Multivariate Multilevel Models and Three-Level Models: Finally, The Finale!

- Topics:
  - Review: univariate centering of L1 person predictors
  - Multivariate latent centering and its extensions
    - Pros/cons, multilevel mediation, and location–scale models
  - A little bit about three-level models
    - Notation and multiple intraclass correlations
    - What slopes can be random over what levels
    - Variable-centering and constant-centering of L1 and L2 predictors

3 Kinds of Fixed Slopes for L1 Predictors

- **Is there a Level-1 Within-Cluster (WC) slope?**
  - If you have a higher $L1x_{pc}$ predictor value *than others in your cluster*, do you also have a higher (or lower) y_{pc} outcome value *than others in your cluster*?
  - If so, the **level-1 within-cluster *part* of the L1 predictor** will reduce the level-1 residual variance (σ_e^2) of the y_{pc} outcome

- **Is there a Level-2 Between-Cluster (BC) slope?**
  - Do clusters with higher average $L1x_{pc}$ predictor values *than other clusters* also have higher (or lower) average y_{pc} outcomes *than other clusters*?
  - If so, the **level-2 between-cluster *part* of the L1 predictor** will reduce level-2 random intercept variance ($\tau_{U_0}^2$) of the y_{pc} outcome

- **Is there a Level-2 Contextual slope: Do the L2 BC and L1 WC slopes differ?**
  - After controlling for the actual value of L1 predictor, is there still **an incremental contribution** from the **level-2 between-cluster *part* of the L1 predictor** (i.e., does a cluster's general tendency matter beyond a person's $L1x_{pc}$ value)?
  - Equivalently, the **Level-2 Contextual slope** = **L2 BC slope** − **L1 WC slope**, so the Level-2 Contextual slope directly tests **if a smushed slope is ok (pry not!)**

