm{E}(\beta_j^2) &= \mathrm{E}(\beta_j)^2 + \mathrm{VAR}(\beta_j) \\
&= \gamma_{10}^2 + \phi_1, \qquad (27)
\end{aligned}
$$

$$\mathrm{VAR}(\beta_j^2) = \mathrm{E}(\beta_j^4) - \mathrm{E}(\beta_j^2)^2, \qquad (28)$$

where the expected value of β_j^4 can be obtained with the moment-generating function for normal distribution (see more details in the Appendix):

$$\mathrm{E}(\beta_j^4) = M^{(4)}(0) = \gamma_{10}^4 + 6\phi_1\gamma_{10}^2 + 3\phi_1^2. \qquad (29)$$

With the moment transformation approach outlined above, inferences can be directly made about the average level of within-cluster variance ($\mathrm{E}(\sigma_j^2)$, $\mathrm{E}(\beta_j^2)$) as well as its heterogeneity across clusters ($\mathrm{VAR}(\sigma_j^2)$, $\mathrm{VAR}(\beta_j^2)$). Importantly, as we propose to use Bayesian estimation via MCMC, it is convenient to implement the moment transformation within each draw and thus monitor the posterior distributions of the moments for within-cluster variance.[3]

[3] This approach is preferred over performing the transformation on the point estimate of the moments for the log-variance or scaling factor (either obtained via ML or Bayesian MCMC). Based on Jensen's inequality, for a convex function g we have $E[g(X)] \geq g(E[x])$. Therefore, transforming the point estimate of the moments for the original distribution will likely yield biased estimates of the moments for the transformed distribution.

Model Extensions

So far we have presented the most basic unconditional models that can be used to model the random intraindividual or intragroup variability for single group analysis. Building upon the above modeling strategy, the models for random variability can be further adapted to accommodate more complicated scenarios. For instance, regarding the intraindividual variability over repeated measures, it may be necessary to control for the time trend of the outcome variables (that is, detrending) when the outcome is believed to systematically change over time. Extending Model A and Model B, it is straightforward to control for the time effect by either including the time variable at Level-1 (for example, Figure 3), or by incorporating a latent growth component into the model (for example, Figure 4). Moving forward, we can envision a collection of SEM-based random variability models to meet a variety of research needs that are commonly seen across the disciplines in social and behavioral sciences (see Table 1). Examples include conditional models with random variability as the predictor, outcome, and/or mediator (for example, Figure 5), between-subjects designs (for example, multiple group comparison regarding random variability, Figure 6), within-subject designs (for example, pre- and posttest random variability), longitudinal growth trajectories of random variability, parallel processes with multiple observed/latent outcomes, and mixture modeling to detect heterogeneous subpopulations differing in terms of random variability. Although more nuanced discussion of each of the extensions is reserved for future methodological and applied research, the modeling strategies and techniques presented in this study present a promising foundation that can potentially grow into a comprehensive framework for random variability modeling that has wide applications.

Illustrative Examples

In this section, we demonstrate the application of the random variability modeling framework with empirical data. Across the examples, we illustrate how the models can be employed to help address research questions about random variability across four different scenarios: (a) unconditional and conditional measured variable models for intraindividual variability; (b) unconditional latent variable models for intraschool variability; (c) conditional latent variable models for intraindividual variability; and (d) conditional latent variable models for intraindividual variability with time trend and lagged residual correlation over time being taken into account. For each example, the models are fit to the data via Mplus or rjags and rstan. The corresponding Mplus code, BUGS model syntax for rjags and Stan model syntax for rstan are provided in Supplemental Material B. The interpretation of the model results is discussed within the corresponding research context.

Example 1: Heterogeneous Intraindividual Variability of Daily Curiosity

For this illustrative example, we applied the proposed models to examine intraindividual consistency in daily curiosity. The empirical data used in this example were collected from the Knowledge Networks Over Time (KNOT) study, an intensive longitudinal investigation of day-to-day intraindividual variability in various outcomes, with curiosity being of specific interest to the investigators (Lydon-Staley et al., 2020). The researchers theorized that people not only differ in their trait-level curiosity, but also vary in the fluctuations of their daily experience of curiosity. Further, they hypothesized that an individual's mental well-being depends not only on a person's trait-level curiosity, but on how consistent a person's curiosity levels are from day to day as well. More details about the KNOT study can be found in Lydon-Staley et al. (2020).

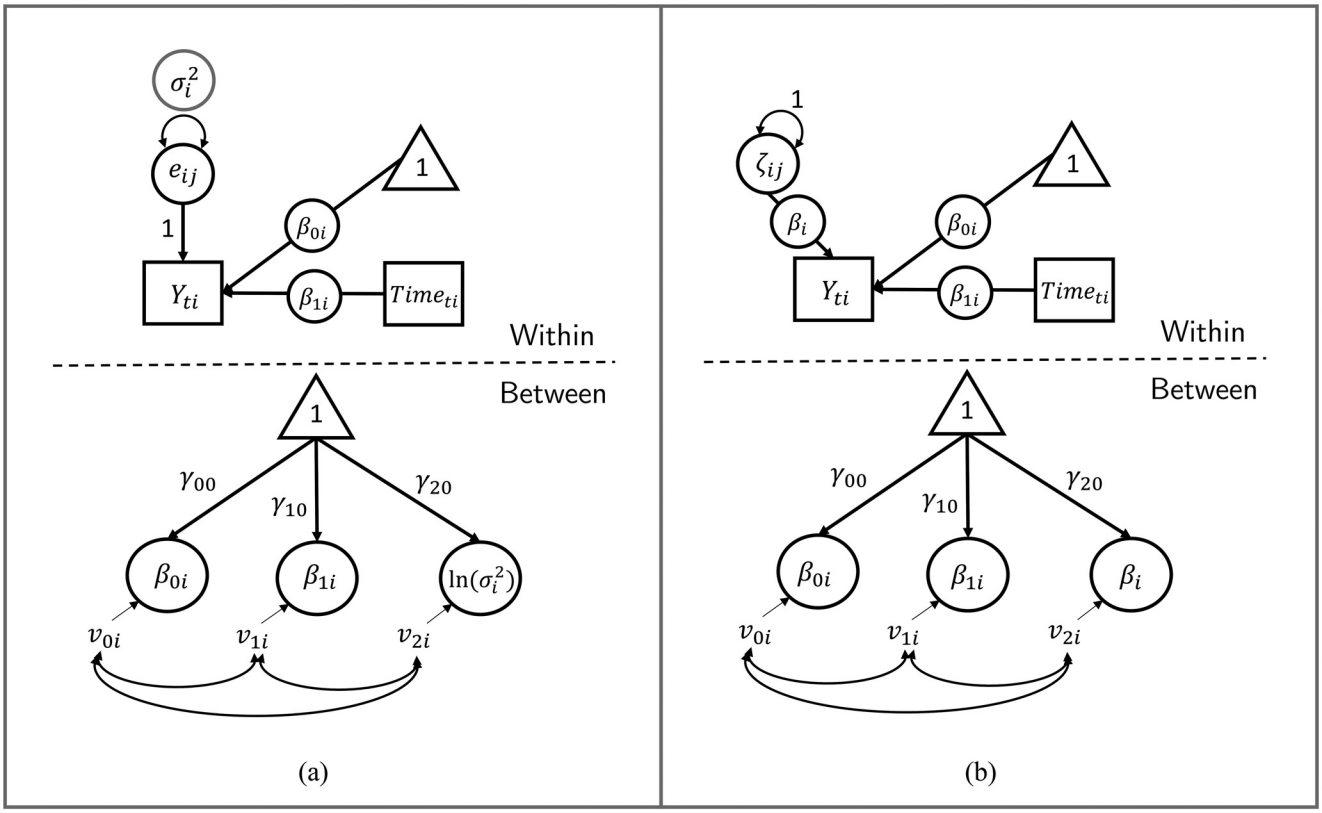
The analytical data used for this illustrative example includes 166 participants (135 females, 29 males, and two with other gender identities; M age = 25.41 years, SD = 7.35) who were instructed to complete a daily diary consisting of various survey items for 21 days after the lab visit. Besides the daily diary, participants also completed a survey to report their trait-level characteristics, including demographics, mental well-being, and general curiosity. Daily curiosity was assessed as part of the daily diary with two items taken from the Curiosity and Exploration Inventory-II (CEI-II; Kashdan et al., 2009): (a) "Today, I viewed challenging situations as an opportunity to grow and learn," and (b) "Everywhere I went today, I was out looking for new things or experiences." Participants responded to both the items on a slider from 0 (*not at all*) to 10 (*very*) in increments of .1, with a daily curiosity score derived as the average score of the two items. A plot showing daily fluctuations in this score for five randomly selected individuals is presented in Figure 7, showing for the sample that while some individuals are more consistent, others vary quite a bit from day to day.

We fit both an unconditional model (Model A) and a conditional model (see Figure 8) to the data in response to the following research questions discussed in Lydon-Staley et al. (2020): (a) Do daily fluctuations curiosity vary across individuals and if so, (b) do day-to-day curiosity fluctuations predict individual well-being after the trait-level curiosity is controlled for? In this example, the model parameters were estimated using Bayesian estimation via MCMC with Mplus v.8.6 (Muthén & Muthén, 1998–2021). The Mplus syntax files are provided in Supplemental Material B. For the unconditional model, a diffuse prior $N(0, 10^{10})$ was used for the fixed effects and an improper uniform prior inverse Wishart $W^{-1}(\mathbf{0}, -3)$ was used for the Level-2 random effects variance-covariance matrix.[4] The priors were set up similarly for the conditional model, except that the improper prior Inv-Gamma$(-1, 0)$ (i.e., a uniform prior on interval $[0, \infty]$) was used for the residual variance of the distal outcome. Three chains were run with a minimum of 5,000 iterations per chain, after which the iterations terminated either when the potential scale reduction (PSR) criterion falls below 1.05, or at the maximum number of iteration (50,000). PSR compares the between-chain variation against the within-chain variation (Gelman & Rubin, 1992), with a PSR value lower than 1.10 generally suggesting that the between-chain variation is small enough relative to the within-chain variation to be considered evidence of stochastic convergence. After a model converged, the posterior distributions were examined, with posterior means used as the parameter

[4] Given the relatively large sample size, the default noninformative priors were employed in the illustrative examples only for illustration purpose. We would suggest researchers make more informed decisions about the priors in practice, by incorporating prior information from theory and existing literature, especially when the sample size is small. For discussions about prior specification in the context of latent variable models, readers may consult Smid et al. (2019) and Zitzmann et al. (2020).

Figure 3

Random Variability Structural Model With One Continuous Observed Outcome Variable Y, Controlling for the Linear Trend Over Time



Note. A time variable ($Time_{ti}$) is included at Level-1 as a predictor. The individual intercept (β_{oi}) and slope (β_{1i}) are both modeled as random coefficients at level 2. (a) Model A: the log-transformed residual within-level variance ($\ln(\sigma_i^2)$) is modeled as a random coefficient; (b) Model B: the scale of the residual within-level variance (β_i) is modeled as a random coefficient at Level 2.

point estimates with 95% highest posterior density intervals (HPDI) as the interval estimates.[5] Both the unconditional model and conditional model converged in seconds (6 s and 12 s, respectively) on Mac OS, with 2.6 GHz 6-Core Intel Core i7 and 32 GB RAM.

