Tutorial for Bayesian Multilevel Structural Equation Modeling using Blimp

Jiwon Kim

Department of Medical Social Sciences

Northwestern University

jiwon.kim@northwestern.edu

Sooyong Lee

Wisconsin Center for Educational Research

The University of Wisconsin-Madison

sooyongl09@gmail.com

Suyoung Kim

Crown Family School of Social Work, Policy, and Practice

University of Chicago

suyoungkim@uchicago.edu

Brian T. Keller

Department of Educational, School, and Counseling Psychology

The University of Missouri

btkeller@missouri.edu

Abstract

This tutorial paper introduces the implementation of Multilevel Structural Equation Modeling (MSEM) using Blimp, a free, user-friendly, and flexible software for Bayesian estimation. Aimed at readers with some familiarity with MSEM and Bayesian methods, the tutorial walks through the specifications of a basic multilevel model in Blimp, progressively incorporating latent variables, random slopes, contextual effects, and moderated mediation effects. For illustrative purposes, we use a simulated multilevel dataset to examine the association between teacher-student relationships, students' interest in math, and students' math achievement.
keywords: Multilevel structural equation modeling, Bayesian estimation, Blimp

Introduction

This tutorial provides practical guidance on conducting Bayesian estimation for Multilevel Structural Equation Modeling (MSEM) using Blimp, a free and accessible software. MSEM integrates the flexibility of Structural Equation Modeling with multilevel modeling to account for nested data structures—such as students within classrooms (Preacher, Zyphur, & Zhang, 2010). This approach enables researchers to examine latent relationships at both individual and group levels. However, estimation using Maximum Likelihood Estimation (MLE) can be challenging, especially when models include random slopes. Such models often require intensive computation, leading to slow estimation, convergence issues, or simplified model structures that may compromise the accuracy and richness of their analyses (Asparouhov & Muthén, 2010).

Bayesian estimation offers a promising alternative. It addresses many of the limitations of MLE by better accommodating model complexity, relaxing distributional assumptions, and providing more stable estimates in small samples (Muthén & Asparouhov, 2012). While statistical software packages like Stan (Stan Development Team, 2020) and WinBUGS (Lunn et al., 2000) support Bayesian estimation, they often demand considerable expertise in both Bayesian statistics and programming. The *blavaan* package in R offers more accessibility by using syntax similar to *lavaan* (Merkle & Rosseel, 2018), but its current functionality for MSEM is limited to random intercept models (Merkle, n.d.). Mplus (Muthén & Muthén, 1998–2021) integrates Bayesian features with a more user-friendly interface, but it comes with a cost and certain limitations in flexibility.

Blimp (Keller & Enders, 2023a) bridges this gap by offering full Bayesian estimation for MSEM with a user-friendly interface and support for complex models. Notably, Blimp supports categorical predictors and mediators in moderation analysis, including multilevel moderated mediation and MNLFA (Moderated Nonlinear Factor Analysis) - features not available in Mplus. Despite its capabilities, Blimp remains underutilized among applied researchers. This tutorial introduces Blimp through an illustrative example—a 1-1-1 mediation model with two random slopes exploring relationships among teacher-student relationships (REL), student interest in math (INT), and math achievement (ACH). Detailed code examples, output, and the Mplus syntax corresponding to the Blimp codes are provided in the supplementary materials (OSF: https://osf.io/frzvu/?view_only=f73070c6e1954f19933c2cae30a8393b), allowing readers to replicate our demonstrations and adapt them to their research. For a more comprehensive overview of Bayesian estimation and MSEM, see Gelman et al. (2013), Wang and Wang (2019), and Muthén and Asparouhov (2012).

Multilevel Structural Equation Modeling

Multilevel Structural Equation Modeling (MSEM) integrates Multilevel Modeling (MLM) and Structural Equation Modeling (SEM) to account for hierarchical data structures and latent variables simultaneously. This framework is specifically designed to address challenges inherent in hierarchical data structures and latent variable measurement, making it particularly valuable for educational, psychological, and social science research.

MSEM is highly flexible and can handle many different research questions for multilevel structures. For instance, consider a study examining how teacher-student relationships (`REL`) influence students' interest in math (`INT`) and their math achievement (`ACH`). At the student level (i.e., within-school, level 1), we might hypothesize a mediation model wherein teacher-student relationships affect students' interest, which then impacts their math achievement. At the school level (i.e., between-school, level 2), similar associations could emerge: schools with stronger overall teacher-student relationships might report higher levels of student math interest and improved math performance. MSEM allows researchers to model such multilevel mediation processes simultaneously at both levels, accounting for the nested data structure.

In the following section, we begin by reviewing how MSEM accounts for both sampling error and measurement error. Next, we discuss incorporating random slopes and modeling multilevel mediation. Finally, we introduce contextual effects and explain how to address cross-level moderation and moderated mediation.

Addressing Sampling Error

The primary strengths of MSEM lie in its ability to handle sampling error in nested data structures. MSEM decomposes the variance and covariance structures into within-group and between-group components. This approach ensures more precise modeling of relationships across different levels. Often referred to as a multilevel path model (Kaplan, 1998), the framework explicitly accounts for the hierarchical nature of the data, allowing researchers to disentangle and model structures at both the individual and group levels effectively.

The decomposition of observed variables into between-group and within-group components for clustered data to handle sampling error is mathematically represented as:

$$y_{ig} = y^B + y^W = \alpha + u_g + e_{ig} \qquad (1)$$

where y_{ig} represents the observed variable, y^B is the between-group component, y^W is the within-group component, α is the grand mean, u_g is the group-level residual, and e_{ig} is the individual-level residual. This decomposition allows for the separate modeling of covariance structures at both between- and within-group levels:

$$\Sigma^T = \Sigma^B + \Sigma^W \qquad (2)$$

where Σ^T is the total covariance matrix, Σ^B is the between-level covariance matrix, and Σ^W is the within-level covariance matrix. In doing so, MSEM separates the variance components attributable to individual differences within groups (within-group variance; individual student differences) and differences between groups (between-group variance; school differences).

When applied to the example scenario, the equation is set up like a mediation model at each level[1] with composite values to address the sampling errors. Consider a path model with three observed variables (i.e., `ACH`, `INT`, and `REL`) measured at the within-school level (Level 1), expressed as:

$$ACH^w = \beta_b^w INT^w + \beta_c^w REL^w + e_{ACH^w} \qquad (3)$$

$$INT^w = \gamma_a^w REL^w + e_{INT^w}$$

where e_{ACH^w} and e_{INT^w} represent the residuals at the student levels, and β_b^w, β_c^w, and γ_a^w are the regression coefficients among the three variables at the student levels. At the between-school level (Level 2), the model can be expressed as:

$$ACH^b = \nu_{ACH} + \beta_b^b INT^b + \beta_c^b REL^b + e_{ACH^b} \qquad (4)$$

$$INT^b = \nu_{INT} + \gamma_a^b REL^b + e_{INT^b}$$

where ν_{ACH} is the intercept and e_{ACH^b} and e_{INT^b} represent the residuals at the school level.

[1] In particular, it is the so-called 1-1-1 mediation model, which will be discussed later.

β_b^b, β_c^b, and γ_a^b are the regression coefficients among the three variables at the school levels. This framework allows researchers to model the structural relationships among the variables at the within-school level while simultaneously accounting for the relationships among the variables at the between-school.

It is important to note that centering plays a crucial role in multilevel modeling because it clarifies the interpretations of the model coefficients (Enders & Tofighi, 2007). At the within level, group-mean centering is used to isolate individual-level effects by removing between-group differences. This ensures that within-level coefficients reflect relationships among individuals within the same group—without interference from higher-level variation. At the between-level, grand-mean centering is typically applied to interpret effects relative to the overall sample mean. This allows between-level coefficients to show how group-level predictors influence outcomes when other variables are at their average values.

Addressing Measurement Error

Beyond sampling error, MSEM also addresses measurement error by incorporating latent variables. Psychological concepts like math interest are considered hypothetical constructs that cannot be directly measured (i.e., latent or unobservable; Wang & Wang, 2019). MSEM employs measurement models with multiple observed indicators to accurately represent latent constructs and account for measurement error.