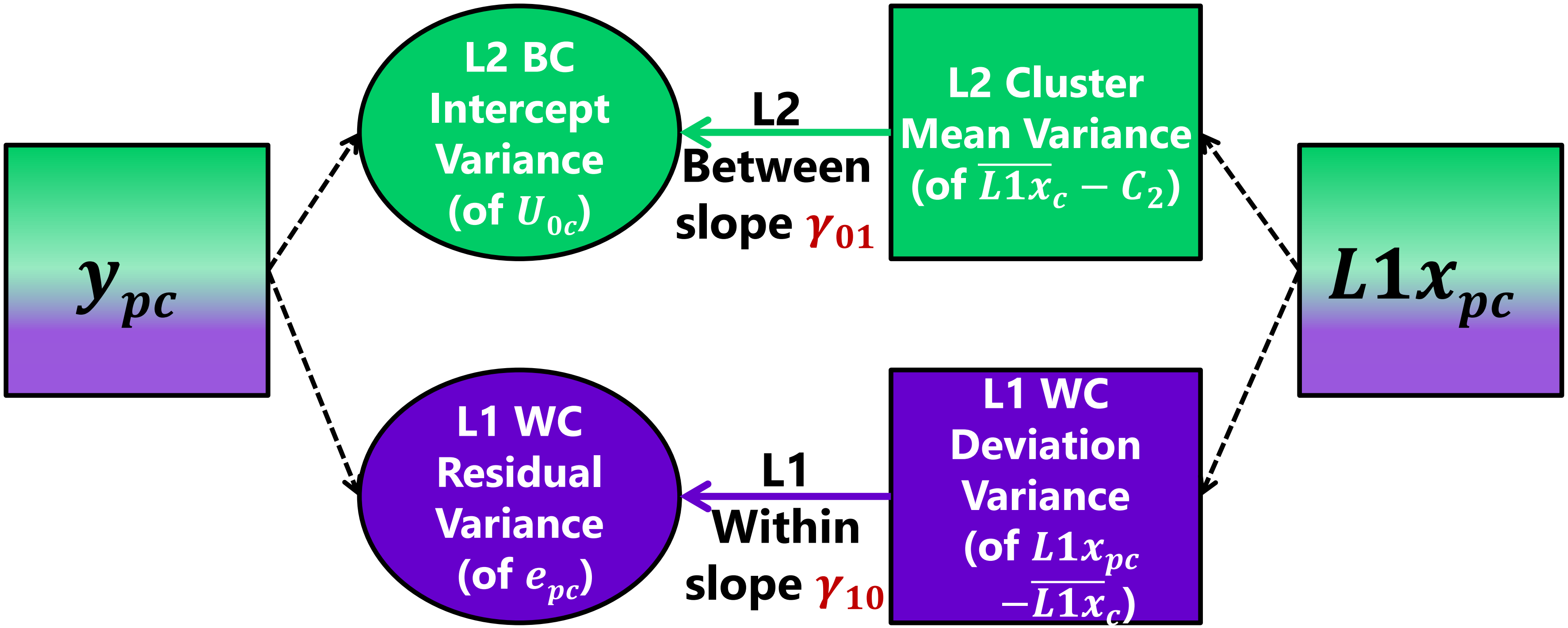
3 Options to Prevent Smushed Slopes

- Within Univariate MLM framework (predict only one outcome):
  1. **Cluster-mean-centering**: manually carve up L1 predictor into its level-specific parts using observed variables (1 predictor per level)
     - More generally, this is "**variable-centering**" because you are **subtracting a variable** (e.g., the cluster mean here; could use other cluster variables)
     - Will always yield **level-1 within slopes** and **level-2 between slopes**!
  2. **Grand-mean-centering**: do NOT carve up L1 predictor into its level-specific parts, but add level-2 mean to distinguish level-specific slopes
     - More generally, this is "**constant-centering**" because you are **subtracting a constant** while still keeping all sources of variance in the L1 predictor
     - **Choice of constant is irrelevant** (changes where 0 is, not what variance it has)
     - Will always yield **level-1 within slopes** and **level-2 contextual slopes**!
- Within Multivariate MLM framework (i.e., via Multilevel-SEM):
  3. **Latent-centering**: Treat the L1 predictor as another outcome → let the model carve it up into **level-specific latent variables**
     - Best in theory, but the type of level-2 slope (between or contextual) depends on model type, syntax type, and the estimator in Mplus! (Hoffman, 2019)

1.Cluster-Mean-Centering

Model-based partitioning of level-1 y_{pc} outcome into level-specific **latent variables**

Manual partitioning of level-1 $L1x_{pc}$ predictor into level-specific **observed variables**

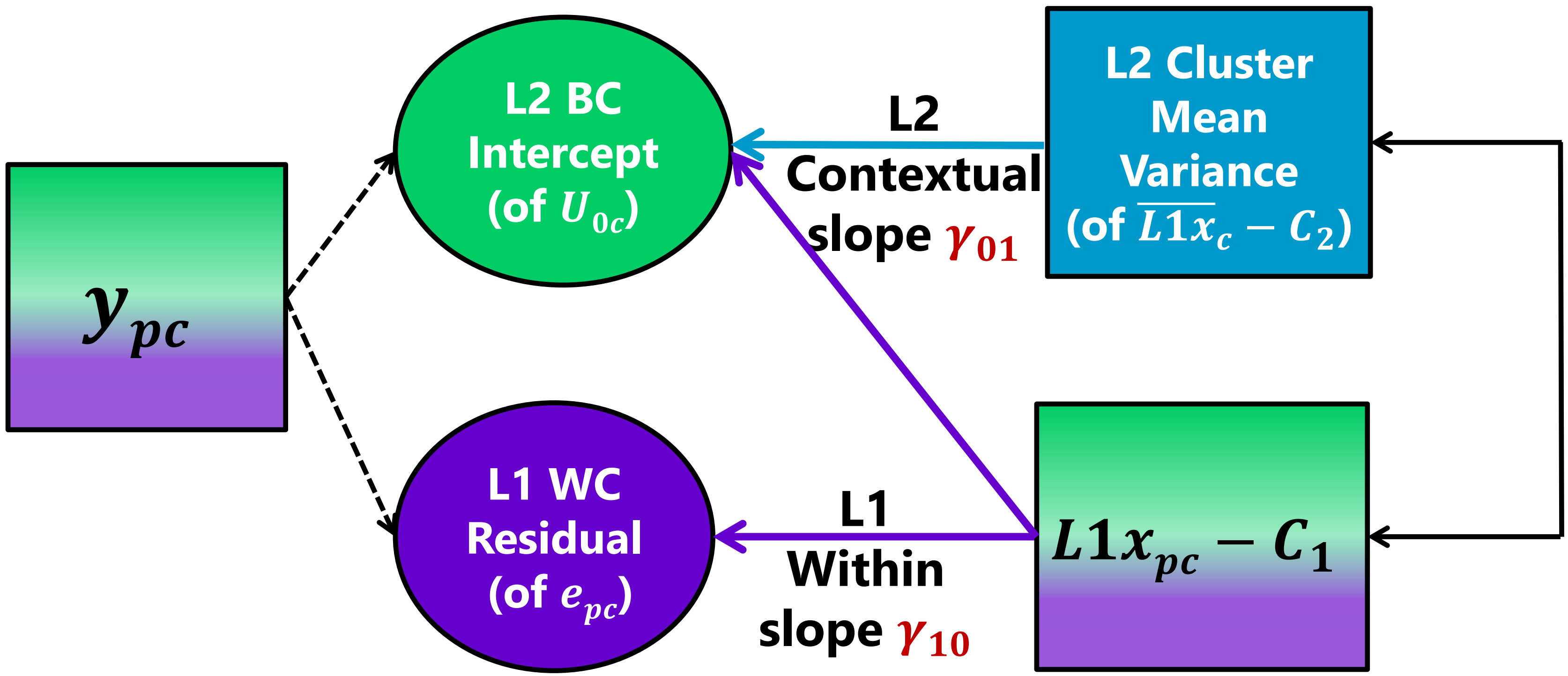


Why not let the model make variance components for $L1x_{pc}$, too?
That is option 3, multivariate MLM (or "multilevel SEM"): stay tuned…

2. Constant-Centering + Cluster Mean

Model-based partitioning of y_{pc} outcome into level-specific **latent variables**