The results of the unconditional model suggest that individuals differ from one another in their daily fluctuations of self-reported curiosity, with an average level of within-individual log-transformed residual variance of .810, and a between-individual variance of the log intraindividual variation of .690:

$$\hat{\gamma}^{(B)} = \begin{bmatrix} 3.111\,(2.842, 3.391) \\ 0.810\,(0.680, 0.951) \end{bmatrix},$$

$$\hat{\boldsymbol{\Phi}}_v = \begin{bmatrix} 3.375\,(2.600, 4.156) & \\ 0.734\,(0.450, 1.025) & 0.690\,(0.525, 0.884) \end{bmatrix}.$$

Translating everything onto the raw variance scale, the average level of intraindividual variance is estimated to be 3.186 with a 95% HPDI [2.657, 3.704], while the variance of intraindividual variance is estimated to be 10.089 with a 95% HPDI [5.240, 18.062]. The results also indicate, as seen in the covariance term below the diagonal, that the average level of daily curiosity is positively associated with the intraindividual fluctuation of curiosity; these results translate to an estimated correlation of approximately .48. Thus, people who have a higher level of average curiosity are expected to be less consistent from day to day in terms of their experience of curiosity feelings.[6]

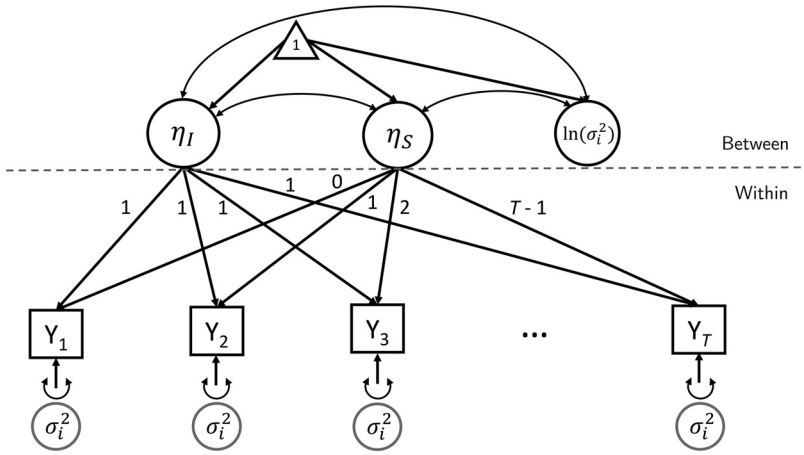
Next, a conditional model was fit to the data (see Figure 8), predicting individual flourishing with self-reported trait-level curiosity at baseline, model-implied individual average curiosity over time,

[5] Researchers can choose to interpret the estimation results either from a frequentist perspective or Bayesian perspective. Readers who are interested in HPDI's coverage properties when it is interpreted as the frequentist interval estimate may refer to Ghosh and Mukerjee (1993), Severini (1991), and Peers (1968) for more detailed discussions.

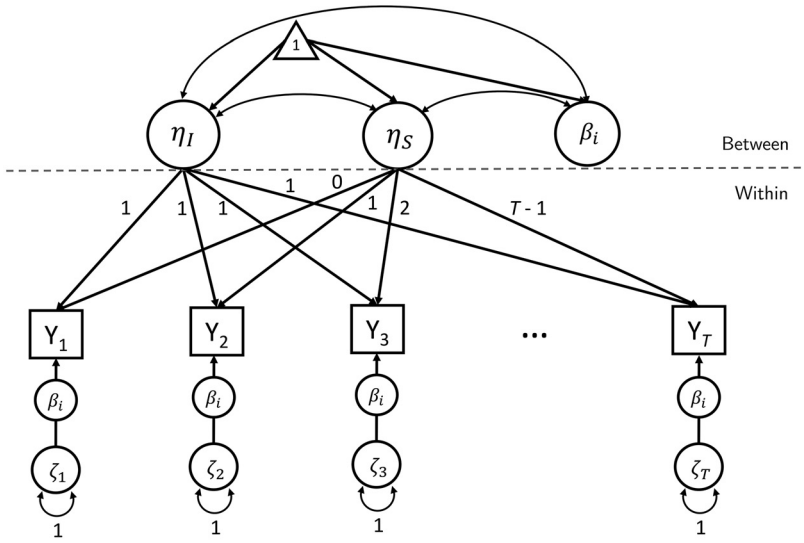
[6] We would caution, however, that such linear associations should not be interpreted causally without ruling out other possibilities that can introduce potential confounding. For instance, the dependence between individual level of curiosity and intraindividual variability in curiosity can be partly due to the bounded measurements (e.g., as discussed by Mestdagh et al., 2018).

Figure 4

Random Variability Structural Model With One Continuous Observed Outcome Variable Y, Controlling for the Linear Trend Over Time With a Latent Growth Model



(a)



(b)

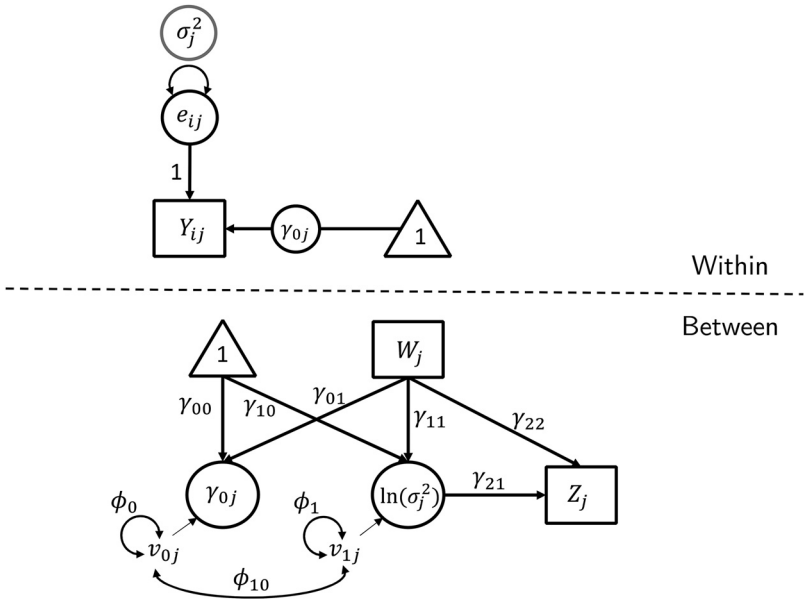
Note. The residual variance of the observed outcome after controlling for the latent growth factors is allowed to vary across individuals. (a) Model A: the log-transformed residual variance (ln (σ_i^2)) is modeled as a random coefficient at individual level; (b) Model B: a phantom variable (ζ_T) is introduced to approach for modeling the random within-individual variance.

and intraindividual fluctuation of daily curiosity. After controlling for the first two terms, the inconsistency of daily curiosity positively predicts personal flourishing ($\hat{\gamma}_{32} = .296$, 95% HPDI [.097, .499]), thus suggesting that the intraindividual variability provides unique information for predicting individuals' flourishing above and beyond the baseline and personal average trait-like characteristics. Specifically, people who show a greater variability from day to day in their curiosity are predicted to have higher-level of self-reported flourishing.

Example 2: Heterogeneous Intraschool Variability of Perceived Math Teacher Support

The second example illustrates an application of the proposed model to study the within-school heterogeneity in student perceptions, using the publicly available data from the Program for International Student Assessment (PISA), a large-scale international study administered by the Organization for Economic Cooperation and Development (OECD). The U.S.

Figure 5
Conditional Random Variability Structural Model With One Continuous Observed Outcome Variable Y

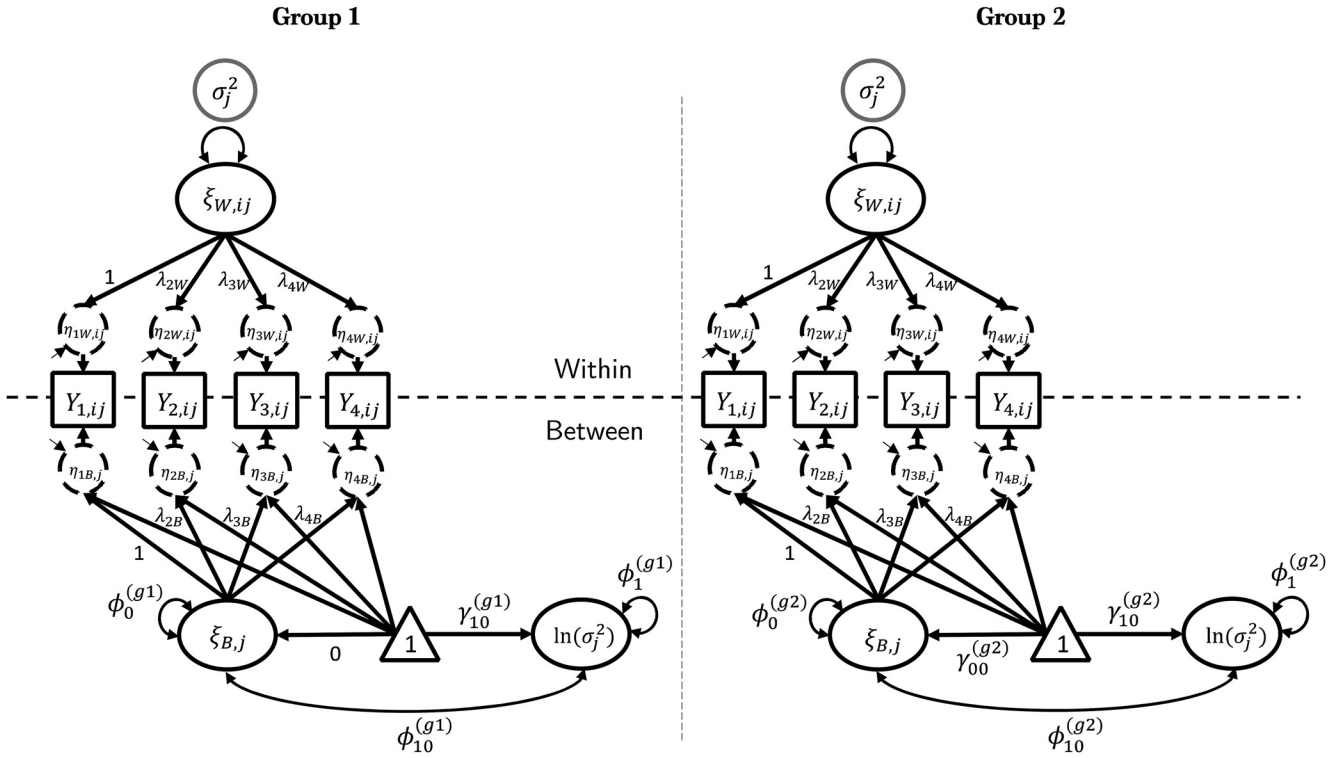


Note. The within-cluster variability is modeled as a mediator between cluster-level characteristic *W* and outcome *Z*. The log-transformed residual variance ($\ln(\sigma^2_i)$) is modeled as a random coefficient at between level.

data from PISA 2012 (OECD, 2013a, 2013b, 2013c) were used for this example, where the outcome of focal interest is the student perceived support from math teachers. The analytic dataset was downloaded via the EdSurvey R package (Bailey et al., 2021), containing the cross-sectional data from 3,232 students (1,580 females; *M* age = 15.52 years, *SD* = .5) in 162 schools. Three items asking how often each described scenario occurs in math classes were used as indicators of latent perceived support: (a) "The teacher gives extra help when students need it," (b) "The teacher helps students with their learning," and (c) "The teacher continues teaching until the students understand." Students responded to the items on a 4-point perceived frequency scale (1 = *every lesson*; 4 = *never or hardly ever*), with all items reverse coded prior to data analyses such that higher values correspond to higher levels of perceived support. Although a latent variable model is analyzed in this example, for visualization purposes Figure 9 shows a grouped boxplot of the distributions of averages of the three perceived support items for a random sample of 16 schools. As seen in the figure, schools differ from one another in terms of the within-school heterogeneity in student perceived support: in some schools students have similar levels of perceived support, while in other schools students have very different perceptions regarding how much support they receive from math teachers. For instance, while School 5 and School 155 have similar levels of average perceived support (*M* = 3.57 vs. *M* = 3.37), students in School 5 appear to be more homogenous in their perspectives than students in School 132 (σ^2 = .12 vs. σ^2 = .58).