Building on the earlier example, the multilevel path model can be extended to include latent variables:

$$\eta_{ACH}^w = \beta_b^w \eta_{INT}^w + \beta_c^w \eta_{REL}^w + e_{\eta_{ACH}^w} \quad (5)$$

$$\eta_{INT}^w = \gamma_a^w \eta_{REL}^w + e_{\eta_{INT}^w}$$

$$\eta_{ACH}^b = \beta_b^b \eta_{INT}^b + \beta_c^b \eta_{REL}^b + e_{\eta_{ACH}^b} \quad (6)$$

$$\eta_{INT}^b = \gamma_a^b \eta_{REL}^b + e_{\eta_{INT}^b}$$

where η^w and η^b indicate the underlying latent constructs for the within-school and between-school levels. The terms e_w and e_b are residuals for endogenous latent constructs (η_{ACH} , η_{INT}) at the student level and school level respectively, which are assumed to follow a normal distribution with a mean of zero and variance. Similarly, the exogenous latent construct, η_{REL} , also follows a normal distribution with a mean of zero and an estimated variance. The regression coefficients maintain the same interpretations as specified in Equations (3) and (4).

The latent constructs η^w and η^b are linked to the observed indicators as follows:

$$ACH_{ig} = {\nu_{ACH_g}}^b + {\Lambda^w}_{ACH}\eta^w_{ACHig} + {\Lambda^b}_{ACH}\eta^b_{ACHg} + \epsilon^w_{ACHig} + \epsilon^b_{ACHg} \quad (7)$$

$$INT_{ig} = {\nu_{INT_g}}^b + {\Lambda^w}_{INT}\eta^w_{INTig} + {\Lambda^b}_{INT}\eta^b_{INTg} + \epsilon^w_{INTig} + \epsilon^b_{INTg}$$

$$REL_{ig} = {\nu_{REL_g}}^b + {\Lambda^w}_{REL}\eta^w_{RELig} + {\Lambda^b}_{REL}\eta^b_{RELg} + \epsilon^w_{RELig} + \epsilon^b_{RELg}$$

where *ig* refers to individual *i* within group *g*. The term ν^b represents the intercepts of indicators at the between-schools, Λ^w_{ig} and Λ^b_g are the factor loadings for the within-school and between-school latent constructs, and ϵ^w_{ig} and ϵ^b_g are within-school and between-school residuals, following normal distribution with the mean of zero and the variance of θ.

Random slopes

In MSEM, the inclusion of random slopes adds a layer of complexity that requires careful consideration. Random slopes allow for the modeling of individual differences in the relationships between variables across different levels of analysis (Kaplan, 1998). By capturing variability in within-level effects, random slopes provide insights into how the strength or direction of relationships between variables varies across groups. For example, the impact of teacher-student relationships on math interest (i.e., γ^w_a) or the influence of math interest on math achievement (i.e., β^w_b) at the student level may differ significantly between schools.

A key conceptual point is that random slopes inherently represent interactions between a Level 1 predictor and the grouping structure (Enders et al., 2020). Consider $\eta_{INT}^{w} = \gamma_a^{w}\eta_{REL}^{w} + e_{\eta_{INT}^{w}}$ in Equation (5). If γ_a^{w} is modeled as a random effect, then:

$$\gamma_{ag}^{w} = \gamma_{00} + \epsilon_{\gamma g} \tag{8}$$

where $\epsilon_{\gamma g}$ represents the deviation of the individual estimate from the grand mean. Substituting into the original regression equation yields:

$$\eta_{INT}^{w} = (\gamma_{00} + \epsilon_{\gamma g})\eta_{REL}^{w} + e_{\eta_{INT}^{w}} \tag{9}$$

$$\eta_{INT}^{w} = \gamma_{00}\eta_{REL}^{w} + \epsilon_{\gamma g}\eta_{REL}^{w} + e_{\eta_{INT}^{w}}$$

Here, the term $\epsilon_{\gamma g}\eta_{REL}^{w}$ reveals the essence of a random slope—it represents a group-specific interaction effect, where the strength of the relationship between η_{REL}^{w} on η_{INT}^{w} depends on the school context (*g*). In cases where the predictor is a latent construct (e.g., η_{REL}^{w}), the random slope becomes a latent-by-latent interaction; with observed predictors, it becomes a latent-by-manifest interaction (Keller, 2022). Either way, the implication is the same: the strength of an effect depends on where (or in which group) it occurs. In other words, the slopes of η_{REL}^{w} is moderated by the group context, much like how interaction terms in a traditional regression model indicate conditional effects.

When multiple random effects are included (e.g., both γ_a^{w} and β_a^{w} in Equation 5), the formulation for random slopes is:

$$\beta_{bg}^{w} = \beta_{00} + \epsilon_{\beta g} \tag{8}$$

$$\gamma_{ag}^{w} = \gamma_{00} + \epsilon_{\gamma g}$$

$$T = \begin{pmatrix} \Theta_{Bg} & \theta_{\beta_b\gamma_a} & \Theta_{\Gamma g} \end{pmatrix} \tag{9}$$

where β_{00} and γ_{00} represent the fixed slope of β_{bg}^{w} and γ_{ag}^{w}, respectively, and $\epsilon_{\beta g}$ and $\epsilon_{\gamma g}$ are disturbance terms assumed to be normally distributed with mean zero and covariance matrix

T. A positive covariance $\theta_{\beta_b \gamma_a}$ means that higher-level units with larger γ_a values are likely to also have larger β_b values. On the other hand, a negative $\theta_{\beta_b \gamma_a}$ implies that units with higher γ_a values are more likely to have smaller β_b values (Kenny, Korchmaros, & Bolger, 2003).

Multilevel Mediation Effects

Mediation analysis is a powerful tool for uncovering the mechanisms between the predictor, mediator, and outcome variables. MSEM extends this capability to hierarchical data structures frequently encountered in psychology and education research (Bauer, Preacher, & Gil, 2006). Mediation effects can be explored at both within- and between-group levels. In our example, mediation analysis can reveal how math teacher relationships (`REL`) influence math achievement (`ACH`) through math interest (`INT`) at both student and school levels. The a-path (i.e., γ_a^w and γ_a^b) captures the effect of the predictor (`REL`) on the mediator (`INT`), and the b-path (i.e., β_b^w and β_b^b) represents the effect of the mediator (`INT`) on the outcome (`ACH`). Their product (*ab*) quantifies the indirect effect, while the direct effect of `REL` on `ACH` is represented by the c'-path. The total effect is the sum of the direct and indirect effects, expressed as c = ab + c' (Bollen, 1987).[2]

Mediation models in MSEM can range in complexity depending on the level at which each variable is measured, with structures like 1-1-1, 2-1-1, or 2-2-1 (Preacher et al., 2010). The present study focuses on a 1-1-1 mediation model, where the mediation process operates at both within-level (Level 1) and between-level (Level 2). The model examines how a Level 1 predictor (`REL`) affects a mediator (`INT`) at the individual level and how the mediator subsequently influences an outcome (`ACH`). This is represented by the a-path (effect of `REL` on `INT`) and the b-path (effect of `INT` on `ACH`). This model enables the analysis of

[2] When the mediator or outcome variable is binary or polytomous, the interpretation of the product of a and b as the indirect effect may not hold. There is significant discussion around this issue. For further exploration of non-continuous mediators or outcomes, see Muthén & Asparouhov (2015). The scope of this study, however, is limited to continuous variables.

relationships at the individual level while simultaneously modeling relationships at the group level, allowing for a comprehensive examination of mediation effects at both levels within a single model.

One important aspect of multilevel mediation in MSEM is that when the mediation effects a and b are treated as random variables and are correlated, the expected value of their product $a \times b$ does not equal the product of their individual expected values, as is typically assumed in single-level mediation models (Kenny et al., 2003). In the context of multilevel models, the expected value of $E(\gamma_{ag}^{w}\beta_{bg}^{w})$ is expressed as $E(\gamma_{bg}^{w})E(\beta_{bg}^{w})+\theta_{\beta_b\gamma_a}$ (Goodman, 1960), where $\theta_{\beta_b\gamma_a}$ represents the covariance between γ_{ag}^{w} and β_{bg}^{w}, accounting for the potential correlation between these effects.

Contextual Effects

Contextual effects are essential for understanding how group-level characteristics influence outcomes beyond individual-level effects. Without accounting for contextual effects, researchers risk conflating individual and group influences, leading to misinterpretation of relationships between variables in hierarchical data (Enders & Tofighi, 2007). For instance, the relationship between a student's interest in math and their achievement may differ from the relationship between their average level of math interest in their school and student achievement. Contextual effects help to disentangle these influences, providing deeper insights into how individual and group dynamics interact.

In multilevel models, contextual effects are calculated by isolating group-level influences while accounting for individual-level variability. While between-school and within-school effects are separately defined, they can be confounded because within-group relationships often mirror group-level patterns. To isolate the contextual effect—also known as the compositional effect—the within-school effect is subtracted from the between-school effect (Enders & Tofighi, 2007; Raudenbush, 2002). For example, the contextual effect for the b-path parameter is calculated as:

$$\beta_b^{contextual} = \beta_b^b - \beta_b^w \tag{10}$$

where ${\beta_b}^b$ is the between-school effect of math interest on achievement, and ${\beta_b}^w$ is the within-school effect. This contextual effect represents the expected difference in math achievement between two students with the same level of math interest, but who attend schools differing by one unit in the average math interest. It is important to note that in the presence of random slopes, Equation 10 holds only in expectation (Laird & Ware, 1982).