$L1x_{pc}$ **is still NOT partitioned**, but cluster mean $\overline{L1x_c} - C_2$ is added to allow an **incremental L2 effect**



L2 BC slope = L1 WC slope + Level-2 Contextual slope

Because original $L1x_{pc}$ still has L2 BC variance, it still carries ***some*** of the L2 BC effect…

Preventing Smushed (BC=WC) Slopes

- **Fixed side: 2 univariate strategies to prevent smushed slopes**
  - If using cluster-MC L1 $\boldsymbol{WCx_{pc}}$, it can only have a **L1 within slope**, and its L2 $\boldsymbol{CMx_c}$ can only have a **L2 between slope** (so there's no problem)
  - If using constant-C L1 $\boldsymbol{L1x_{pc}}$, its L1 slope will be smushed (assume BC=WC) if you don't add its L2 $\boldsymbol{CMx_c}$ to allow a **L2 contextual slope = BC − WC**

- **Random side: Only 1 univariate strategy is likely possible!** *(see Rights & Sterba, MBR in press, for details)*
  - If using cluster-MC L1 $\boldsymbol{WCx_{pc}}$, its L2 random slope variance **only** captures L2 BC differences in its L1 WC slope (so there's no problem)
    - Creates a pattern of quadratic heterogeneity of variance **across** $\boldsymbol{WCx_{pc}}$ **ONLY**
  - If using constant-C L1 $\boldsymbol{L1x_{pc}}$, its L2 random slope variance **also** creates **intercept heterogeneity of variance** (beyond BC diffs in L1 WC slope)
    - Enforces **SAME** quadratic heterogeneity of variance across **L1** $\boldsymbol{WCx_{pc}}$ and **L2** $\boldsymbol{CMx_c}$
  - If using $\boldsymbol{L1x_{pc}}$, you need a "contextual" random slope to allow a **different** pattern of variance heterogeneity across $\boldsymbol{CMx_c}$ than $\boldsymbol{WCx_{pc}}$ (for BC − WC)
    - Requires a L2 BC random "slope-ish" variance for **L2** $\boldsymbol{CMx_c}$ – good luck estimating it!

NEW Option 3. Latent-Centering

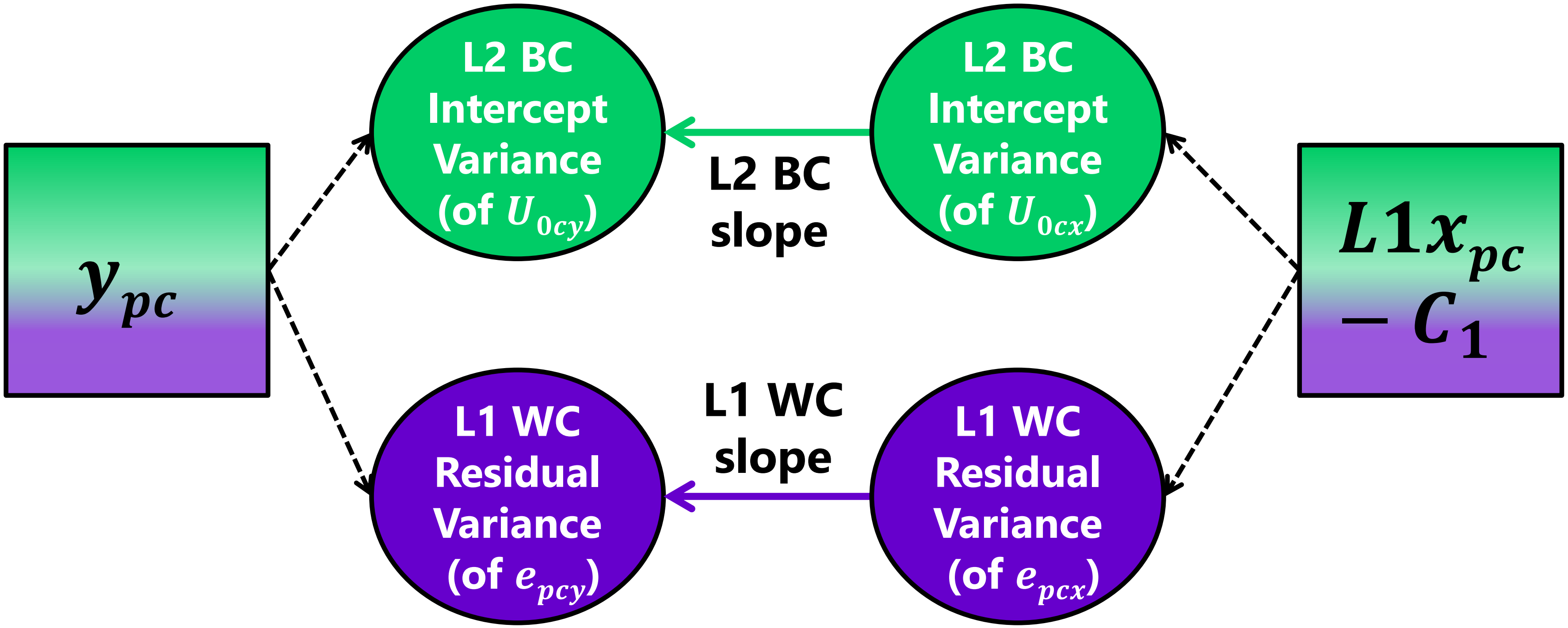
- We let the model partition the L1 predictor $\boldsymbol{L1x_{pc}}$ into two latent variables that directly represent its **L2 between**-cluster (BC) and **L1 within**-cluster (WC) sources of variation, just as we did for y_{pc} !
  - At a minimum: Fit an empty means, random intercept model for $L1x_{pc}$ (centered ahead of time at a constant so that 0 is still meaningful)
  - Level-2 BC differences are represented by L2 random intercept for $L1x_{pc}$ *(instead of observed cluster mean, $\boldsymbol{\overline{L1x}_c} - \boldsymbol{C_2}$, as in cluster-MC)*
  - Level-1 WC differences are represented by L1 residual for $L1x_{pc}$ *(instead of observed cluster mean deviation, $\boldsymbol{L1x_{pc}} - \boldsymbol{\overline{L1x}_c}$, as in cluster-MC)*