To further investigate the variability of within-school heterogeneity in perceived support from math teachers across schools, we fit a random variability model to the data using the phantom variable approach (see Figure 10). We hypothesized that the observed responses were governed by a latent construct of perceived support both at the within-level as well as at the between-level. The between-level latent construct is measured by the school-level averages and the within-level latent construct is measured by the individual students' deviations from their school-level average. At the school level, the aggregated latent perceived support can be interpreted as an aspect of the school climate. For instance, a school that better trains teachers to provide support to students in math class may have a higher school-level perceived support, which in turn yields higher school averages for each item. At the student level, the latent construct speaks to how much support an individual student believes that they receive from the math teachers, relative to the school average. Within the same school, a student who perceives a higher-level of support is expected to respond more favorably on each item compared with other students. For this example, we are specifically interested in how intraschool heterogeneity in this student-level latent perception varies across schools. For simplicity of illustration, the item scores reported on the 4-point frequency scale are treated here as continuous variables.

Figure 6
Between-Subject Design for Random Variability Models With One Latent Outcome



Note. Between-subject design for random variability models with one latent outcome, comparing the within-cluster variability between two types of clusters (classified into Group 1 and Group 2). Measurement invariance can also be tested with this model.

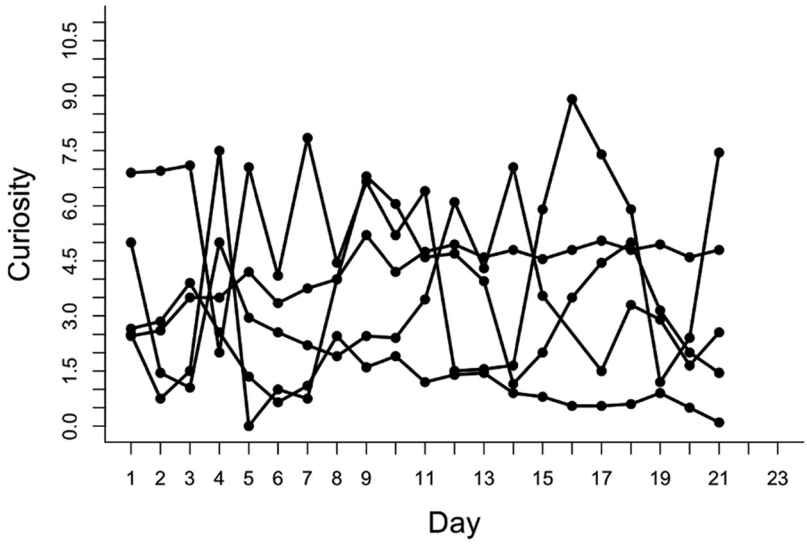
Before introducing the random variability model, we begin with a conventional MSEM model, where the within-cluster variance is fixed across all schools. The fit indices of the fixed variability model supported the hypothesized measurement structure with cross-level measurement invariance: $\chi^2(2) = .49$, $p = .783$; within-level SRMR = .004; between-level SRMR = .032. Next, we proceeded to fit a random variability model (see Figure 10), and the parameters are estimated using Bayesian estimation via MCMC with Mplus v.8.6 (Muthén & Muthén, 1998–2021). The Mplus syntax files are provided in Supplemental Material B. For the model set-up, the default priors in Mplus were utilized: diffuse prior $N(0, 10^{10})$ for all the free factor loadings, fixed effects, and item intercepts; improper prior Inv-Gamma(−1, 0) for both within-level and between-level individual item residual variances; and improper uniform prior $W^{-1}(\mathbf{0}, -3)$ for the between-level random effects variance-covariance matrix. The model was estimated with three chains. A minimum number of 4,000 iterations per chain was requested to avoid premature termination of iterations; the model converged with PSR = 1.082. The posterior means were then used as the point estimates for model parameters with 95% HPDI reported as the interval estimates. The model converged in 39 s on Mac OS, with 2.6 GHz 6-Core Intel Core i7 and 32 GB RAM. To check the sensitivity to prior specification, we also fit the model with a more informative prior (e.g., $N(6, 1)$ for the fixed effect of the scaling factor), which yielded similar results. Therefore, we will only discuss the results obtained with the default noninformative priors in this section:

$$\hat{\gamma}^{(B)} = \begin{bmatrix} 0 \\ 0.658\,(0.630,\ 0.684) \end{bmatrix},$$

$$\hat{\mathbf{\Phi}}_{\nu} = \begin{bmatrix} 0.039\,(0.026,\ 0.055) & \\ -0.028\,(-0.039,\ -0.020) & 0.021\,(0.014,\ 0.029) \end{bmatrix}.$$

The model estimation results also suggest that schools have varying levels of intraschool heterogeneity in student perceived support. On average, the within-school scaling factor is .658 with 95% HPDI[.630, .684], while the variance of the within-school scaling factor is estimated to be .021 with 95% HPDI [.014, .029]. That is to say, for some schools the scaling factor within school may be as low as .374 ($.658 - 1.96 \times \sqrt{.021}$; more homogeneous) and for some it may be as high as .942 ($.658 + 1.96 \times \sqrt{.021}$; more heterogeneous). To interpret the results on the variance scale, the posterior mean for the within-school variance is .454 with 95% HPDI [.418, .488], and the posterior mean for the variance of the within-school variance is .037 with 95% HPDI [.025, .051]. Therefore, on average, the within-school variance in latent perceived support is .454, and the variance of this within-school variance

Figure 7
Pattern of Daily Self-Reported Curiosity Over 21 Days for Five Randomly Selected Participants



across different schools is .037. Further, based on the covariance term, the results yield a negative association between the school-level perceived support and within-school heterogeneity of the student perceptions ($r = -.98$), suggesting that schools that have higher overall level of latent perceived support tend to be more homogeneous within the school in terms of how students feel they are supported by the math teachers.[7]

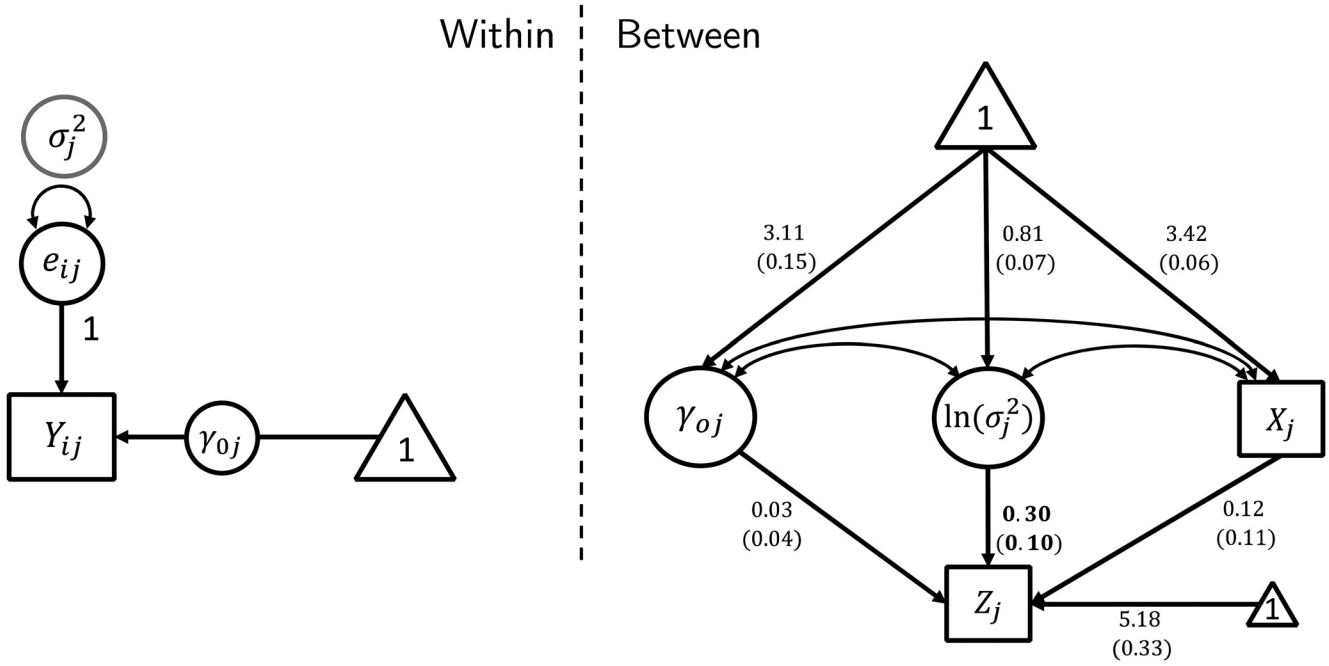
Example 3: Intraindividual Sleep Quality Consistency and Physical Health

In this example, the random variability models are employed to study intraindividual sleep quality consistency, specifically its relation with individual physical health. The data used for this example are from the Biomarker Project of Midlife in the United States Study series (MIDUS Refresher; Weinstein et al., 2012–2016). MIDUS is a national longitudinal study on American adults' health and well-being (Radler, 2014). Over decades MIDUS has been collecting data from various resources, such as lab assessment, self-administered surveys, bio-marker collection, and daily diary report. For this example, we utilize the biological assessment data collected in MIDUS Biomarker Project (MIDUS Refresher). As part of this study, a group of participants were instructed to wear a preprogrammed Actiwatch® activity monitor for seven days in a row; besides the daily activity data, they also provided one-time fasting blood draw, urine, and saliva samples for biomarker assays. By continuously monitoring individuals' movement, the Actiwatch generated a series of sleep quality measures for each day, including sleep onset latency (OL), sleep efficiency (EFF), and wake after sleep onset (WSO). Meanwhile, the biomarker specimen assays yielded various indicators reflecting different aspects of internal metabolic functioning, among which there are interleukin-6 (IL6), c-reactive protein (CRP), and fibrinogen (FGN), that are commonly treated as inflammation markers. Thus, this example addresses the relation between sleep inconsistency and physical health, an area in which scientists have shown increasing interest (e.g., Dzierzewski et al., 2020).