Cross-level moderation and moderated mediation

In multilevel mediation models with random slopes, cross-level moderation effects can be explored by incorporating Level 2 covariates. These models treat the within-cluster a-path and b-path as random, and their variability is explained by a Level 2 covariate, such as socioeconomic status (SES). This relationship can be formulated as:

$$\beta_{bg}^w = \beta_{b0} + \beta_{b1} {SES^b}_g + \epsilon_{\beta_{bg}} \tag{11}$$

$$\gamma_{ag}^w = \gamma_{a0} + \gamma_{a1} {SES^b}_g + \epsilon_{\gamma_{ag}}$$

where ${SES^b}_g$ is a level 2 variable. β_{b1} and γ_{a1} capture cross-level moderation effects — indicating how the strength of the within-level relationship varies as a function of SES.

The indirect effect with a cross-level covariate is defined as:

$$E({\gamma_a}^w {\beta_b}^w | SES_g = ses) = (\beta_{b0} + \beta_{b1} ses)(\gamma_{a0} + \gamma_{a1} ses) + \sigma_{ab} \tag{12}$$

, where σ_{ab} is the covariance between the two random slopes. When moderated mediation is present, researchers can evaluate simple indirect effects at different levels of the moderator (Bauer et al., 2006). For example, when SES equals its mean (SES = 0), the indirect effect simplifies to $\beta_{b0}\gamma_{a0} + \sigma_{ab}$. Conditional effects at specific levels of SES (e.g., ±1 SD from the grand mean) can be evaluated by computing the expected value of the indirect effect at those SES values, as presented in Equation (13). Importantly, this standard deviation should reflect the appropriate level of analysis: Level 2 SD for between-level effects, and Level 1 SD for within-level effects.

$$E(\gamma_a{}^w \beta_b{}^w | w_1 = ses + SD) = (\beta_{b0} + \beta_{b1} w_1)(\gamma_{a0} + \gamma_{a1} w_1) + \sigma_{ab} \quad (13)$$

$$E(\gamma_a{}^w \beta_b{}^w | w_2 = ses - SD) = (\beta_{b0} + \beta_{b1} w_2)(\gamma_{a0} + \gamma_{a1} w_2) + \sigma_{ab}$$

Incorporating Level 2 covariates enables the assessment of moderated mediation effects across different levels of the moderator. This approach helps clarify how the mediation pathways (the a-path and b-path) vary as a function of `SES`. It provides deeper insights into the dynamics of the mediation process at both the within-cluster and between-cluster levels, revealing how individual and group-level factors interact to influence outcomes.

Bayesian Inference

As discussed so far, MSEMs are well-established within the frequentist framework (i.e., ML). However, as model complexity grows—whether due to multiple random slopes, numerous latent factors, or high-dimensional structures—ML methods often struggle. Convergence issues and excessive computation times become common, especially with small sample sizes or complex model specifications. Bayesian estimation offers a compelling alternative. Instead of relying on direct numerical integration, Bayesian methods utilize Markov Chain Monte Carlo (MCMC) algorithms (e.g., Gibbs sampling) to estimate model parameters. This iterative approach samples from conditional distributions to construct full posterior distributions, capturing the range of plausible parameter values based on the data and prior assumptions (Gelman et al., 2013). Statistical software programs like Blimp and M*plus* implement these algorithms efficiently.

This pragmatic use of Bayesian estimation—primarily for overcoming computational hurdles—has been described as "Bayes as Computational Frequentism" (Levy & McNeish, 2023).[3] In this approach, Bayesian tools are leveraged to obtain frequentist-like point estimates and uncertainty measures when ML estimation becomes impractical. The inferential framework remains fundamentally frequentist; the Bayesian machinery serves primarily as a computational workaround.

In this context, posterior means or medians serve as analogs to ML estimates, and posterior standard deviations act as Bayesian counterparts to standard errors (Kruschke & Liddell, 2018; McElreath, 2020). Likewise, credible intervals are often interpreted similarly to confidence intervals—although they differ philosophically, their presentation and use can align in practical terms. To minimize the influence of prior assumptions, researchers using this approach typically rely on diffuse or weakly informative priors, especially in large samples where the data tend to dominate the priors.

While diffuse priors are common in this computational framework, one of Bayesian estimation's true strengths lies in its ability to incorporate informative priors. This shifts the focus from computational convenience to Bayes as Updating or Augmenting the Likelihood. Informative priors, derived from prior research or theoretical reasoning, help stabilize estimation—particularly in small-sample scenarios—and can improve estimate precision (McNeish, 2016). As sample size increases, the impact of priors fades, and Bayesian estimates converge toward their frequentist equivalents—a phenomenon known as asymptotic equivalence (Gelman et al., 2013).

Another advantage of Bayesian methods is their flexibility in distributional assumptions, which is especially useful in contexts like mediation analysis, where indirect effects (products of coefficients) rarely follow a normal distribution (Yuan & MacKinnon, 2009). Bayesian estimation naturally accommodates these non-normalities, yielding more

[3] Bayesian analyses can be conceptualized as computational frequentism, likelihood augmentation, updating, joint distribution analysis, multilevel model building. See Levy & McNeish (2023) for further discussion.

accurate credible intervals for indirect effects—even in complex or nested models.

From a philosophical standpoint, Bayesian offers a more intuitive interpretation of uncertainty by treating parameters as random variables with posterior distributions (Etz & Vandekerckhove, 2018; McElreath, 2020). Bayesian approaches also offer practical tools for assessing convergence, such as the Gelman-Rubin diagnostic, often referred to as Potential Scale Reduction (PSR) or $\hat{R}$, which assesses if multiple MCMC chains have mixed well (Gelman & Rubin, 1992). An $\hat{R}$ value below 1.05 or 1.10 typically indicates adequate convergence, whereas higher values suggest the need for additional iterations or model adjustments (Brooks & Gelman, 1998; Gelman et al., 2013).

In summary, Bayesian estimation offers numerous advantages for MSEMs, including improved convergence for complex models, flexibility in addressing distributional assumptions, and straightforward estimation of mediation effects in the multilevel frameworks. Whether used for computational pragmatism or full Bayesian reasoning, this approach enables richer and more robust insights into multilevel data structures—particularly when traditional estimation methods fall short.

Application of MSEM with Blimp

This section illustrates the application of MSEM using Blimp, based on the hypothesized mediation example previously discussed. The analysis examines relationships among three latent factors—math achievement (`ACH`), math interest (`INT`), and teacher-student relationships (`REL`)—as well as one covariate, socioeconomic status (`SES`). Each latent factor is measured by four continuous indicators. The dataset used for this analysis, which is generated by the authors specifically for illustrative purposes, consists of 100 schools (clusters), each with 30 students.

In this hypothesized model, the within-school associations between `REL` and `INT`, and between `INT` and `ACH`, are assumed to vary across schools. These associations are modeled as random slopes within the MSEM framework, enabling the estimation of within-level mediation effects that differ between schools. To account for the variability in these effects, `SES` is introduced as a cross-level moderator. While `SES` is measured at the within-school level, it is applied at both the within- and between-school levels. Specifically, at the between-school level, `SES` moderates the within-school level indirect effect of teacher-student relationships (`REL`) on math achievement (`ACH`) through math interest (`INT`). Note that the data are simulated and observational, and any causal interpretations are intended for illustrative purposes. We include them to demonstrate typical approaches to interpreting the results of the mediation models, but caution is needed when drawing causal conclusions in non-experimental contexts.

The following section provides a step-by-step guide through four models, each accompanied by the corresponding Blimp syntax to illustrate the application of MSEM. All relevant syntax, outputs, data, and corresponding *Mplus* syntax for the Blimp analyses are available at https://osf.io/frzvu/?view_only=f73070c6e1954f19933c2cae30a8393b.

- Model 1: Introduces a basic multilevel path model.
- Model 2: Expands on Model 1 by incorporating a measurement model, allowing for latent constructs.
- Model 3: Builds on Model 2 by adding random slopes, enabling the testing of both between- and within-school indirect effects as well as contextual effects.
- Model 4: Further extends Model 3 by introducing a Level 2 covariate as a cross-level moderator, facilitating the examination of moderated mediation effects.