- Requires multivariate software that can predict more than one column (either multilevel-SEM, aka M-SEM, or single-level SEM) if you want to still predict y_{pc} from x_{pc} (not just have them covary)
  - Best in theory given a “large enough” sample at both levels, but it gets complicated quickly: The type of level-2 slope (between or contextual) depends on type of model, syntax, and estimator in Mplus! (Hoffman, 2019)

3. Latent-Centering in Multivariate MLM

Model-based partitioning of level-1 y_{pc} outcome into level-specific **latent variables**

Model-based partitioning of level-1 $L1x_{pc}$ predictor (= outcome now) into level-specific **latent variables**



Univariate MLM software can be tricked into multivariate MLM if the relationships between X and Y at each level are phrased as covariances, but not if you want directed regressions (or moderators thereof)

Table 1. Summary of Modeling Choices and Level 2 Results

Level 1 source of variance and type of effects	Mplus syntax for the Level 1 effect	Resulting Level 2 effect
Univariate MLM with ML or Bayesian estimation		
Variable-centered observed variable		
Fixed effects only	Level 1 direct	Between[a]
Fixed effects only	Level 1 placeholder	Between[a]
Fixed and random effects	Level 1 placeholder	Between[a]
Constant-centered observed variable		
Fixed effects only	Level 1 direct	Contextual[a]
Fixed effects only	Level 1 placeholder	Contextual[a]
Fixed and random effects	Level 1 placeholder	Contextual[a]
Multivariate MLM using Mplus multilevel structural equation modeling with ML estimation		
Within-level latent variable		
Fixed effects only	Level 1 direct	Between[b]
Uncentered observed variable		
Fixed effects only	Level 1 placeholder	Contextual[b]
Fixed and random effects	Level 1 placeholder	Contextual[b]
Multivariate MLM using Mplus 8.1+ multilevel structural equation modeling with Bayesian estimation		
Within-level latent variable		
Fixed effects only	Level 1 direct	Between[b]
Fixed effects only	Level 1 placeholder	Between[b]
Fixed and random effects	Level 1 placeholder	Between[b]
Multivariate MLM using general structural equation modeling with ML or Bayesian estimation		
Latent residual of observed variable		
Fixed effects only	Residual direct	Contextual[b]
Fixed effects only	Structured residual	Between[b]
Fixed and random effects	Residual direct through placeholder	Contextual[b]
Fixed and random effects	Structured residual through placeholder	NA

Note: MLM = multilevel model; ML = maximum likelihood; NA = not available.
[a]These Level 2 effects are fixed effects for observed Level 2 mean predictors (included in all univariate models).
[b]These Level 2 effects are effects for latent Level 2 intercept predictors (included in all multivariate models).

Table 1 from Hoffman 2019

We've only used these options so far

Random slopes are also **smushed** with this method

Random slopes are **unsmushed** with this method

Single-level SEM is not very useful for clustered data (but it can be for longitudinal data)

Troubleshooting Tips: Are My Level-2 Slopes Between or Contextual?

- Start with a simplified multivariate MLM in which each pile of variance for y_{pc} is predicted by only one pile of variance for x_{pc} at a time
  - Goal: Recover bivariate relations without contamination by how slopes change when they are "unique" effects controlling for other predictors

- Concern is relevant when same variables have **slopes at both levels**
  - e.g., piles of variance for $\boldsymbol{x_{pc}}$ → L2 $\boldsymbol{BCx_c}$ intercept, L1 $\boldsymbol{WCx_{pc}}$ residual
  - e.g., piles of variance for $\boldsymbol{y_{pc}}$ → L2 $\boldsymbol{BCy_c}$ intercept, L1 $\boldsymbol{WCy_{pc}}$ residual
  - If there is a L1 $\boldsymbol{WCx_{pc}}$ → $\boldsymbol{WCy_{pc}}$ slope, then fixed slopes for the $\boldsymbol{BCx_c}$ → $\boldsymbol{BCy_c}$ intercept relations could be **L2 contextual slopes instead of L2 between slopes**