The analytic data for this example consisted of the complete Actiwatch sleep data and biomarker data collected from 90 adult males (M age = 53.66 years, SD = 13.22). Plots showing the patterns for each sleep quality indicator over the 7-day study period are provided in Figure 11. These images suggest that participants have varying levels of consistency in their sleep quality; some stay more stable while others fluctuate to a greater extent in the three observed sleep quality indicators. The random variability model is thus applied to further examine the intraindividual consistency in sleep quality. With repeated measures of OL, EFF, and WSO nested within individuals, we hypothesized both a within-level latent sleep quality as well as a between-level latent sleep quality. The between-level latent sleep quality is the overall sleep quality for an individual measured by the average OL, EFF, and WSO over the days, and the within-level latent sleep quality refers to an individual's sleep quality on each day measured by the fluctuations

[7] Such a high correlation suggests a strong linear dependence between the school levels of latent perceived support and within-school heterogeneity in student perceived support. Therefore, researchers may not wish to include both as predictors in a conditional model because of the collinearity. With this model, a high correlation at the latent variable level implies a high correlation between the school-level means and within-school variability in the observed indicators, which is evident in the observed data (r ranges from –0.67 to –0.81). Importantly, the estimation of the correlation between latent variables is essentially driven by the strong observed correlations. Because measurement error is removed, we also expect to see an even higher correlation at the latent variable level. The observed correlation in and of itself, however, as discussed earlier, can be a result of bounded measurements (Mestdagh et al., 2018). It is also possible that the relation between the school-level means and within-school variability in the observed indicators is nonlinear in nature and thus a linear association does not accurately characterize the true relation in population. As in the previous example, for illustration purpose, we do not further investigate the underlying mechanism for this linear association in the current manuscript and it should not be interpreted as a causal relation.

Figure 8
The Conditional Model Predicting Individual Flourishing



Note. The conditional model predicting individual flourishing (Z) with self-reported trait-level curiosity at baseline (X), model-implied individual average curiosity over time (γ_{0j}), and intraindividual fluctuation of daily curiosity ($\ln(\sigma^2)$). The posterior mean is listed as point estimate and the posterior standard deviation is displayed in parentheses.

in OL, EFF, and WSO relative to individual's average level. Additionally, we also hypothesized that there is a latent inflammation variable at the between-level measured by the three biomarkers: IL6, CRP, and FGN. With a just-identified Level-1 measurement model, we tested the fit of the hypothesized measurement model at the between-level. The baseline model with both saturated Level-1 and Level-2 covariance structure was compared with the model with saturated Level-1 covariance structure but hypothesized Level-2 measurement model. The results suggested the Level-2 measurement structure fits the data well: $\chi^2(8) = 8.019$, $p = .432$; between-level SRMR = .074. We thus proceeded with the hypothesized measurement structure in subsequent random variability modeling, but allowing the loadings for sleep quality to differ cross-level. An unconditional model with random log variance was first fit to the data using Bayesian estimation in Mplus, following the similar procedure as we did in the second example. The posterior mean of the log variance was –2.823 with 95% HPDI [–3.254, –2.355]; the posterior mean for the variance of the log variance was 2.592 with 95% HPDI [1.868, 3.360]. It thus suggests that within-individual consistency in sleep quality varies from person to person. In addition, the covariance between overall sleep quality and (in)consistency in sleep quality was estimated to be –.714 with 95% HPDI [–.991, –.429], indicating a negative association—people who have higher average sleep quality tend to experience less fluctuation (i.e., are more consistent) in their daily sleep quality.

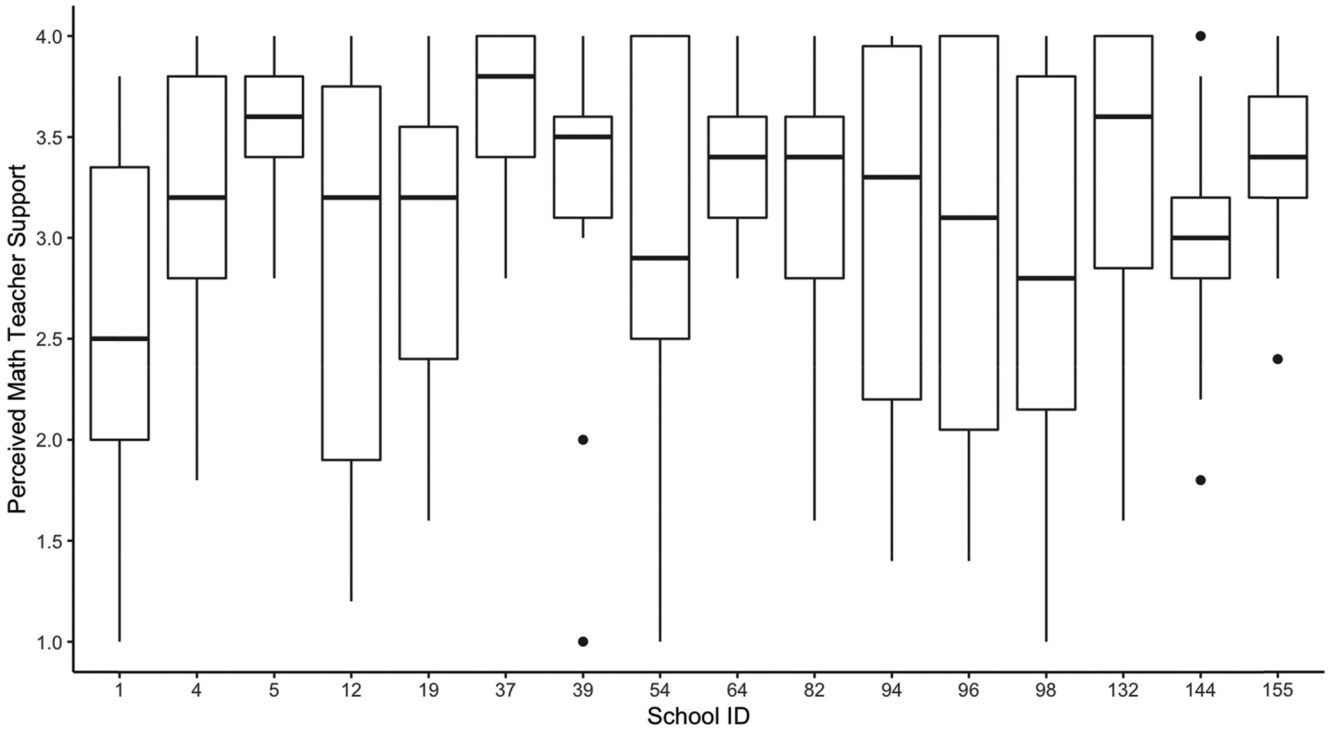
To address whether sleep consistency predicts individual inflammation reactions, we fit a conditional model to the data using Bayesian estimation (see Figure 12). The corresponding Mplus code is supplied in Supplemental Material B. The model was estimated with two chains and a minimum number of 13,000 iterations per chain. The MCMC posterior draws were thinned at every third iteration; the model converged with PSR = 1.09. The default diffuse priors in Mplus were used for the model parameters: diffuse prior $N(0, 10^{10})$ for all the factor loadings, fixed effects, item intercepts, Level-2 path coefficients, and Level-2 random effect variance-covariances; improper prior Inv-Gamma(–1, 0) for both the within-level and between-level individual item residual variances, as well as the Level-2 disturbance variance for latent inflammation. The model converged in 1.02 min on Mac OS, with 2.6 GHz 6-Core Intel Core i7 and 32 GB RAM. The posterior means and standard deviation for the model parameters are included in Figure 12. There appears to be a weakly negative association between average sleep quality and inflammation reaction in this sample, although the 95% HPDI [–3.276, .836] of this model parameter covers zero. Additionally, after controlling for the long-term average sleep quality, the intraindividual (in)consistency of daily sleep quality is barely predictive of inflammation reaction with a point estimate of .031 and 95% HPDI [–.892, .354].

Example 4: Intraindividual Positive Affect Consistency

In previous examples when intraindividual variability was of interest, we assumed there to be no systematic trend over time in the repeated measures, and thus the fluctuation at each measurement point was treated as idiosyncratic in nature. In some research scenarios, however, it may be necessary to account for the trend over time so that intraindividual consistency can be meaningfully

Figure 9

Distribution of Student Perceived Math Teacher Support Within Each School for a Random Sample of 16 Schools



Note. Perceived support in this graph is computed as the average score of the three observed indicators for illustration purpose.

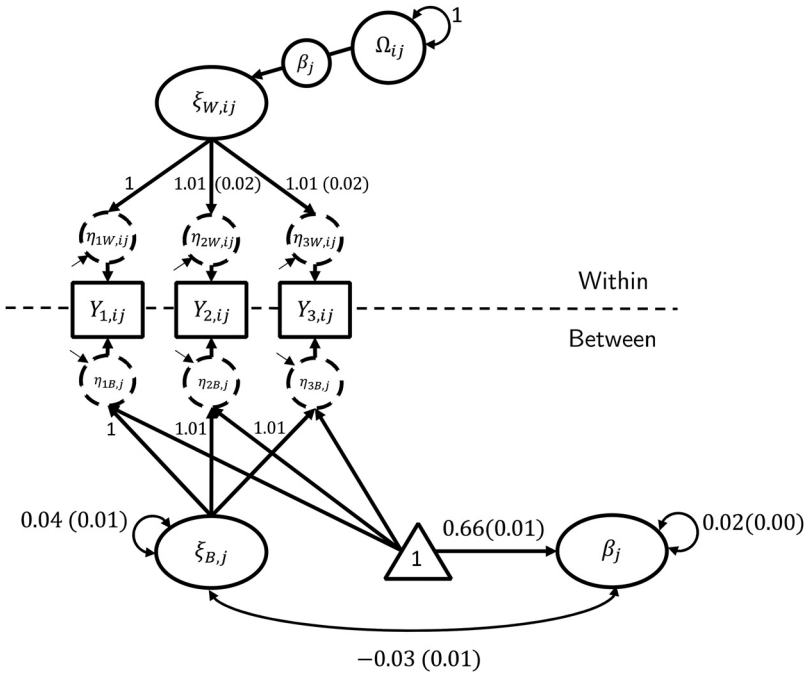
interpreted. As mentioned earlier, it is straightforward to extend the proposed SEM-based models to account for the systematic change over time. In this example, we demonstrate how to detrend the data by integrating a latent growth component into the proposed random variability models. The data used for this illustrative example come from the Daily Stress Project of MIDUS II (i.e., NSDE; Ryff & Almeida, 2004–2009) and MIDUS III (Ryff et al., 2013–2014). Participants in NSDE were interviewed on the telephone every day for 8 consecutive days and responded to questions about their daily experiences. People who completed the phone interview in MIDUS II were invited to participate in MIDUS III for follow-up phone interviews and self-administered surveys. The analytic sample for this example consists of those who participated in the follow-up wave (MIDUS III) as well as provided complete daily diary data over phone interview in MIDUS II (n = 1,251; 711 females and 540 males; M age = 56.6 years, SD = 11.84). The outcome of interest is positive affect, a latent construct indicated by 13 survey items administered during a phone interview. On each day, the respondents were instructed to indicate how much of the time on that day they felt: (a) in good spirits, (b) cheerful, (c) extremely happy, (d) calm and peaceful, (e) satisfied, (f) full of like, (g) close to others, (h) like you belong, (i) enthusiastic, (j) attentive, (k) proud, (l) active, and (m) confident. Participants responded to each of the items on a 5-point frequency scale (0 = *none of the time*; 4 = *all of the time*). Based on the descriptive plot in Figure 13 showing the mean of the 13 items, it appears that the participants do not only experience varying levels of overall positive affect but also demonstrate different levels of intraindividual consistency in daily positive affect. The hypothetical research scenario is to investigate to what extent people vary in terms of intraindividual consistency in the underlying latent positive affect.