Basic Settings

Before diving into multilevel modeling, it is essential to understand the basic syntax for Blimp. Blimp uses # symbol for single-line comments. The analysis begins by specifying the file directory using the `DATA` command, followed by listing the relevant variables using the VARIABLE command, where four indicators are listed for each construct (ACH, INT, REL), one covariate (SES), cluster id variable (schid), and student id variable (id). Blimp requires users to specify the names of any latent variables involved in the model. In this example, ACH_w represents the latent variables at Level 1. For Level 2 latent variable, the cluster id variable must be specified as schid = ACH_b.

The Level 2 identifier is specified using the CLUSTERID command. Missing data is specified through the MISSING command. It may also be noted that Blimp uses a seed for reproducibility (SEED)—a positive integer that allows users to replicate the results of the pseudo-random number generator. For Bayesian estimation, the BURN and ITERATIONS commands are used to specify the number of burn-in samples and total number of iterations, respectively. The CHAIN command determines the number of MCMC chains used. The total number of iterations is calculated as the number of iterations per chain multiplied by the number of chains. Unless otherwise specified, the default prior distributions in Blimp are used for Bayesian estimation (see the Blimp User's Guide (Keller & Enders, 2023b) for details). Once these basic settings are defined, Blimp is ready to perform Bayesian multilevel analysis without requiring additional configuration.

Blimp Syntax

```
DATA: data.csv;                         #Read in data
VARIABLES: ACH1:ACH4 INT1:INT4 REL1:REL4 SES schid id;  #List variables
LATENT: ACH_w schid = ACH_b;            #Specify latent variables
CLUSTERID: schid;                       #Specify between-level identifier
MISSING: NA;                            #Missing Data
SEED: 46;                               #Set a seed
BURN: 10000;       #Specify a number of burn-in iterations
ITERATIONS: 10000;#Specify number of iterations after burn-in period
CHAIN: 4;          #Specify number of chains
```

Model 1. Multilevel path model

Before moving onto the main model (MSEM), we first elaborate on the multilevel path model, where only sampling errors are addressed. As illustrated in Figure 1, this is a 1-1-

1 multilevel mediation model, where mediation occurs across both the within-level (Level 1) and the between-level (Level 2). This model explores how the predictor (REL) influences the mediator (INT), and in turn, how the mediator impacts the outcome (ACH). These pathways, known as the a-path (REL to INT) and b-path (INT to ACH), are designed to capture variability both across individuals and between clusters.

In the following Blimp syntax, the LATENT command is used to define the three between-level (Level 2) variables: ACH_b, INT_b, and REL_b. Then, in the MODEL command, group-mean centering at the within-level (Level 1) is used to compute the within-level deviations by subtracting the between-level means. For instance, within-level math achievement is computed as ACH_w = ACH - ACH_b, representing individual deviation of math achievement from its group mean (ACH_b). At the between-level (Level 2), Blimp uses grand-mean centering to capture the between-cluster deviations of each variable. For instance, the syntax ACH_bc = ACH_b - gm_ACH captures the between-level deviation from the grand mean (gm_ACH). Here, gm_ACH is the label assigned to the intercept of ACH_b. Because all predictors are group-mean centered, gm_ACH reflects the grand mean of ACH by definition.

The mediation pathways at the within-level are specified by regressing the math achievement (ACH_w) on the within-level deviations of the mediator (math interest; INT_w) and the predictor (math teacher relations; REL_w) using the syntax ACH ~ INT_w REL_w and INT ~ REL_w. These correspond to the mediation paths in the within-cluster level box in Figure 1: path a_w (from REL_w to INT_w), path b_w (from INT_w to ACH_w), and path c_w (from REL_w to ACH_w). Likewise, the mediation pathways at the between-level are specified by regressing the group means of the outcome and mediator on the between-level deviations of the predictor and mediator, using the syntax ACH_b ~ INT_bc REL_bc and INT_b ~ REL_bc. These paths correspond to a_b, b_b, c_b and in the between-cluster box in Figure 1.

Blimp allows users to organize the syntax by level using the optional “blocks” (e.g., *within_l1* and *between_l2*). While this is not required, these blocks help improve the clarity and structure of the model specification. Another noteworthy feature is the @ operator, which labels parameters. For instance, ACH ~ 1@ACH_b indicates that the intercept of ACH is labeled as ACH_b, which is also defined as a latent variable at the between-level. Note that the within-level intercepts are not freely estimated but are fixed to between-level latent variables, which vary across clusters.

Blimp Syntax

```
LATENT:
 #Define between-school latent variables
  schid = ACH_b INT_b REL_b;

 MODEL:
 within_l1:
 #Define group mean-centered deviations
  ACH_w = ACH - ACH_b;
  INT_w = INT - INT_b;
  REL_w = REL - REL_b;

 #Fit within Models
  ACH ~ 1@ACH_b INT_w REL_w;
  INT ~ 1@INT_b REL_w;
  REL ~ 1@REL_b;

 between_l2:
 #Define grand mean-centered deviations
  ACH_bc = ACH_b - gm_ACH;
  INT_bc = INT_b - gm_INT;
  REL_bc = REL_b - gm_REL;

 #Fit between Models
  ACH_b ~ 1@gm_ACH INT_bc REL_bc;
  INT_b ~ 1@gm_INT REL_bc;
  REL_b ~ 1@gm_REL;
```

Note. ACH, INT, and REL represent the average scores of math achievement, math interest, and math teacher relations, respectively, each measured by four indicators (ACH1-ACH4 for math achievement, INT1-INT4 for math interest, and REL1-REL4 for math teacher relations).

[Figure 1 about here]

[Table 1 about here]

Table 1 presents the model estimates, 95% credible intervals, and the $\hat{R}$ values for convergence. Since all $\hat{R}$ values are below the threshold of 1.05, this indicates that the model has successfully achieved satisfactory convergence. First, the within-school paths from `REL` to `INT` and `ACH` are estimated to be 0.279 (95% CI: 0.240, 0.319) and 0.224 (95% CI: 0.184, 0.262), respectively, with the path from `INT` to `ACH` estimated at 0.430 (95% CI: 0.395, 0.465). The credible intervals for all paths do not include zero, indicating that the effects are significantly different from zero. These results suggest that, given this model, for each additional unit increase in `REL` within an individual (relative to others in the school), `INT` and `ACH` are expected to increase by approximately 0.279 and 0.224 units. Furthermore, an increase in `INT` leads to a corresponding increase in `ACH` by 0.430 units.

In the between-school model, the path from `REL` to `INT` and `ACH` is estimated at 0.621 (95% CI: 0.455, 0.783) and 0.277 (95% CI: 0.090, 0.458) and is significantly different from zero. This suggests that for each additional unit increase in the average `REL` at the school level, the school's average achievement is expected to increase by 0.62 and 0.28 units. The between-school effects of `INT` on `ACH` are 0.55 (95% CI: 0.361, 0.737), and statistically significant.

Model 2. Multilevel Structural Equation model

In Model 2, the multilevel path model is extended to incorporate a measurement model, addressing both measurement and sampling errors as shown in Figure 2. Blimp treats latent variables as missing data and estimates them iteratively through data augmentation (Lee & Shi, 2000; Palomo, Dunson, & Bollen, 2007; Merkle & Rosseel, 2018), rather than integrating them out from a joint multivariate distribution with observed variables (Kaplan & Depaoli, 2012; Muthén & Asparouhov, 2012; Merkle, Fitzsimmons, Uanhoro, & Goodrich, 2020). Therefore, the `LATENT` command is used to define and name these unobserved variables. We can then use the `->` to load multiple indicators onto the latent variable. For example, `ACH_w -> ACH1:ACH4` indicates that the latent factor `ACH_w` is measured by the observed variable `ACH1` through `ACH4`. In our syntax, we omit the `ACH1:ACH4|0` constraint because we estimate residual variances at both the within- and between-levels. However, if between-level residual variances are intended to be constrained to zero, the `ACH1:ACH4|0` specification should be explicitly added (i.e., `ACH_w -> ACH1:ACH4|0`). This ensures that all residual variance is attributed to the within-level only.

At the between-level, the intercepts of the indicators are not fixed and freely estimated, as shown in `ACH_b -> ACH1:ACH4`. The zero intercepts for the latent variables do not have to be specified, as they are automatically included by default for model identification (e.g., `ACH_b ~ 1@0`). The same approach is applied to all other latent factors in the model. By using latent factors, we improve the precision of our estimates by explicitly modeling measurement error, which provides a more accurate representation of the relationships among the variables in both the within-level and between-level models, provided the latent variable model accurately reflects the true data generating process and is not misspecified (Rhemtulla, van Bork, & Borsboom, 2020).