- **How to check?** Compare L2 slope results from two models:
  - A) $\boldsymbol{WCx_{pc}}$ → $\boldsymbol{WCy_{pc}}$ **fixed slope** (no random slope var or cross-level interactions)
  - B) $\boldsymbol{WCx_{pc}}$ → $\boldsymbol{WCy_{pc}}$ **covariance** (no random slope var or cross-level interactions)
  - If any L2 slopes changed notably, they must be L2 **contextual** (because they are **controlled for the L1 slope only in A,** whereas **between** slopes don't control for L1)

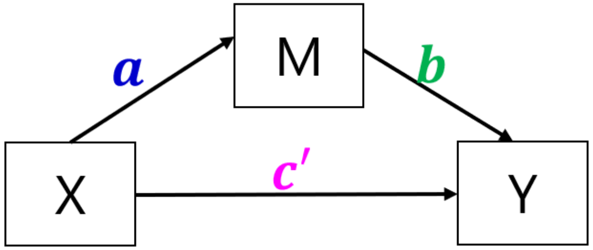
Univariate vs. Multivariate MLM (M-SEM)

- **Cons of Multivariate MLMs (M-SEM):**
  - Current software does not have REML or denominator DF → not good for small L2 samples (see McNeish & Stapleton, 2016; McNeish 2017)
  - Interactions among what used to be observed cluster means become interactions among latent variables (random effects) → harder to estimate
  - Whether your L2 slopes are between or contextual varies by software used, syntax specification, and method of estimation! (see Hoffman 2019)

- **Pros of Multivariate MLMs (via M-SEM):**
  - Univariate MLM uses observed variables for x_{pc} but latent variables for y_{pc}; multivariate MLMs use latent variables both x_{pc} and y_{pc} (more reasonable)
  - Simulation research suggests that the L2 fixed slopes in M-SEM are less biased (because cluster means are not perfectly reliable as assumed), but the L2 fixed slopes also more inconsistent across samples, particularly for variables with lower ICCs (little intercept info) and smaller level-1 samples
    - e.g., your readings: Lüdtke et al.: 2008, 2011 and Preacher et al.: 2010, 2011, and 2016
  - Only way to test single-stage mediation at multiple levels simultaneously or use cluster differences in within-cluster variance ("location–scale" models)

Implications for Multilevel Mediation

- Mediation is more complex in multilevel samples and only logically possible at two levels for **one combination**, as shown below
  - By mediation, I mean "M is part of the reason why X → Y" theoretically
  - Although indirect effects can always be computed, they may not make sense
  - Below: Is each variable measured at Level 2 or Level 1 (= both L1+L2)



X predictor	M mediator	Y outcome	L1 mediation?	L2 mediation?
2	2	2	no	yes
2	2	1	no	yes
2	1	2	no	yes
2	1	1	no	yes
1	2	2	no	yes
1	2	1	no	yes
1	1	2	no	yes
1	**1**	**1**	**yes**	**yes**

Multilevel Models: Differing Variances?

- M-SEM in Mplus specifically offers extensions for "location–scale mixed-effects models" by which to examine L2 diffs in L1 residual variance $\boldsymbol{\sigma_e^2}$ for cluster c as an outcome OR a predictor
  - **Location model** = what you already know as MLM; new part is **scale model** = how differences in variance can be quantified and predicted
  - **Scale model:** log($\boldsymbol{\sigma_{ec}^2}$) = fixed int + cluster predictors + random scale
    - A "random scale (factor)" is a separate random intercept for L1 residual variance = cluster differences in extent of within-cluster residual variance

- e.g., L1 student evaluations of a L2 instructor as the target, or L1 employee evaluations of a L2 supervisor as the target
  - L1 $\boldsymbol{\sigma_e^2}$ for cluster c = "unreliability" → disagreement about same target
  - L2 diffs in L1 $\boldsymbol{\sigma_{ec}^2}$ → differential amounts of target disagreement
  - So why do some targets have more disagreement? Add target predictors!
  - See also your readings: Lester et al., 2021 and Hoffman & Walters, 2022

A Little Bit about Three-Level Models

- Number of levels needed is determined by the dimensions of sampling in the outcome, NOT in the predictors
  - e.g., students within schools within districts?
    - Student outcomes (did that student graduate?) need a three-level model
    - School outcomes (school graduation rate) only need a two-level model (any student information would have to be aggregated to be a predictor)

- Nesting/crossing pattern depends on sampling design!
  - e.g., multiple people within multiple countries over multiple years?
    - If same countries and same people are measured repeatedly:
      - L1 = year, L2 = person, L3 = country
    - If same countries are measured but different with people each year:
      - L1 = person, L2 = year, L3 = country
    - If different countries and different people are measured each year:
      - L1 = person, L2 = country, L3 = year

Empty Means, 3-Level Random Intercept Model: Example for Three-Level Clustered Data

Notation: t = L1 teacher, s = L2 school, d = L3 district

Level 1: $y_{tsd} = \beta_{0sd} + e_{tsd}$

Residual = teacher-specific deviation from school's predicted outcome

Level 2: $\beta_{0sd} = \delta_{00d} + U_{0sd}$

School Random Intercept = school-specific deviation from district's predicted outcome