Traditionally, an average score across the 13 items is treated as the unobserved positive affect, treating the aggregate as though there is no measurement error and assuming every item is equally good indicator of the underlying latent construct. Alternatively, we chose to model latent positive affect with a one-factor CFA model at each time point (i.e., within-level) to explicitly account for measurement error and validate the measurement structure via formal testing. With measurement error accommodated, the fluctuation of the latent positive affect at each measurement point was then assumed to result from both a potential systematic linear trend over time and an idiosyncratic process that reflects intraindividual consistency. To account for the growth trend, a second-order latent linear growth model (LGM) was imposed at the person level with measurement invariance[8] constrained across time as shown in Figure 14 (see Hancock et al., 2001, for a treatment on second-order latent growth models). Because the same set of items were used across the days, we further allowed the residuals for each individual

[8] The measurement invariance over intensive longitudinal measurements can also be tested by fitting a cross-classified factor analysis model and testing the variance of random time effects on factor loadings and intercepts. For more details about this approach, readers can refer to McNeish et al. (2021).

Figure 10

Multilevel Structural Equation Model With Random Scaling Factor Fitted to the PISA 2012 Data, Assessing the Variation Between Schools Regarding Within-School Heterogeneity in Student Perceived Support From Math Teachers



Note. Three survey items (Y_1–Y_3) are used as the indicators for the latent constructs at both the within-level and between-level. The factor loadings are constrained to be equal at both levels. The posterior mean is displayed as the point estimate, with posterior standard deviation included in parentheses.

item to covary across different measurement points, as people are likely to respond to the same item similarly on different days even after controlling for the underlying positive affect. Importantly, different from the conventional parameterization of second-order LGM, in this model we assumed that the disturbance terms of the latent construct at each time point come from a normal distribution that varies from individual to individual. The log-transformed variance of the disturbance terms was in turn modeled as a random coefficient at the person-level, which was allowed to covary with the latent growth factors.

To start the modeling process, a conventional second-order LGM with fixed disturbance variances across individuals was first fit to the data to check whether the specification of the hypothetical measurement model and growth model was reasonable. The model had acceptable model-data fit: $\chi^2(5059) = 15094.189$, $p < .001$; RMSEA = .040, 90% CI [.039, .041]; SRMR = .073. Therefore, building upon this second-order LGM, we further modeled the random variability at the person-level with the log variance approach (see Figure 14). The second-order LGM random variability model was estimated using Bayesian estimation with MCMC using `rjags` in R (Plummer, 2016; RStudio Team, 2020). The model syntax defined in `BUGS` language (Lunn et al., 2009) for this example is provided in Supplemental Material B.[9] As in previous examples, we employed noninformative priors for all model parameters: $N(0, 10^7)$ for all the free factor loadings, item intercepts, and Level-2 fixed effects, Inverse Wishart distribution $W^{-1}(\mathbf{I}_8, 9)$ for the item residual variance-covariance matrices, and $W^{-1}(\mathbf{I}_3, 4)$ for the variance-covariance matrix for the between-level random coefficients. We estimated the model with three chains and 20,000 iterations per chain after 10,000 iterations of burn-in. The model converged with PSR = 1.011, after 11.35 hr of running time on Windows 10, with 1.90 GHz 4-Core Intel Core i7 and 16 GB RAM. The posterior means and 95% HPDI for the Level-2 fixed effects were:

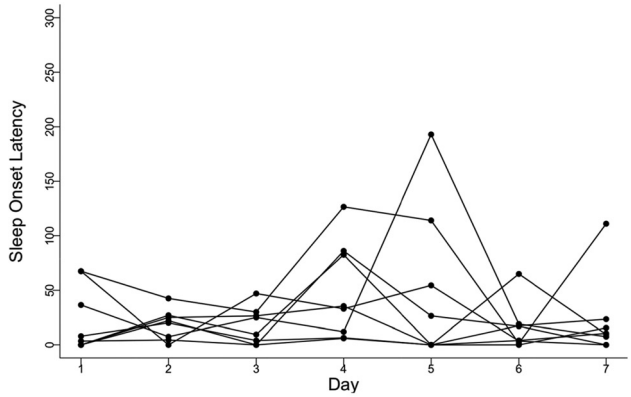
$$\hat{\boldsymbol{\gamma}}^{(\mathrm{B})} = \begin{bmatrix} 2.997\,(2.960,\ 3.034) \\ 0.004\,(-0.000,\ 0.008) \\ -3.133\,(-3.241,\ -3.022) \end{bmatrix},$$

corresponding to the average latent intercept, latent slope, and latent (in)consistency, respectively. The posterior means and 95% HPDI for the Level-2 random effect variance-covariances were:

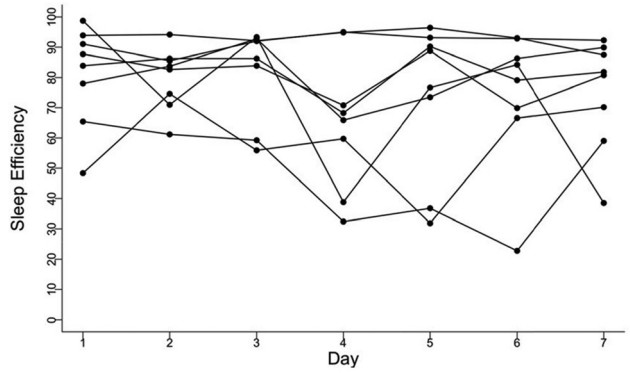
[9] As a reference for interested readers, the corresponding Stan model syntax is also supplied in Supplemental Material B, which yielded very similar results as rjags. It is noted that the Hamiltonian Monte Carlo (HMC) sampling in Stan took a longer run-time for this example, compared with the Gibbs sampler in JAGS. The running time for three chains with 5,000 iterations per chain and 1,000 iterations of warm-up is 62.59 hr on a Windows laptop.

Figure 11

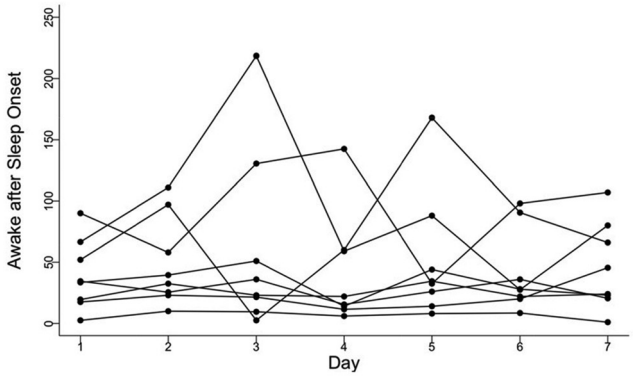
Patterns of the Three Observed Sleep Quality Indicators on Their Raw Scale Over the 7-Day Period, Plotted for a Random Sample of Eight Males



(a)



(b)



(c)

$$\hat{\Phi}_v = \begin{bmatrix} 0.326\,(0.294,\ 0.358) & & \\ 0.000\,(-0.002,\ 0.003) & 0.003\,(0.003,\ 0.004) & \\ -0.270\,(-0.328,\ -0.208) & 0.002\,(-0.004,\ 0.009) & 2.018\,(1.760,\ 2.303) \end{bmatrix},$$

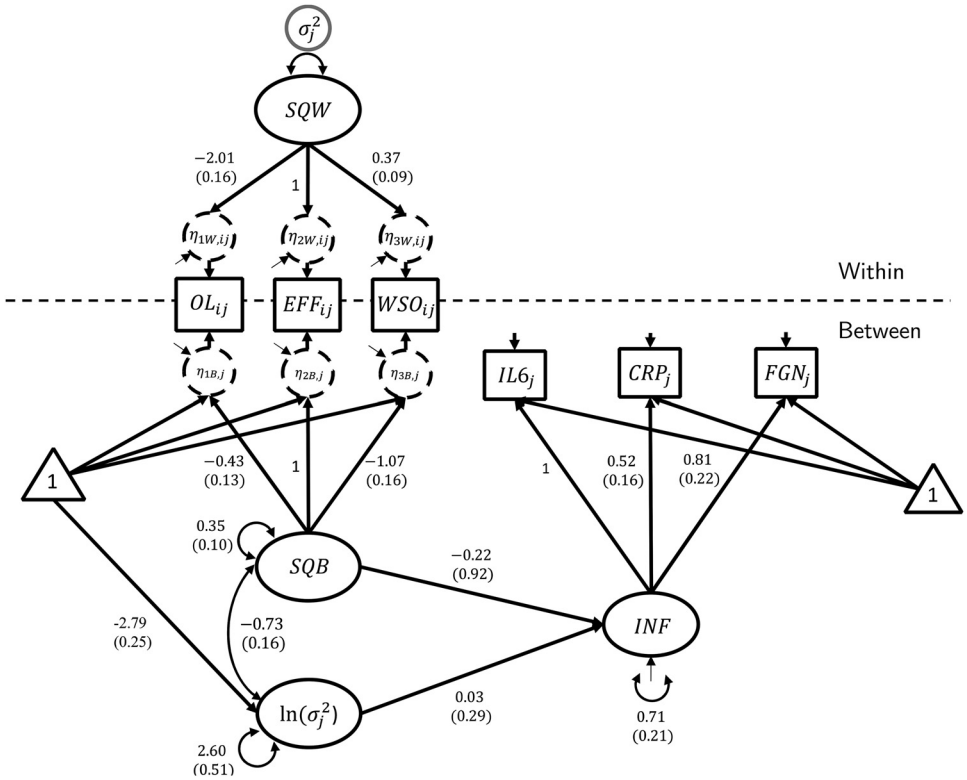
with the rows and columns corresponding to the latent intercept, latent slope, and latent inconsistency in this order. The results suggest no evidence of a linear growth trend for positive affect over time; nevertheless, after controlling for trend, the extent to which individuals fluctuate in daily positive affect differed from person to person. The average inconsistency defined as the within-individual log disturbance variance was –3.133 with 95% HPDI [–3.241, –3.022], and the variance of the inconsistency was 2.018 with 95% HPDI [1.760, 2.303]. Further, the individual inconsistency in daily positive affect was negatively associated with the initial level of positive affect, such that people who started with higher positive affect were likely to experience more stable positive affect over the time span. To interpret the intraindividual inconsistency on a variance scale, we also monitored the posterior distribution of the mean and variance of the within-individual disturbance variance via the transformed moments approach discussed earlier. On average, the within-individual disturbance variance was estimated to be .120 with 95% HPDI [.103, .138], while the variance of the within-individual disturbance variance was estimated to be .097 with 95% HPDI [.051, .157].