Blimp Syntax

```
LATENT:
 ACH_w INT_w REL_w ;           #Latent at the within-level
 schid = ACH_b INT_b REL_b;   #Latent at the between-level

 MODEL:
 Within-level Model:
 ACH_w -> ACH1:ACH4;     #Define latent factors
 INT_w -> INT1:INT4;
 REL_w -> REL1:REL4;

 ACH_w ~ INT_w REL_w;          #Set regressions
 INT_w ~ REL_w;

 Between-level Model:
 ACH_b -> ACH1:ACH4;           #Define latent variables
```

```
INT_b -> INT1:INT4;
REL_b -> REL1:REL4;

ACH_b ~ INT_b REL_b;        #Set regressions
INT_b ~ REL_b;
```

Note. ACH, INT, and REL represent the latent factors of math achievement, math interest, and math teacher relations, respectively, each underlying four indicators (ACH1-ACH4 for math achievement, INT1-INT4 for math interest, and REL1-REL4 for math teacher relations). The suffixes 'b' and 'w' denote the between-school and within-school levels, respectively.

[Figure 2 about here]

[Table 2 about here]

Table 2 presents the model estimates for the mediation model with latent factors. Unlike Model 1, the latent factors are employed at both levels to account for measurement error. The within- and between-level loadings from each latent factor to the corresponding indicators are approximately 1.0, with the first loading excluded as it is fixed to 1.0. Factor loadings can be interpreted in the same way as in the typical Structural Equation Modeling (SEM). They indicate the strength of the association between the observed indicators and the latent factor. When standardized, the square of the standardized factor loadings represents the proportion of variance in the indicator that is explained by the latent factor. In Table 2, the standardized factor loadings are approximately 0.5–0.6, suggesting that 25–36% of the variance in the indicators is explained by the latent factors, both within-school and between-school.

Regarding the regression coefficients, the within-school paths from REL to INT and ACH are estimated to be 0.338 (95% CI: 0.288, 0.389) and 0.245 (95% CI: 0.197, 0.294), respectively, with the path from INT to ACH estimated at 0.475 (95% CI: 0.431, 0.520). In the between-school model, the path from REL to INT and ACH is estimated at 0.601 (95% CI: 0.432, 0.789) and 0.220 (95% CI: 0.035, 0.415) and both paths are significantly different from zero. This suggests that for each additional unit increase in the average REL at the school level, the school's average achievement is expected to increase by 0.22 units and the school's average interest in math increases by 0.60. The between-school effects of `INT` on `ACH` are 0.577 (95% CI: 0.358, 0.819).

Compared to the estimates from Model 1, these coefficients are generally higher. This difference arises because Model 2 accounts for measurement errors by separating the latent constructs from the error components, thereby effectively reducing the noise introduced by imperfect measurements. As a result, the coefficients in Model 2 are more accurate, whereas those in Model 1 are attenuated due to unaddressed measurement error, provided that Model 2 is correctly specified and adequately captures the underlying measurement structure.

Model 3. Model with random slopes to indirect paths

In Model 3 (see Figure 3), random slopes are introduced to account for the variability in within-school effects across schools, allowing for a more accurate estimation of the mediation effect by modeling how these pathways differ across schools. In this model, we also estimate the indirect effects for between- and within-school levels, as well as examine contextual effects.

Random slopes

To estimate random slopes, random slopes are specified in the `LATENT` command (i.e., `RSa`, `RSb`). The Blimp syntax for the random slope model includes several specifications to capture variability in relationships across schools. In the within-level model, the notation `INT_w@RSb` assigns a random slope (`RSb`) to the effect of `INT_w` on `ACH_w`, allowing this effect to vary between schools. Similarly, `REL_w@RSa` attaches the random slope (`RSa)` to the effect of `REL_W` on `INT_w`, allowing this effect to vary across schools. At the between-level, `RSb ~ 1@b_w` sets the intercept of the random slope `RSb` to `b_w`, representing the average effect of `INT_w` on `ACH_w` across schools. Similarly, `RSa ~ 1@a_w` sets the intercept of `RSa` to `a_w`, representing the average effect of `REL_w` on `INT_w` across schools. The bottom three lines of syntax specify the variances and covariances of two random slopes. Specifically, the syntax `RSb ~~ RSb` and `RSa ~~ RSa` specify the variances of two random slopes that capture how much these effects vary across schools. The `RSb ~~ RSa@rCor` sets the "correlation" between two random slopes, labeled as `rCor`. This correlation term reveals whether variability in the effect of `INT_w` on `ACH_w` is related to variability in the effect of `REL_w` on `INT_w` across schools.

Blimp Syntax

```
LATENT:
 ACH_w INT_w REL_w;
 schid = ACH_b INT_b REL_b RSa RSb;

 MODEL:
 Within-level model:
  ACH_w -> ACH1:ACH4;
  INT_w -> INT1:INT4;
  REL_w -> REL1:REL4;

  ACH_w ~ INT_w@RSb REL_w;
  INT_w ~ REL_w@RSa;

 Between-level model:
  ACH_b -> ACH1:ACH4;
  INT_b -> INT1:INT4;
  REL_b -> REL1:REL4;

  ACH_b ~ MI_b@b_b REL_b ;
  INT_b ~ REL_b@a_b ;

  RSb ~ 1@b_w;
  RSa ~ 1@a_w;
  RSa ~~ RSa;
  RSb ~~ RSb;
  RSa ~~ RSb@rCor;
```

Note. `ACH`, `INT`, and `REL` represent the latent factors of math achievement, math interest, and math teacher relations, respectively, each underlying four indicators (`ACH1-ACH4` for math achievement, `INT1-INT4` for math interest, and `REL1-REL4` for math teacher relations). The suffixes 'b' and 'w' denote the between-school and within-school levels, respectively.

[Figure 3 about here]

[Table 3 about here]

Table 3 presents the random slope estimates from Blimp. The means of random slope (`RSa` and `RSb`) indicate the average within-school effects for the a- and b-path at the within-school level. The variances and correlation of `RSa` and `RSb` reflect how much the within-school effects vary across schools and how they relate to each other.

The means of `RSa` and `RSb`, with estimated medians of 0.331 and 0.458 respectively, indicate the average within-school effects for two separate pathways, labeled as a-path and b-path. These estimates are supported by relatively narrow 95% credible intervals, suggesting a reasonable degree of certainty: `RSa` ranges from 0.213 to 0.452, and `RSb` ranges from 0.359 to 0.559.

The variances for `RSa` and `RSb` are 0.309 and 0.212, respectively. The larger variance for `RSa` suggests that the a-path has greater variability in its impact between schools compared to the b-path. This indicates that while `RSb` effects are relatively stable across schools, `RSa` effects fluctuate more significantly, potentially pointing to factors that differentially influence the a-path at the school level.

The correlation between `RSa` and `RSb` is 0.027 (95% CI: -0.033, 0.091). This reflects a slightly positive association implying that when one pathway shows a higher within-school effect, the other tends to have a higher effect as well. However, the strength of this relationship is not particularly strong, suggesting that while there is some linkage between `RSa` and `RSb`, they largely capture distinct aspects of within-school variability. Overall, the table provides insights into how the within-school effects vary across schools and how the pathways relate to each other, emphasizing the heterogeneity in the data at the school level.

Mediation effects

Blimp provides features to compute additional parameters, such as indirect effects (including moderated indirect effects) and contextual effects. These parameters can be computed directly in the PARAMETERS section. For instance, to compute the mediation effects at the within- and between-levels, the following syntax can be used:

```
med_w = a_w * b_w + rCov;
med_b = a_b * b_b;
```

where med_w and med_b are defined as the within-school and the between-school mediation effect, respectively; and rCov denotes the covariance between two random slopes.

By default, Blimp estimates the relationship between random slopes as a correlation. To compute the within-school mediation effect involving random slopes, the correlation must be converted into the covariance. This can be achieved by labeling the correlation as RSb ~~ RSa@rCor, where rCorr represents the correlation between the two random slopes. Then, the covariance is then computed as $\sigma_{ab} = \rho_{ab}\sqrt{{\sigma_a}^2{\sigma_b}^2}$, where the total variance of two random slopes are obtained as RSa.totalvar and RSa.totalvar.

Blimp Syntax

```
PARAMETERS:
 #Covariance between two random slopes
 rCov = rCor * sqrt(RSa.totalvar * RSb.totalvar);
 #Mediation effects
 med_w = a_w * b_w + rCov;
 med_b = a_b * b_b;
```

Table 3 presents the estimated mediation effects at both the within-school (MED_W) and between-school (MED_B) levels. These effects describe the indirect influence of teacher-student relationships (REL) on math achievement (ACH) through math interest (INT). The within-school mediation effect (MED_W) is estimated at 0.151 (95% CI: 0.090, 0.223). This significant positive mediation effect implies that, within individual schools, improvements in teacher-student relations are associated with higher student interest in mathematics, which in turn leads to enhanced math achievement. The credible interval does not include zero, indicating a robust indirect effect that is consistently positive within schools. At the between-school level, the mediation effect (MED_B) is estimated at 0.343 (0.201, 0.525). This suggests that schools with stronger overall teacher-student relationships (REL) tend to have higher levels of student math interest (INT), which subsequently translates to improved math achievement (ACH) at the school level. Similar to the within-school effect, the between-school mediation effect is statistically significant as the credible interval does not include zero. However, it is slightly smaller in magnitude compared to the within-school mediation effect, suggesting that the impact of teacher-student relationships (REL) on math interest (INT) and subsequent achievement (ACH) is somewhat attenuated when aggregated across schools.