Level 3: $\delta_{00d} = \gamma_{000} + V_{00d}$

Fixed Intercept =grand mean of district means

District Random Intercept = district-specific deviation from fixed intercept

4 Total Parameters:

Model for the Means (1):

- Fixed Intercept γ_{000}

Model for the Variance (3):

- Level-1 Variance of e_{tsd} → σ_e^2
- Level-2 Variance of U_{0sd} → $\tau_{U_0}^2$
- Level-3 Variance of V_{00d} → $\tau_{V_{00}}^2$

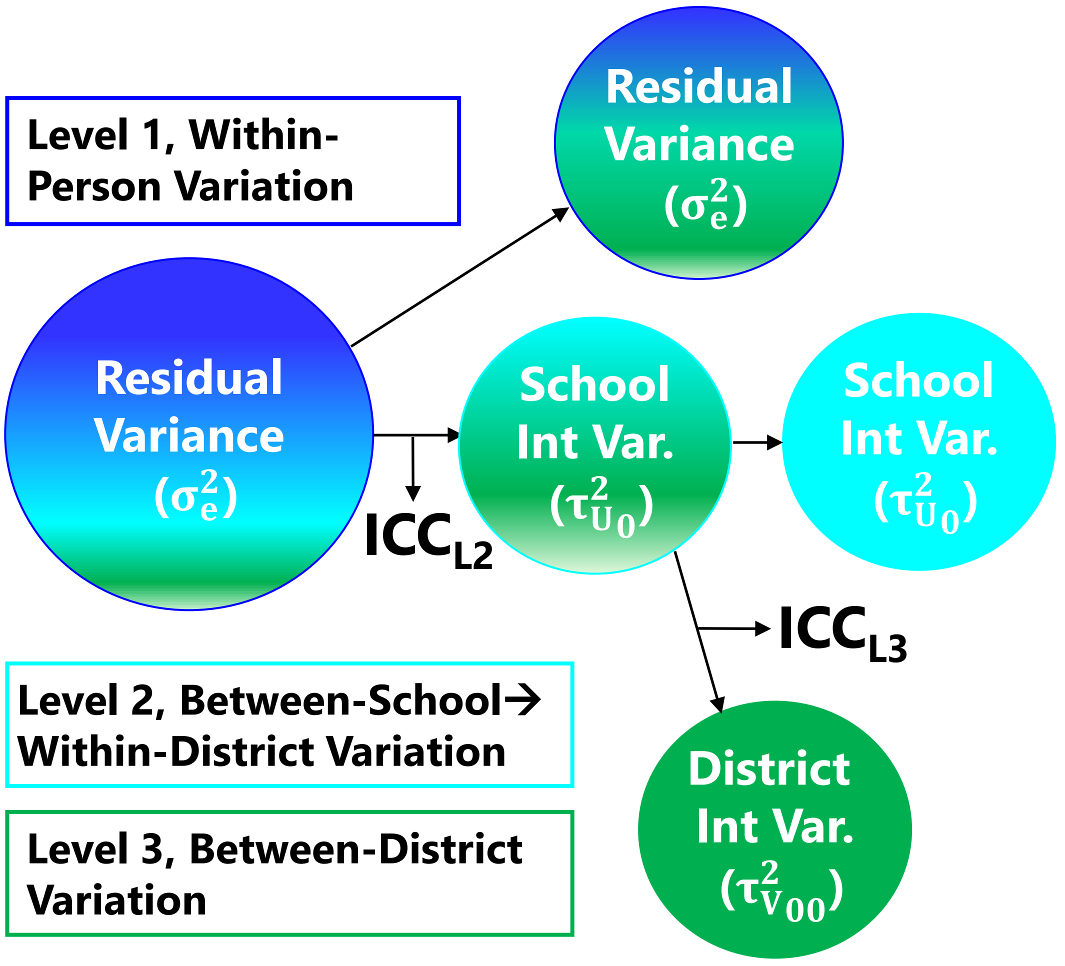
Composite equation:

$y_{tsd} = \gamma_{000} + V_{00d} + U_{0sd} + e_{tsd}$

Btw: My bad for reusing "V"

Example 3-Level Random Intercept Model

- Example empty means, random intercept 3-level model of L1 teachers within L2 schools within L3 districts:



ICCs in a 3-Level Random Intercept Model: L1 Teachers within L2 Schools within L3 Districts

- **ICC for level 2 (and level 3) relative to level 1:**

- $$\mathrm{ICC_{L2}} = \frac{\text{Between-School}}{\text{Total}} = \frac{\text{L3+L2}}{\text{L3+L2+L1}} = \frac{\boldsymbol{\tau^2_{V_{00}} + \tau^2_{U_0}}}{\boldsymbol{\tau^2_{V_{00}} + \tau^2_{U_0} + \sigma^2_e}}$$

  → This ICC expresses the similarity of **teachers from the same school** (and by definition, from the same district) → of the ***total outcome variation***, how much of it is ***between schools and districts (due to organizations)***?