Discussion

As a characteristic that potentially plays an important role in human development, group performance, and beyond, there has been growing interest in questions related to variability across a wide variety of research fields. Unfortunately, whether in the form of intraindividual consistency/fluctuation over repeated measures, or intragroup cohesion/heterogeneity among group members, this interest has been met with methodological barriers that have impeded such scientific inquiries, primarily due to traditional statistical methods' focus on modeling means while treating parameters related to variability as fixed, peripheral, and/or a complete nuisance. In short, opportunities have been lost without comprehensive methodological tools and guidance on how to operationalize variability, how to model variability, and how to interpret statistical results informing relevant conclusions thereof. The current study is in response to this collective need, introducing a system of structural models for random variability modeling applicable to investigations of intraindividual or intragroup variability. Specifically, by specifying a random coefficient for either the log-transformed variance or the scaling factor, variability can be conveniently modeled as a latent variable. These approaches, built upon the developments in SEM and MSEM, have the flexibility to embed variability within a broader covariance and mean structure, which can then be further extended to accommodate many complex research scenarios.

Aiming to complement currently existing methods that can accommodate heterogeneous variability, the modeling approaches discussed in this study have wider applicability across different research contexts involving observed/latent outcomes, covariates, mediators, and distal outcomes. While in some research scenarios single observed measures can be meaningfully studied and interpreted, in many cases—especially within the social and behavioral sciences—researchers are often interested in underlying latent constructs that are not directly observable (e.g., perception, attitude, ability). Existing analytical methods for studying variability, however, generally assume no measurement error, and therefore either only focus on single observed variables or an aggregate across multiple observed indicators as a (potentially crude) operationalization

Figure 12
Conditional Model Predicting Inflammation Reaction With Sleep Quality Consistency and Average Sleep Quality



Note. OL = sleep onset latency; EFF = sleep efficiency; WSO = wake after sleep onset; INF = inflammation; SQW = within-level sleep quality; SQB = between-level sleep quality.

of a latent construct of interest. By treating the observed score as a perfect measure of the underlying latent construct, the outcome itself as well as its variability are conflated with variation due to measurement error, which may in turn lead to biased results and reduced power for statistical tests regarding variability. Studying variability within the framework of SEM, on the other hand, allows the possibility of incorporating and testing measurement models, such that the structural features of variability are not driven by fallible measures.

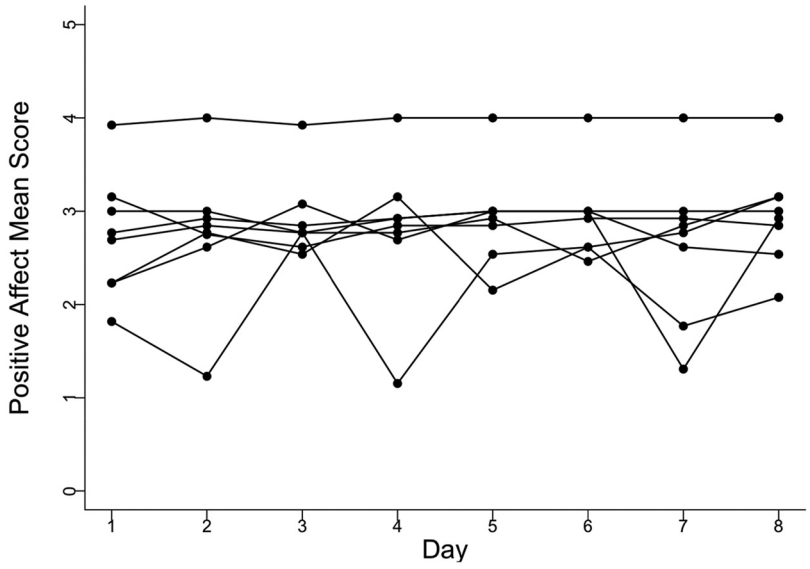
Of course in some research contexts the focal outcome can indeed be meaningfully represented by a single measure variables (e.g., response time, blood pressure, heart rate), and therefore researchers do not need to consider the integration of measurement models. In such cases, model-based approaches are still preferable to multistage strategies that treat point estimates of variation as if they were observed variables, and LSM is a clever multilevel approach that can meet such research needs. It is important to note that the SEM-based strategy discussed in this article is not incompatible with the MLM-based approach. Indeed, the development of MSEM shows that when the two frameworks come together, we can enjoy greater flexibility than when we rely on a single framework alone. In fact, LSM can be conveniently specified and estimated as a MSEM or DSEM model (e.g., McNeish, 2020; Nestler, 2021). Readers may also find that Model A for observed outcomes with random log-variance shares great similarities with LSM,[10] but with the key exception that Model A can be expanded to model the random log-variance as predictor, outcome, and/or mediator simultaneously in one step. Therefore, whether the research focus is on observed variables or latent variables, the SEM-based framework for random variability serves as a powerful complement to currently available analytical tools.

Most importantly, the SEM-based approaches for random variability have great potential to grow into an even more comprehensive modeling framework that can be flexibly adapted to a wide range of research scenarios (see Table 1). Based on the foundational modeling structure presented in the current study, in which variability is parameterized as a random latent variable

[10] In its most general form, Model A does not specify the within-level residual variance as varying across within-level units, which is different from the LSM parameterization (see Equation 2). Although it is possible to also include time varying covariates for intraindividual variability in Model A, we chose not to do so considering it does not make much theoretical sense for intragroup variability (e.g., it is hard to interpret if the intragroup variability is individual-specific).

Figure 13

Daily Positive Affect Mean Score Over the 8-Day Study Period for Eight Randomly Selected Subjects



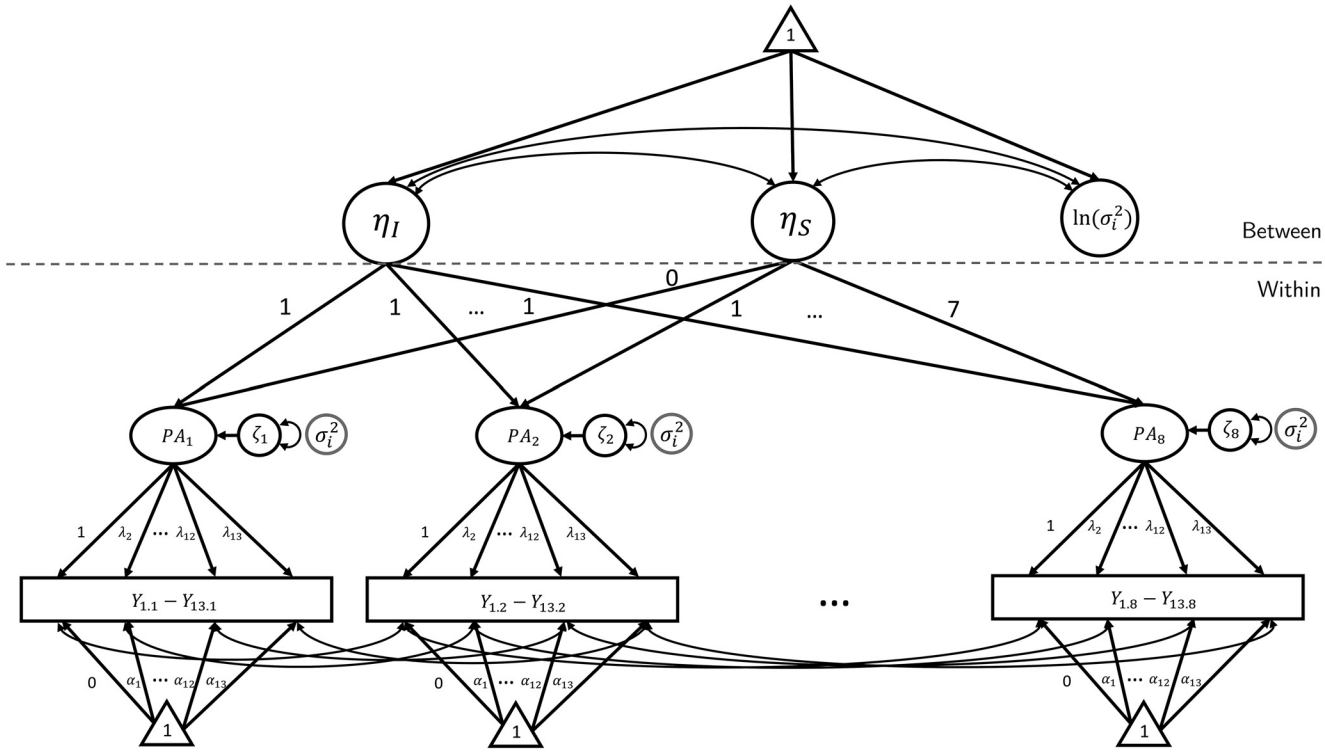
Note. The mean score is only used here for easier data visualization.

in a structural model, various features of variability can be investigated, opening up a plethora of new possibilities for research on variability across different fields. Future studies can further investigate and illustrate, for example, models for between-subjects designs, within-subject designs, longitudinal models for variability, parallel processes with multiple outcomes considered simultaneously, mixture modeling for random variability, and much more.

With the many advantages and great potential of the proposed modeling framework, there are some caveats to be noted as well. The current study introduced and illustrated two different parameterization strategies for modeling random variability: variability as the natural log-transformed variance and variability as the random scaling factor by introducing a phantom variable. Conceptually, both are viable for random variability modeling, although the different parameterization strategies represents different distributional assumptions about the variance. Utilizing the log transformation approach implicitly assumes that the within-level variance follows a lognormal distribution (via the exponential function), while modeling a random scaling factor with phantom variable assumes the within-level variance follows a scaled noncentral χ^2 distribution (via the quadratic function). As such, the exponential function is monotonically increasing over its entire domain while the quadratic function is only monotonically increasing on $(0, \infty)$. Therefore, the Level-2 parameters of the scaling factor may not be interpretable if the random coefficient β_j takes on both positive and negative values. Depending on how far β_j falls from zero, additional positivity constraints may be needed (e.g., to force the MCMC draws to sample from a truncated distribution). Although this issue is of less concern in practice when individuals actually fluctuate to a great extent or group members differ considerable from one another, it may yield misleading results when there is minimal within-level variability for the individuals or groups in the sample. In such cases, the scaling factor approach may overestimate the within-level variability without the positivity constraints. Future research could look into this issue by systematically evaluating the performance of the phantom variable approach with and without these constraints under different scenarios.