Contextual effects

Contextual effects can be computed similarly to mediation effects for a-, b-, and indirect effects. These effects are derived by subtracting the within-school effects from the between-school effects (e.g., a_con = a_b - a_w;).

Blimp Syntax

```
PARAMETERS:
 #Contextual effects
 a_con = a_b - a_w;
 b_con = b_b - b_w;
 med_con = med_b - med_w ;
```

Table 3 presents the estimated contextual effects for a-, b-, and indirect paths, respectively. The calculated contextual effects (A_CON and B_CON) provide additional insight into the discrepancies between within-school and between-school dynamics. The contextual effect for REL to INT (A_CON = 0.273) and for INT to ACH (B_CON = 0.115) are both positive, indicating that the between-school effects are stronger after accounting for the within-school relationships. This suggests that schools with higher average teacher-student relationship quality tend to foster greater student interest in math and achievement at a collective level, above and beyond individual-level effects.

The contextual mediation effect is significant (MED_CON = 0.192, 95% CI: 0.033, 0.383). This positive contextual mediation effect suggests that the indirect relationship between teacher-student relationships (REL) and math achievement (ACH), mediated through student interest in math (INT), is more pronounced at the between-school level than at the within-school level, although it is not significant. These contextual effects emphasize the importance of examining how group-level dynamics, such as the overall quality of teacher-student relationships within a school, can shape broader patterns of student outcomes. While individual-level relationships are important, the results suggest that the collective environment within a school—how teachers and students interact on average—may have an even greater impact on factors such as student interest and achievement. This suggests that the dynamics at the group or school level may not merely mirror individual-level patterns but instead offer additional insights into how a positive school climate collectively influences students. Focusing on these group-level effects allow us to gain a better understanding of how changes in the overall school environment can drive some significant improvements in student outcomes.

Model 4. Model with cross-level moderation effects

MSEM can incorporate cross-level moderation to explore how the relationships between variables vary based on between-school moderators. As illustrated in Figure 4, Model 4 employs individual socioeconomic status (SES) as a covariate, decomposed into student-level (SES_w) and school-level (SES_b) similar to Model 1. In this model, both SES_w and SES_b are modeled to influence math interest (INT) and math achievement (ACH) at both student- and school-levels. For moderated mediation, the school-level socioeconomic status (SES_b) influences the random slopes of a- and b-path. It indicates that the student-level effect of math teacher relations on math interest (a-path) may vary depending on the school-level SES, and the student-level effect of math interest on math achievement (b-path) may also depend on the school-level SES. By modeling these interactions, the analysis captures how school-level socioeconomic context shapes the relationships at the student level.

To estimate moderated mediation effects, conditional mediation effects can be calculated for different levels of the moderator, such as SES_b. Typically, these conditional effects are computed at ± one standard deviation from the mean of SES_b, enabling an examination of how the mediation effects vary depending on the level of the moderator. In Blimp, the same approach is used to calculate moderated mediation effects. When SES_b is one standard deviation above the mean, the conditional mediation effect is `(a_w + SESb_a * sqrt(SES_b.totalvar))*(b_w + SESb_b * sqrt(SES_b.totalvar))`. When SES_b is one standard deviation below the mean, it is `(a_w - SESb_a * sqrt(SES_b.totalvar)) * (b_w - SESb_b * sqrt(SES_b.totalvar))`.

Blimp Syntax

```
LATENT:
  ACH_w INT_w REL_w ;
  schid = ACH_b INT_b REL_b SES_b RSa RSb;

  MODEL:
  Within-level model:
  SES_w = SES - SES_b;
  SES ~ 1@SES_b;

  ACH_w -> ACH1:ACH4;
  INT_w -> INT1:INT4;
  REL_w -> REL1:REL4;

  ACH_w ~ INT_w@RSb REL_w SES_w;
  INT_w ~ REL_w@RSa SES_w;
  REL_w ~~ SES;

  Between-level model:
  SES_bc   = SES_b   - gm_SES;
  SES_b   ~ 1@gm_SES ;

  ACH_b -> ACH1:ACH4;
  INT_b -> INT1:INT4;
  REL_b -> REL1:REL4;

  ACH_b ~ INT_b@b_b REL_b SES_bc;
```

```
 INT_b ~ REL_b@a_b  SES_bc;

 REL_b ~~ SES_b ;

 RSb ~ 1@b_w SES_bc@SESb_b;
 RSa ~ 1@a_w SES_bc@SESb_a;

 RSb ~~ RSb;
 RSa ~~ RSa;
 RSb ~~ RSa@rCor;

PARAMETERS:
 rCov = rCor * sqrt(RSa.totalvar * RSb.totalvar);
 #Moderated mediation effects
 mome_p1 = (a_w + SESb_a*sqrt(SES_b.totalvar)) * (b_w +
SESb_b*sqrt(SES_b.totalvar)) + rCov;
 mome_m1 = (a_w - SESb_a*sqrt(SES_b.totalvar)) * (b_w -
SESb_b*sqrt(SES_b.totalvar)) + rCov;
```

[Figure 4 about here]

[Table 4 about here]

Cross-level moderation

Table 4 summarizes the results of the cross-level moderation effects of between-school SES on the a-path (RS2: REL to INT) and b-path (RS1: INT to ACH) within the multilevel mediation model. SES serves as a Level 2 covariate, representing whether schools are more affluent or economically disadvantaged, and its effects on the random slopes provide insights into the moderating role of school socioeconomic status on these mediation pathways.

The effect of SES on RSa (REL to INT), with an estimated median of 0.357 (95% CI: 0.253, 0.465) indicates a significant, positive moderation. This suggests that in schools with higher socioeconomic status, the positive effect of teacher-student relationships (REL) on interest in math (INT) is stronger. Essentially, in more affluent schools, improving teacher-student relationships seems to have a greater impact on fostering student interest in math, which may be attributed to additional resources or supportive environments available in such settings.

The effect of SES on RSb (INT to ACH), with an estimated median of 0.088 (95% CI: -0.015, 0.188), also reveals an insignificant but positive moderation. This suggests that the impact of math interest (INT) on math achievement (ACH) is considerably stronger in schools with higher SES, although this interpretation depends on strong causal assumptions. In affluent schools, increased student interest in math more effectively translates into improved math achievement, potentially due to enhanced access to learning materials, better instructional quality, or more encouragement and support from both teachers and parents. However, these interpretations are limited by the non-experimental nature of the simulated data, and causal inferences should be made with caution.

Moderated mediation effects

Table 4 presents the results of the moderated mediation effects, assessing how the mediation effects vary based on the level of school socioeconomic status (SES) at ±1 standard deviation from the mean. When SES is at 1 standard deviation above the mean (MOME_P1), the estimated mediation effect is 0.349 (95% CI: 0.219, 0.518). These significant positive mediation effects suggest that when school SES is high, the indirect effect from teacher-student relationships (REL) through math interest (INT) to math achievement (ACH) is strong. Specifically, in more affluent schools, enhanced teacher-student relationships may potentially be associated with increased student interest in mathematics, which effectively contributes to higher math achievement. Again, these interpretations are intended for illustrative purposes and should not be taken as evidence of causal relationships. Notably, the credible intervals do not include zero, indicating a robust positive effect at high levels of SES.

Conversely, when SES is at 1 standard deviation below the mean (MOME_M1), the moderated mediation effect is estimated at 0.009 (95% CI: -0.101, 0.076). The moderated mediation effect is statistically insignificant since the confidence interval includes zero. This suggests that in lower-SES schools, although teacher-student relationships are positively associated with student interest and, subsequently, math achievement, the effect is weaker compared to higher-SES schools.

Discussion

Social and behavioral researchers frequently confront the challenge of analyzing complex, nested data structures that require advanced statistical techniques. This tutorial offers a step-by-step guide for implementing Bayesian Multilevel Structural Equation Modeling (MSEM) using Blimp, a user-friendly software that supports flexible model specifications. Using an example of math achievement (ACH), math interest (INT), and teacher-student relationships (REL), we guide readers through the specification and estimation of progressively complex models in Blimp, demonstrating the power of Bayesian MSEM for addressing multilevel mediation and complex data structures.