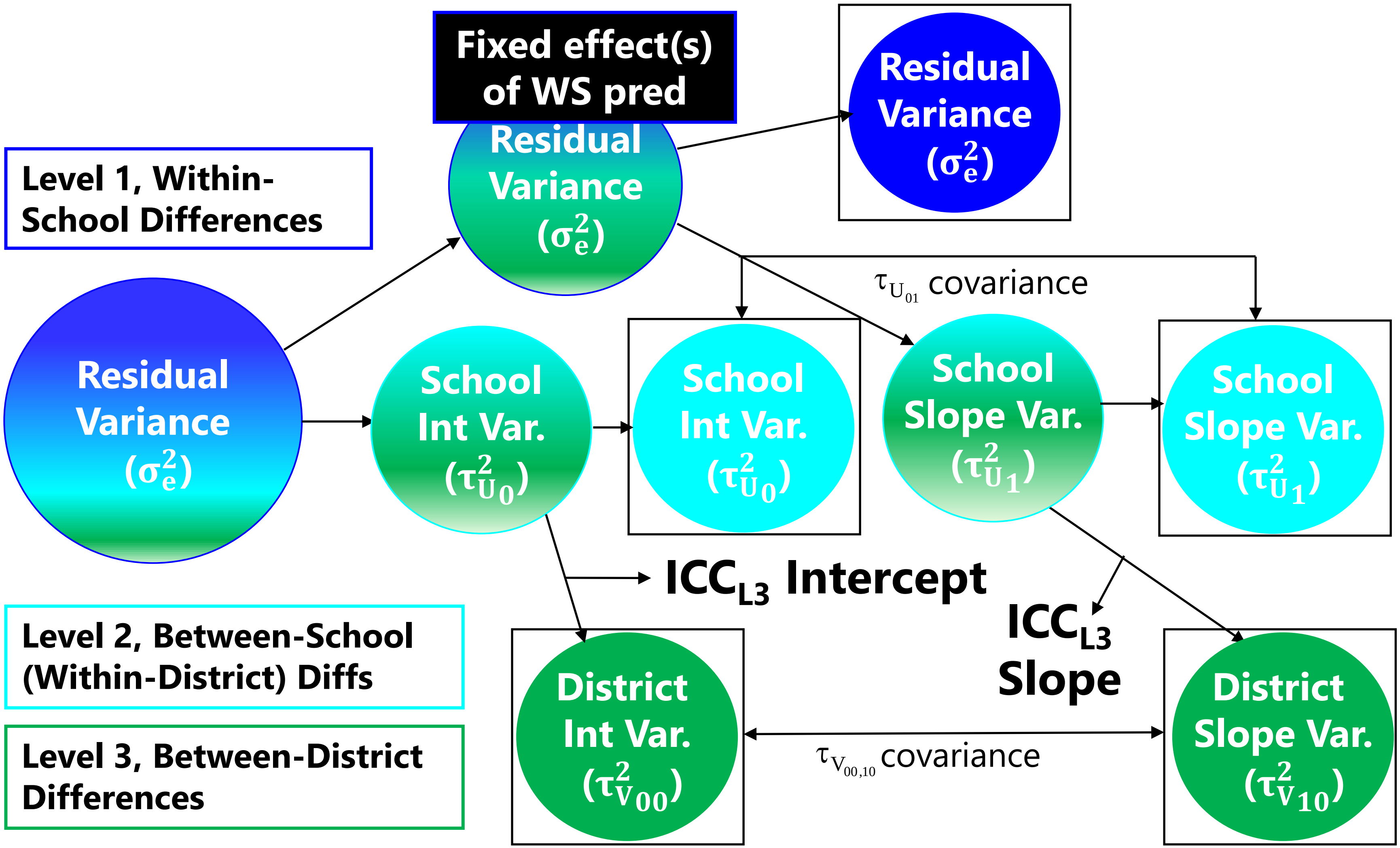
- **ICC for level 3 relative to level 2 (ignoring level 1):**

- $$\mathrm{ICC_{L3}} = \frac{\text{Between-District}}{\text{Between-School}} = \frac{\text{L3}}{\text{L3+L2}} = \frac{\boldsymbol{\tau^2_{V_{00}}}}{\boldsymbol{\tau^2_{V_{00}} + \tau^2_{U_0}}}$$

  → This ICC expresses the similarity of **schools from the same district** (ignoring within-school variation over teachers) → of ***that total between-school outcome variation***, how much of that is actually ***between districts***?