Although beyond the scope of the current study, a thorough treatment of model fit assessment within the context of SEM-based random variability modeling would be welcome. As the model for random variability is rather complex to begin with, model misspecification at any layer can adversely impact parameter estimation across the entire model. In the illustrative examples, we only briefly mentioned model fit. For instance, in the case of a multilevel CFA model with random variability (Model C and Model D), we examined the model (mis)fit in three separate aspects whenever applicable: (a) measurement model at the within level, (b) measurement model at the between level, and (c) heterogeneous within-level variability structure. The first two sources of model misfit are assessed by comparing the saturated model to the corresponding reduced model with a restricted structure at either the within level or between level, via ML-based model fit indices (e.g., χ^2). The third piece of model misfit can be assessed by comparing the random variability model to the reduced model that constrained zero variation of the within-level variability, via the information criterion (e.g., DIC). Although the literature has provided useful guidance regarding model fit assessment for MSEM (Ryu, 2014; Ryu & West, 2009) and Bayesian SEM (Garnier-Villarreal & Jorgensen, 2020; Levy, 2011), it is not entirely clear how to properly assess model data fit for random

Figure 14
Random Variability Model Fitted to the Fourth Illustrative Example



Note. $Y_{k.t}$ indicates the observed indicator k at time t. PA = positive affect. The factor loadings and intercepts are constrained to be equal across 8 days. The residual term and variance for the observed indicators are omitted in the diagram. A linear latent growth is specified at the between-level, with a linear intercept (η_I) and a linear slope (η_S), along with the random log variance.

variability models in the context of Bayesian MSEM. Therefore, it would be important for future methodological research to investigate optimal procedures and fit indices for assessment of random variability models.

In summary, the current study introduced and delineates SEM-based approaches for modeling variability as a random latent variable. These not only complement and unify the existing analytical methods for variability, but they also open up many possibilities for more advanced methodological and applied developments. It is our hope that the proposed models provide applied researchers with relatively straightforward tools that can be implemented to answer a wide range of research questions about variability. We also look forward to future research on SEM-based variability modeling that further expands this framework.

References

Adolph, K. E., Cole, W. G., & Vereijken, B. (2015). Intraindividual variability in the development of motor skills in childhood. In M. Diehl, K. Hooker, & M. Sliwinski (Eds.), *Handbook of intraindividual variability across the lifespan* (pp. 59–83). Routledge/Taylor & Francis Group.

Asparouhov, T., Hamaker, E. L., & Muthén, B. (2018). Dynamic structural equation models. *Structural Equation Modeling*, *25*(3), 359–388. https://doi.org/10.1080/10705511.2017.1406803

Asparouhov, T., & Muthén, B. (2007). Computationally efficient estimation of multilevel high-dimensional latent variable models. *Proceedings of the 2007 Joint Statistical Meetings, Section on Statistics in Epidemiology* (pp. 2531–2535). American Statistical Association. https://www.statmodel.com/download/JSM2007000746.pdf

Asparouhov, T., & Muthén, B. (2016). General random effect latent variable modeling: Random subjects, items, contexts, and parameters. In J. R. Harring, L. M. Stapleton, & S. N. Beretvas (Eds.), *Advances in multilevel modeling for educational research: Addressing practical issues found in real-world applications* (pp. 163–192). Information Age Publishing, Inc.

Bailey, P., Emad, A., Huo, H., Lee, M., Liao, Y., Lishinski, A., Nguyen, T., Xie, Q., Yu, J., Zhang, T., Buehler, E., Lee, S., Sikali, E., Bundsgaard, J., C'deBaca, R., & Christensen, A. A. (2021). EdSurvey: Analysis of NCES education survey and assessment data (R package version 2.6.9) [Computer software]. https://CRAN.R-project.org/package=EdSurvey

Barnard, J., McCulloch, R., & Meng, X. L. (2000). Modeling covariance-matrices in terms of standard deviations and correlations, with applicationto shrinkage. *Statistica Sinica*, *10*(4), 1281–1311.

Bauer, D. J. (2003). Estimating multilevel linear models as structural equation models. *Journal of Educational and Behavioral Statistics*, *28*(2), 135–167. https://doi.org/10.3102/10769986028002135

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M. A., Guo, J., Li, P., & Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, *76*(1), 1–32. https://doi.org/10.18637/jss.v076.i01

Cox, T. H., & Blake, S. (1991). Managing cultural diversity: Implications for organizational competitiveness. *The Academy of Management Perspectives*, *5*(3), 45–56. https://doi.org/10.5465/ame.1991.4274465

Curran, P. J. (2003). Have multilevel models been structural equation models all along? *Multivariate Behavioral Research*, *38*(4), 529–569. https://doi.org/10.1207/s15327906mbr3804_5

Dzierzewski, J. M., Donovan, E. K., Kay, D. B., Sannes, T. S., & Bradbrook, K. E. (2020). Sleep inconsistency and markers of inflammation. *Frontiers in Neurology*, *11*, 1042. https://doi.org/10.3389/fneur.2020.01042

Dzubur, E., Ponnada, A., Nordgren, R., Yang, C. H., Intille, S., Dunton, G., & Hedeker, D. (2020). MixWILD: A program for examining the effects of variance and slope of time-varying variables in intensive longitudinal data. *Behavior Research Methods*, *52*(4), 1403–1427. https://doi.org/10.3758/s13428-019-01322-1

Easley, C. A. (2001). Developing valuing and managing diversity in the new millennium. *Organization Development Journal*, *19*(4), 38–51.

Fagot, D., Mella, N., Borella, E., Ghisletta, P., Lecerf, T., & De Ribaupierre, A. (2018). Intra-individual variability from a lifespan perspective: A comparison of latency and accuracy measures. *Journal of Intelligence*, *6*(1), 16–33. https://doi.org/10.3390/jintelligence6010016

Garnier-Villarreal, M., & Jorgensen, T. D. (2020). Adapting fit indices for Bayesian structural equation modeling: Comparison to maximum likelihood. *Psychological Methods*, *25*(1), 46–70. https://doi.org/10.1037/met0000224

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, *7*(4), 457–472. https://doi.org/10.1214/ss/1177011136

Ghosh, J. K., & Mukerjee, R. (1993). Frequentist validity of highest posterior density regions in the multiparameter case. *Annals of the Institute of Statistical Mathematics*, *45*(2), 293–302. https://doi.org/10.1007/BF00775815

Goldstein, H., & McDonald, R. P. (1988). A general model for the analysis of multilevel data. *Psychometrika*, *53*(4), 455–467. https://doi.org/10.1007/BF02294400

Hamaker, E. L. (2012). Why researchers should think "within-person": A paradigmatic rationale. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 43–61). Guilford Press.

Hancock, G. R., Kuo, W. L., & Lawrence, F. R. (2001). An illustration of second-order latent growth models. *Structural Equation Modeling*, *8*(3), 470–489. https://doi.org/10.1207/S15328007SEM0803_7

Hedeker, D., & Mermelstein, R. (2007). Mixed-effects regression models with heterogeneous variance: Analyzing ecological momentary assessment data of smoking. In T. D. Little, J. A. Bovaird, & N. A. Card (Eds.), *Modeling ecological and contextual effects in longitudinal studies of human development* (pp. 183–206). Oxford University Press.

Hedeker, D., & Mermelstein, R. J. (2012). Mood changes associated with smoking in adolescents: An application of a mixed-effects location scale model for longitudinal Ecological Momentary Assessment (EMA) data. In G. R. Hancock & J. R. Harring (Eds.), *Advances in longitudinal methods in the social and behavioral sciences* (pp. 59–79). Information Age Publishing.

Hedeker, D., Mermelstein, R. J., & Demirtas, H. (2008). An application of a mixed-effects location scale model for analysis of ecological momentary assessment (EMA) data. *Biometrics*, *64*(2), 627–634. https://doi.org/10.1111/j.1541-0420.2007.00924.x

Hedeker, D., & Nordgren, R. (2013). MIXREGLS: A program for mixed-effects location scale analysis. *Journal of Statistical Software*, *52*(12), 1–38. https://doi.org/10.18637/jss.v052.i12

Hox, J. J., Moerbeek, M., & van de Schoot, R. (2018). *Multilevel analysis: Techniques and applications* (3rd ed.). Routledge.

Hultsch, D. F., & MacDonald, S. W. (2004). Intraindividual variability in performance as a theoretical window onto cognitive aging. In R. A. Dixon, L. Backman, & L.-G. Nilsson (Eds.), *New frontiers in cognitive aging* (pp. 65–88). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198525691.003.0004

Jehn, K. A., Northcraft, G. B., & Neale, M. A. (1999). Why differences make a difference: A field study of diversity, conflict and performance in workgroups. *Administrative Science Quarterly*, *44*(4), 741–763. https://doi.org/10.2307/2667054

Jongerling, J., Laurenceau, J. P., & Hamaker, E. L. (2015). A multilevel AR (1) model: Allowing for inter-individual differences in trait-scores, inertia, and innovation variance. *Multivariate Behavioral Research*, *50*(3), 334–349. https://doi.org/10.1080/00273171.2014.1003772

Kapur, K., Li, X., Blood, E. A., & Hedeker, D. (2015). Bayesian mixed-effects location and scale models for multivariate longitudinal outcomes: An application to ecological momentary assessment data. *Statistics in Medicine*, *34*(4), 630–651. https://doi.org/10.1002/sim.6345

Kashdan, T. B., Gallagher, M. W., Silvia, P. J., Winterstein, B. P., Breen, W. E., Terhar, D., & Steger, M. F. (2009). The curiosity and exploration inventory-II: Development, factor structure, and psychometrics. *Journal of Research in Personality*, *43*(6), 987–998. https://doi.org/10.1016/j.jrp.2009.04.011

Kaufhold, S. P., & van Leeuwen, E. J. C. (2019). Why intergroup variation matters for understanding behaviour. *Biology Letters*, *15*(11), 20190695. https://doi.org/10.1098/rsbl.2019.0695

Knebel, D., Ayali, A., Guershon, M., & Ariel, G. (2019). Intra- versus intergroup variance in collective behavior. *Science Advances*, *5*(1), eaav0695. https://doi.org/10.1126/sciadv.aav0695

Kupers, E., Lehmann-Wermser, A., McPherson, G., & van Geert, P. (2019). Children's creativity: A theoretical framework and systematic review. *Review of Educational Research*, *89*(1), 93–124. https://doi.org/10.3102/0034654318815707

Lang, J. W., Bliese, P. D., & de Voogt, A. (2018). Modeling consensus emergence in groups using longitudinal multilevel methods. *Personnel Psychology*, *71*(2), 255–281. https://doi.org/10.1111/peps.12260

Leckie, G., French, R., Charlton, C., & Browne, W. (2014). Modeling heterogeneous variance–covariance components in two-level models. *Journal of Educational and Behavioral Statistics*, *39*(5), 307–332. https://doi.org/10.3102/1076998614546494

Lee, Y., & Nelder, J. A. (2006). Double hierarchical generalized linear models (with discussion). *Journal of the Royal Statistical Society. Series C, Applied Statistics*, *55*(2), 139–185. https://doi.org/10.1111/j.1467-9876.2006.00538.x

Levy, R. (2011). Bayesian data-model fit assessment for structural equation modeling. *Structural Equation Modeling*, *18*(4), 663–685. https://doi.org/10.1080/10705511.2011.607723

Lunn, D., Spiegelhalter, D., Thomas, A., & Best, N. (2009). The BUGS project: Evolution, critique and future directions. *Statistics in Medicine*, *28*(25), 3049–3067. https://doi.org/10.1002/sim.3680