We began with a basic multilevel path model, establishing a 1-1-1 mediation framework that distinguished the within- and between-group relationships. This initial model illustrated how Bayesian MSEM can address sampling error in nested data using observed variables, providing a conceptual foundation for understanding multilevel mediation processes. We then extended the model by incorporating latent variables to account for measurement error. This enabled constructs to be defined using multiple observed indicators, leading to more precise parameter estimates and a clearer understanding of the relationships between variables across both levels.

Next, we extended the model to include random slopes, allowing mediation paths to vary across clusters. This component demonstrated how Bayesian MSEM can accommodate heterogeneity in effects across groups, an important feature when modeling real-world variation. We further incorporated contextual effects, helping to separate the individual-level from group-level influences. Finally, we implemented a model with cross-level moderated mediation, where a Level 2 covariate—school-level socioeconomic status (SES)—was used to explain variability in the mediation paths. This demonstrated how group-level moderators can shape individual-level processes, emphasizing the integrative capacity of Bayesian MSEM for testing complex interaction effects across levels.

Throughout these four models, the tutorial emphasized practical syntax implementation in Blimp, clarified key statistical concepts (e.g., latent measurement, random effects, mediation, and moderation), and highlighted interpreting model results and understanding their implications in applied research contexts. These model-based examples underscore how Bayesian estimation allows for the analysis of complex models—including those involving random slopes and indirect effects—that might be challenging to estimate using traditional methods.

While the tutorial centers on two-level MSEM, the modeling framework and interpretation strategies can be extended to more complex settings, including longitudinal data structures, cross-lagged panel models, and higher-level hierarchies. Altogether, this resource equips researchers with both conceptual and practical tools for applying Bayesian MSEM to a range of empirical questions.

References

Asparouhov, T., & Muthén, B. (2010). Bayesian analysis of latent variable models using Mplus (Technical report, Version 4). Retrieved from http://www.statmodel.com

Bauer, D. J., Preacher, K. J., & Gil, K. M. (2006). Conceptualizing and testing random indirect effects and moderated mediation in multilevel models: new procedures and recommendations. Psychological methods, 11(2), 142.

Bentler, P. M. (2006). *EQS 6 structural equations program manual*. Encino, CA: Multivariate Software, Inc.

Bollen, K. A. (1987). Total, direct, and indirect effects in structural equation models. *Sociological methodology*, 37-69.

Brooks, S. P., & Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of computational and graphical statistics*, *7*(4), 434-455.

Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: a new look at an old issue. *Psychological methods*, *12*(2), 121.

Enders, C. K., Du, H., & Keller, B. T. (2020). A model-based imputation procedure for multilevel regression models with random coefficients, interaction effects, and nonlinear terms. *Psychological methods*, *25*(1), 88.

Etz, A., & Vandekerckhove, J. (2018). Introduction to Bayesian inference for psychology. Psychonomic Bulletin & Review, 25(1), 5-34.

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical science*, *7*(4), 457-472.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). Bayesian data analysis (3rd ed.). Chapman and Hall/CRC.

Goodman, L. A. (1960). On the exact variance of products. Journal of the American statistical association, 55(292), 708-713.

Kaplan, D. (1998). Methods for multilevel data analysis. *Modern methods for business research*, 337-357.

Kaplan, D., & Depaoli, S. (2012). Bayesian structural equation modeling.

Keller, B.T. (2022). An introduction to factored regression models with BLIMP. Psych, 4(1), p.10–37.

Keller, B.T., & Enders, C.K. (2023). An investigation of factored regression missing data methods for multilevel models with cross-level interactions. Multivariate Behavioral Research, 58(5), p.938–963.

Keller, B.T., & Enders, C.K. (2023). Blimp user's guide (Version 3). Blimp Software: Los Angeles, CA, USA. Retrieved from www.appliedmissingdata.com/blimp.

Kenny, D. A., Korchmaros, J. D., & Bolger, N. (2003). Lower level mediation in multilevel models. Psychological methods, 8(2), 115.

Krull, J. L., & MacKinnon, D. P. (2001). Multilevel modeling of individual and group level mediated effects. *Multivariate behavioral research*, *36*(2), 249-277.

Kruschke, J. K., & Liddell, T. M. (2018). The Bayesian new statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. Psychonomic Bulletin & Review, 25(1), 178-206.

Laird, N. M., & Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, 963-974.

Lee, S.-Y., & Shi, J.-Q. (2000). Joint bayesian analysis of factor scores and structural parameters in the factor analysis model. Annals of the Institute of Statistical Mathematics, 52, 722–736.

Levy, R., & McNeish, D. (2023). Perspectives on Bayesian inference and their implications for data analysis. *Psychological Methods*, *28*(3), 719.

Lunn, D. J., Thomas, A., Best, N., & Spiegelhalter, D. (2000). WinBUGS - A Bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing, 10*(4), 325–337. https://doi.org/10.1023/A:1008929526011

McElreath, R. (2020). Statistical rethinking: A Bayesian course with examples in R and Stan (2nd ed.). Chapman and Hall/CRC.

McNeish, D. (2016). On using Bayesian methods to address small sample problems. Structural Equation Modeling: A Multidisciplinary Journal, 23(5), 750-773.

Merkle, E. C. (n.d.). *Two-level SEM*. blavaan. Retrieved March 25, 2025, from https://ecmerkle.github.io/blavaan/articles/multilevel.html#references

Merkle, E. C., & Rosseel, Y. (2018). blavaan: Bayesian structural equation models via parameter expansion. *Journal of statistical software*, *85*, 1-30.

Merkle, E. C., Fitzsimmons, E., Uanhoro, J., & Goodrich, B. (2021). Efficient Bayesian structural equation modeling in Stan. *Journal of Statistical Software*, *100*, 1-22.

Muthén, L.K. & Muthén, B.O. (1998-2021). Mplus User's Guide. Eighth Edition. Los Angeles, CA: Muthén & Muthén

Muthén, B., & Asparouhov, T. (2012). Bayesian structural equation modeling: A more flexible representation of substantive theory. *Psychological Methods, 17*(3), 313–335. https://doi.org/10.1037/a0026802

Muthén, B., & Asparouhov, T. (2015). Causal effects in mediation modeling: An introduction with applications to latent variables. Structural equation modeling: a multidisciplinary journal, 22(1), 12-23.

Palomo, J., Dunson, D. B., & Bollen, K. (2007). Bayesian structural equation modeling. In Handbook of latent variable and related models (pp. 163–188). Elsevier.

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

Raudenbush, S. W. (2002). Hierarchical linear models: Applications and data analysis methods. *Advanced Quantitative Techniques in the Social Sciences Series/SAGE*.

Rhemtulla, M., van Bork, R., & Borsboom, D. (2020). Worse than measurement error: Consequences of inappropriate latent variable measurement models. *Psychological Methods*, *25*(1), 30.

Stan Development Team. (2020). *Stan modeling language: User's guide and reference manual* (Version 2.24). Retrieved from https://mc-stan.org

Wang, J., & Wang, X. (2019). *Structural equation modeling: Applications using Mplus*. John Wiley & Sons.

Yuan, Y., & MacKinnon, D.P. (2009). Bayesian mediation analysis. Psychological Methods, 14(4), p.301–322.

Table 1. Parameter Estimates for 1-1-1 Mediation Model with Observed Variables

	Est (Median)	SD	Lower CI	Upper CI	$\hat{R}$
Parameters			Within Level		
INT ← REL	**0.279**	0.020	0.240	0.319	1.0003
ACH ← INT	**0.430**	0.018	0.395	0.465	1.0001
ACH ← REL	**0.224**	0.020	0.184	0.262	1.0001
Parameters			Between Level		
INT ← REL	**0.621**	0.083	0.455	0.783	1.0003
ACH ← INT	**0.550**	0.096	0.361	0.737	1.0004
ACH ← REL	**0.277**	0.094	0.090	0.458	1.0007

Note. CI = 95% credible intervals. Bolded estimates have credible intervals that do not include zero, indicating statistical significance.