Lydon-Staley, D. M., Zurn, P., & Bassett, D. S. (2020). Within-person variability in curiosity during daily life and associations with well-being. *Journal of Personality*, *88*(4), 625–641. https://doi.org/10.1111/jopy.12515

McDonald, R. P., & Goldstein, H. (1989). Balanced and unbalanced designs for linear structural relations in two-level data. *British Journal of Mathematical & Statistical Psychology*, *42*(2), 215–232. https://doi.org/10.1111/j.2044-8317.1989.tb00911.x

McNeish, D. (2020). Specifying location-scale models for heterogeneous variances as multilevel SEMs. *Organizational Research Methods*, *24*(3), 630–653. https://doi.org/10.1177/1094428120913083

McNeish, D., & Hamaker, E. L. (2020). A primer on two-level dynamic structural equation models for intensive longitudinal data in Mplus. *Psychological Methods*, *25*(5), 610–635. https://doi.org/10.1037/met0000250

McNeish, D., MacKinnon, D. P., Marsch, L. A., & Poldrack, R. A. (2021). Measurement in intensive longitudinal data. *Structural Equation Modeling*, *28*(5), 807–822. https://doi.org/10.1080/10705511.2021.1915788

McNeish, D., & Wolf, M. G. (2020). Thinking twice about sum scores. *Behavior Research Methods*, *52*(6), 2287–2305. https://doi.org/10.31234/osf.io/3wy47

Mestdagh, M., Pe, M., Pestman, W., Verdonck, S., Kuppens, P., & Tuerlinckx, F. (2018). Sidelining the mean: The relative variability index as a generic mean-corrected variability measure for bounded variables. *Psychological Methods*, *23*(4), 690–707. https://doi.org/10.1037/met0000153

Muthén, B. O. (1989). Latent variable modeling in heterogeneous populations. *Psychometrika*, *54*(4), 557–585. https://doi.org/10.1007/BF02296397

Muthén, B. O. (1994). Multilevel covariance structure analysis. *Sociological Methods & Research*, *22*(3), 376–398. https://doi.org/10.1177/0049124194022003006

Muthén, L. K., & Muthén, B. O. (1998–2021). *Mplus user's guide* (8th ed.).

Nestler, S. (2020). Modelling inter-individual differences in latent within-person variation: The confirmatory factor level variability model. *British Journal of Mathematical & Statistical Psychology*, *73*, 452–473. https://doi.org/10.1111/bmsp.12196

Nestler, S. (2021). Modeling intraindividual variability in growth with measurement burst designs. *Structural Equation Modeling*, *28*(1), 28–39. https://doi.org/10.1080/10705511.2020.1757455

OECD. (2013a). *PISA 2012 assessment and analytical framework: mathematics, reading, science, problem solving and financial literacy*. OECD Publishing. https://doi.org/10.1787/9789264190511

OECD. (2013b). *PISA 2012 dataset* [Data file and code book]. https://www.oecd.org/pisa/data/pisa2012database-downloadabledata.htm

OECD. (2013c). *PISA 2012 technical report*. OECD Publishing. https://www.oecd.org/pisa/pisaproducts/PISA-2012-technical-report-final.pdf

Peers, H. W. (1968). Confidence properties of Bayesian interval estimates. *Journal of the Royal Statistical Society. Series B. Methodological*, *30*(3), 535–544. https://doi.org/10.1111/j.2517-6161.1968.tb00752.x

Petersen, L. E., Dietz, J., & Frey, D. (2004). The effects of intragroup interaction and cohesion on intergroup bias. *Group Processes & Intergroup Relations*, *7*(2), 107–118. https://doi.org/10.1177/1368430204041397

Plummer, M. (2016). rjags: Bayesian graphical models using MCMC (R package version 4-6) [Computer software]. https://CRAN.R-project.org/package=rjags

Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2004). Generalized multilevel structural equation modeling. *Psychometrika*, *69*(2), 167–190. https://doi.org/10.1007/BF02295939

Rabe-Hesketh, S., Skrondal, A., & Zheng, X. (2012). Multilevel structural equation modeling. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 512–531). Guilford Press.

Radler, B. T. (2014). The midlife in the United States (MIDUS) series: A national longitudinal study of health and well-being. *Open Health Data*, *2*(1), e3. https://doi.org/10.5334/ohd.ai

Ram, N., Gatzke-Kopp, L., Gerstorf, D., & Coccia, M. (2015). Intraindividual variability across the life span: Moving toward computational developmental science. In M. Diehl, K. Hooker, & M. Sliwinski (Eds.), *Handbook of intraindividual variability across the lifespan* (pp. 16–34). Routledge/Taylor & Francis Group.

Rast, P., & Ferrer, E. (2018). A mixed-effects location scale model for dyadic interactions. *Multivariate Behavioral Research*, *53*(5), 756–775. https://doi.org/10.31234/osf.io/3czur

Rast, P., Hofer, S. M., & Sparks, C. (2012). Modeling individual differences in within-person variation of negative and positive affect in a mixed effects location scale model using BUGS/JAGS. *Multivariate Behavioral Research*, *47*(2), 177–200. https://doi.org/10.1080/00273171.2012.658328

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

Rosseel, Y. (2017). *Multilevel structural equation modeling with lavaan* [pdf slides]. http://users.ugent.be/~yrosseel/lavaan/zurich2017/MULTILEVEL/lavaan_multilevel_zurich2017.pdf

RStudio Team. (2020). *RStudio: Integrated development for R*. RStudio, PBC. http://www.rstudio.com/

Ryff, C. D., & Almeida, D. M. (2004–2009). *Midlife in the United States (MIDUS 2): Daily Stress Project* [Data set and code book]. Inter-university Consortium for Political and Social Research. https://doi.org/10.3886/ICPSR26841.v2

Ryff, C. D., Almeida, D. M., Ayanian, J., Binkley, N., Carr, D. S. C., Davidson, R. C., Grzywacz, J., Karlamangla, A., Krueger, R., Lachman, M., Love, G., Mailick, M., Mroczek, D., Radler, B., Seeman, T., Sloan, R., Thomas, D., Weinstein, M., & Williams, D. (2013–2014). *Midlife in the United States (MIDUS 3)* [Data set and code book]. Inter-university Consortium for Political and Social Research. https://doi.org/10.3886/ICPSR36346.v7

Ryu, E. (2014). Model fit evaluation in multilevel structural equation models. *Frontiers in Psychology*, *5*, 81. https://doi.org/10.3389/fpsyg.2014.00081

Ryu, E., & West, S. G. (2009). Level-specific evaluation of model fit in multilevel structural equation modeling. *Structural Equation Modeling*, *16*(4), 583–601. https://doi.org/10.1080/10705510903203466

Schultzberg, M., & Muthén, B. (2018). Number of subjects and time points needed for multilevel time-series analysis: A simulation study of dynamic structural equation modeling. *Structural Equation Modeling*, *25*(4), 495–515. https://doi.org/10.1080/10705511.2017.1392862

Severini, T. A. (1991). On the relationship between Bayesian and non-Bayesian interval estimates. *Journal of the Royal Statistical Society. Series B. Methodological*, *53*, 611–618. https://doi.org/10.1111/j.2517-6161.1991.tb01849.x

Smid, S. C., McNeish, D., Miočević, M., & van de Schoot, R. (2019). Bayesian versus frequentist estimation for structural equation models in small sample contexts: A systematic review. *Structural Equation Modeling*, *27*(1), 131–161. https://doi.org/10.1080/10705511.2019.1577140

Snijders, T., & Bosker, R. (2011). *Multilevel analysis: An introduction to basic and advanced multilevel modeling* (2nd ed.). SAGE Publications, Inc.

Stapleton, L. M., Yang, J. S., & Hancock, G. R. (2016). Construct meaning in multilevel settings. *Journal of Educational and Behavioral Statistics*, *41*(5), 481–520. https://doi.org/10.3102/1076998616646200

Tanner, M. A., & Wong, W. H. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, *82*(398), 528–540. https://doi.org/10.1080/01621459.1987.10478458

Tanner, M. A., & Wong, W. H. (2010). From EM to data augmentation: The emergence of MCMC Bayesian computation in the 1980s. *Statistical Science*, *25*(4), 506–516. https://doi.org/10.1214/10-STS341

van Dijk, M., & van Geert, P. (2015). The nature and meaning of intra-individual variability in development in the early life span. In M. Diehl, K. Hooker, & M. J. Sliwinski (Eds.), *Handbook of intraindividual variability across the life span* (pp. 37–58). Routledge/Taylor & Francis Group.

van Knippenberg, D., De Dreu, C. K., & Homan, A. C. (2004). Work group diversity and group performance: An integrative model and research agenda. *Journal of Applied Psychology*, *89*(6), 1008–1022. https://doi.org/10.1037/0021-9010.89.6.1008

Wang, L. P., Hamaker, E., & Bergeman, C. S. (2012). Investigating inter-individual differences in short-term intra-individual variability. *Psychological Methods*, *17*(4), 567–581. https://doi.org/10.1037/a0029317

Weinstein, M., Ryff, C. D., & Seeman, T. E. (2012). 2016). *Midlife in the United States (MIDUS Refresher): Biomarker project, 2012-2016* [Data set and code book]. Inter-university Consortium for Political and Social Research. https://doi.org/10.3886/ICPSR36901.v6

Williams, D. R., Mulder, J., Rouder, J. N., & Rast, P. (2021). Beneath the surface: Unearthing within-person variability and mean relations with Bayesian mixed models. *Psychological Methods*, *26*(1), 74–89. https://doi.org/10.31234/osf.io/gwatq

Zitzmann, S., Helm, C., & Hecht, M. (2020). Prior specification for more stable Bayesian estimation of multilevel latent variable models in small samples: A comparative investigation of two different approaches. *Frontiers in Psychology*. Advance online publication. https://doi.org/10.3389/fpsyg.2020.611267

(*Appendix follows*)

Appendix

Moment Generating Function

The moment generating function (MGF) for random variable X that follows a normal distribution $N(\mu, \sigma^2)$ is:

$$M_X(t) = \exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right).$$

It is also known that for the random variable X with MGF $M_X(t)$, we have:

$$E(X^n) = M_X^{(n)}(0),$$

where $M_X^{(n)}(0)$ is the n-th derivative of $M_X(t)$ with respect to t, evaluated at $t = 0$:

$$M_X^{(n)}(t) = \frac{d^n}{dt^n} M_X(t)\Big|_{t=0}.$$

Therefore, we thus can compute the expectation for X^4 as:

$$\begin{aligned} E(X^4) &= M_X^{(4)}(0) \\ &= \frac{d^4}{dt^4}\exp(\mu t + \frac{1}{2}\sigma^2 t^2)\,|_{t=0} \\ &= \sigma^8 0^4 + 4\mu\sigma^6 0^3 + 6\sigma^4(\sigma^2 + \mu^2)0^2 + 4\mu\sigma^2(3\sigma^2 + \mu^2)0 \\ &\quad + 3\sigma^4 + 6\mu^2\sigma^2 + \mu^4 = 3\sigma^4 + 6\mu^2\sigma^2 + \mu^4 \end{aligned}$$

Received June 1, 2021
Revision received October 6, 2021
Accepted December 5, 2021 ■