Table 2. Parameter Estimates for 1-1-1 Mediation Model with Latent Variables

	Est (Median)	SD	Lower CI	Upper CI	Std.Est.
Parameters			Within Level		
INT ← REL	**0.338**	0.026	0.288	0.389	0.292
ACH ← INT	**0.475**	0.023	0.431	0.520	0.457
ACH ← REL	**0.245**	0.025	0.197	0.294	0.203
ACH2 by ACH	**1.024**	0.022	0.982	1.067	0.659
ACH3 by ACH	**0.991**	0.021	0.950	1.033	0.658
ACH4 by ACH	**0.984**	0.021	0.943	1.026	0.656
INT2 by INT	**0.959**	0.022	0.917	1.004	0.621
INT3 by INT	**0.954**	0.022	0.913	0.998	0.621
INT4 by INT	**0.960**	0.022	0.918	1.004	0.631
REL2 by REL	**0.972**	0.027	0.921	1.027	0.578
REL3 by REL	**0.939**	0.026	0.889	0.992	0.557
REL4 by REL	**1.006**	0.028	0.953	1.062	0.604
Parameters			Between Level		
INT ← REL	**0.601**	0.091	0.432	0.789	0.667
ACH ← INT	**0.577**	0.117	0.358	0.819	0.589
ACH ← REL	**0.220**	0.096	0.035	0.415	0.251
ACH2 by ACH	**1.150**	0.093	0.977	1.347	0.532
ACH3 by ACH	**1.102**	0.092	0.934	1.293	0.524
ACH4 by ACH	**1.136**	0.087	0.978	1.319	0.544
INT2 by INT	**1.048**	0.088	0.887	1.236	0.521
INT3 by INT	**1.112**	0.085	0.954	1.286	0.554
INT4 by INT	**1.054**	0.083	0.908	1.234	0.532
REL2 by REL	**0.940**	0.072	0.808	1.087	0.556
REL3 by REL	**1.023**	0.071	0.899	1.176	0.604
REL4 by REL	**0.902**	0.067	0.780	1.047	0.540

Note. CI = 95% credible intervals. Bolded estimates have credible intervals that do not include zero, indicating statistical significance.

Table 3. Parameter Estimates for 1-1-1 Mediation Model with Two Random Slopes

	Est (Median)	SD	Lower CI	Upper CI
Parameters		Between Level		
RSa Mean	**0.331**	0.061	0.213	0.452
RSb Mean	**0.458**	0.051	0.359	0.559
RSa Var	**0.309**	0.056	0.221	0.441
RSb Var	**0.212**	0.039	0.151	0.301
RSa, RSb Cor	0.027	0.031	-0.033	0.091
MED_B	**0.343**	0.083	0.201	0.525
MED_W	**0.151**	0.034	0.090	0.223
A_CON	**0.273**	0.110	0.063	0.494
B_CON	0.115	0.128	-0.121	0.380
MED_CON	**0.192**	0.089	0.033	0.383

Note. CI = 95% credible intervals. Bolded estimates have credible intervals that do not include zero, indicating statistical significance.

Table 4. Parameter Estimates for 1-1-1 Mediation Model with Cross-Level Moderation Effects

	Est (Median)	SD	Lower CI	Upper CI
Parameters		Between Level		
RSa ← SES	**0.357**	0.054	0.253	0.465
RSb ← SES	0.088	0.051	-0.015	0.188
MOME_P1	**0.349**	0.076	0.219	0.518
MOME_M1	0.009	0.045	-0.101	0.076

Note. CI = 95% credible intervals. Bolded estimates have credible intervals that do not include zero, indicating statistical significance.

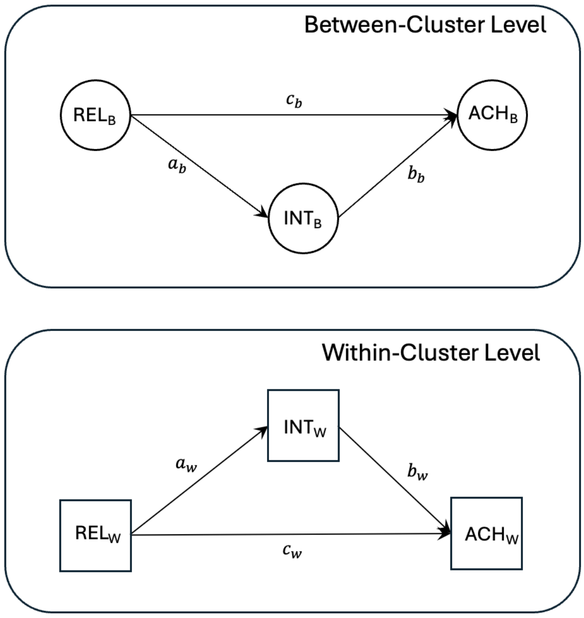

Figure 1. 1-1-1 Mediation Model with Observed Variables

Note. residual terms are omitted.

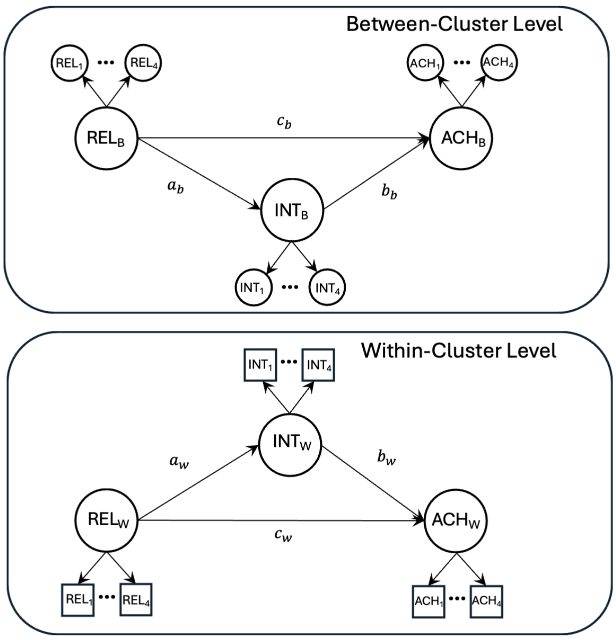

Figure 2. 1-1-1 Mediation Model with Latent Variables

Note. residual terms are omitted.

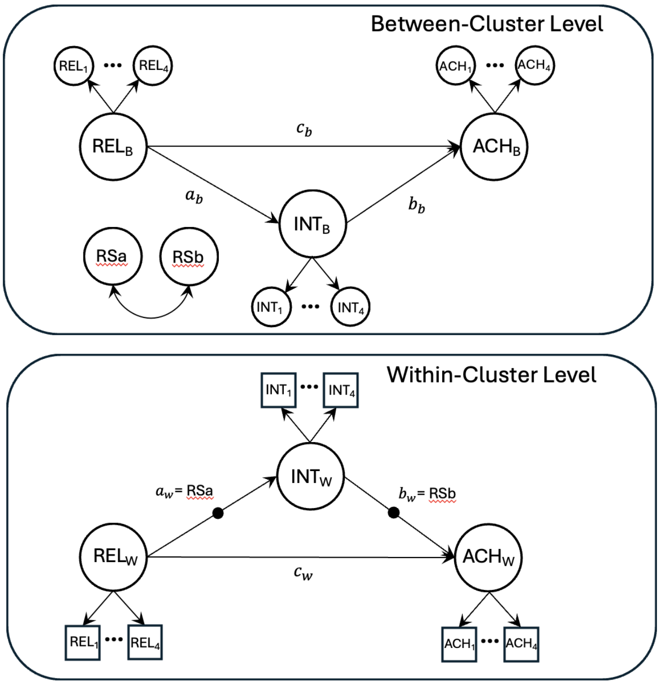

Figure 3. 1-1-1 Mediation Model with Two Random Slopes

Note. residual terms are omitted.

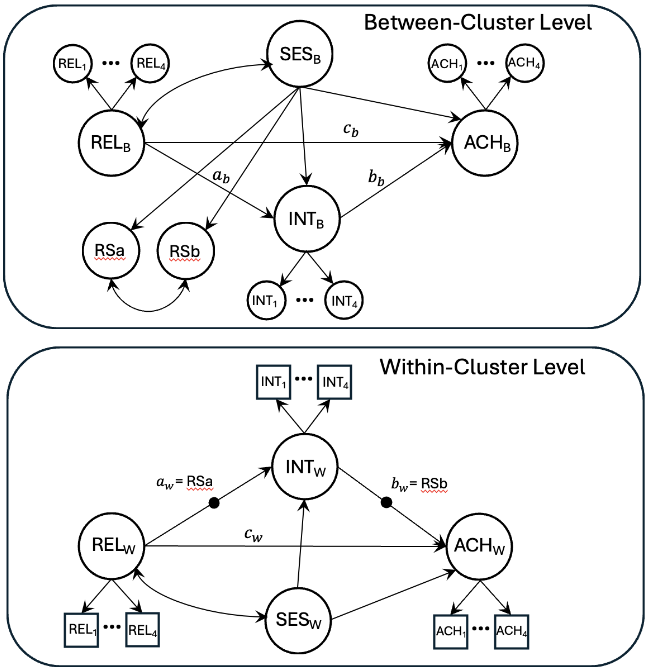

Figure 4. 1-1-1 Mediation Model with Cross-Level Moderation Effects

Note. residual terms are omitted.