Example 3-Level Random Change Model

- Can have random effects of L1 predictors over L2 and L3:



Example 3-Level Random Slope Model

Notation: t = L1 teacher, s = L2 school, d = L3 district

L1: $y_{tsd} = \beta_{0sd} + \beta_{1sd}(WSx_{tsd}) + e_{tsd}$

Residual = teacher-specific deviation from school's predicted slope (var = σ^2_e)

L2: $\beta_{0sd} = \delta_{00d} + U_{0sd}$

$\beta_{1sd} = \delta_{10d} + U_{1sd}$

School Random Intercept and Slope = school-specific deviations from district's predicted intercept, slope ($\tau^2_{U_0}, \tau^2_{U_1}, \tau_{U_{01}}$)

L3: $\delta_{00d} = \gamma_{000} + V_{00d}$

$\delta_{10d} = \gamma_{100} + V_{10d}$

District Random Intercept and Change = district-specific deviations from fixed intercept, slope ($\tau^2_{V_{00}}, \tau^2_{V_{10}}, \tau_{V_{00,10}}$)

Fixed Intercept, Fixed WSx Slope

Composite equation (9 parameters):

$$y_{tsd} = (\gamma_{000} + V_{00d} + U_{0sd}) + (\gamma_{100} + V_{10d} + U_{1sd})(WSx_{tsd}) + e_{tsd}$$

More on Random Slopes in 3-Level Models

- Any L1 predictor can have a random slope over L2, L3, or over both levels at once, but I recommend working your way **UP the higher levels** for assessing random effects…
  - e.g., Does the effect of L1 teacher experience vary over L2 schools?
  - If so, does the effect of L1 experience vary over L3 districts, too? → Is there a **commonality** in L1 experience slopes from same L1 district?
- … because random effects of L1 pred at L3 but not L2 are possible but unlikely (e.g., L1 teacher experience slope is exactly the same across L2 schools)

- L2 predictors can also have random effects over L3
  - e.g., Does the effect of a L2 school predictor vary over L3 districts?
- L1, L2, and L1 by L2 cross-level interactions can all have random effects over L3, too, at least in theory
  - But tread carefully! The more random effects you have, the more likely you are to have convergence problems ("**G** matrix not positive definite")

Conditional 3-Level Model Specification

- Remember separating between- and within-cluster effects?
  <u>*Now there are **3** potential fixed effects for any **L1** predictor!*</u>
  - e.g., Effect of L1 teacher experience on L1 mean student achievement
  - **L1**: **Teachers** with more experience may have better mean student outcomes (*than other teachers in same school*)
  - **L2**: **Schools** with more experienced teachers may have better school mean student outcomes (*than other schools in the district*)
  - **L3**: **Districts** with more experienced teachers may have better district mean student outcomes (*than other districts*)

- <u>And **2** potential fixed effects for any **L2** predictor:</u>
  - e.g., Effect of L2 principal experience on L2 mean student achievement
  - **Level 2**: **Schools** with more experienced principals may have better school mean student outcomes (than other schools in the district)
  - **Level 3**: **Districts** with more experienced principals may have better district mean student outcomes (than other districts)

Option 1: Separate Level-Specific Effects Using **Variable-Centering**

- **L1 Teachers:** *Teacher experience relative to rest of school*
  - → $WSexp_{tsd}$ = Exp_{tsd} – $SchoolMeanExp_{sd}$
  - → Directly tests if L1 within-school effect ≠ 0?
  - → **Total** within-school effect of more experience ***than others in school*** ≠ 0?

- **L2 Schools:** *School mean experience relative to rest of district*
  - → $WDexp_{sd}$ = $SchoolMeanExp_{sd}$ – $DistrictMeanExp_{d}$
  - → Directly tests if L2 within-district effect ≠ 0?
  - → **Total** effect of more mean experience ***than other schools in district*** ≠ 0?

Btw, compute L3 mean over L1 directly (not as L3 mean of L2 means)

- **L3 Districts:** *District mean experience relative to all districts*
  - → $BDexp_{d}$ = $DistrictMeanExp_{d}$ – (whatever constant)
  - → Directly tests if L3 between-district effect ≠ 0?
  - → **Total** effect of more mean experience ***than other districts*** ≠ 0?

Option 1: Separate Level-Specific Effects Using **Variable-Centering**

Notation: t = L1 teacher, s = L2 school, d = L3 district
SM = school mean, DM = district mean, C = centering constant

L1: $y_{tsd} = \beta_{0sd} + \beta_{1sd}(Exp_{tsd} - SMexp_{sd}) + e_{tsd}$

L2: $\beta_{0sd} = \delta_{00d} + \delta_{01d}(SMexp_{sd} - DMexp_{d}) + U_{0sd}$

$\beta_{1sd} = \delta_{10d} + U_{1sd}$

Level 3: $\delta_{00d} = \gamma_{000} + \gamma_{001}(DMexp_{d} - C) + V_{00d}$

$\delta_{01d} = \gamma_{010} + V_{01d}$

$\delta_{10d} = \gamma_{100} + V_{10d}$

Fixed intercept, Between-district exp main effect

Within-district exp main effect

Within-school exp main effect

Option 2: Contextual Effects Per Level Using **Constant-Centering**

Only ok if L1 will have not have a random slope

- **L1 Teachers:** *Teacher experience (relative to constant)*
  → $L1exp_{tsd} = Exp_{tsd} - C_1$
  → Directly tests if within-school effect ≠ 0, but only if L2 and L3 are there too!
  → **Total** within-school effect of more experience ***than others in school*** ≠ 0?

- **L2 Schools:** *School mean experience (relative to constant)*
  → $L2exp_{sd} = SchoolMeanExp_{sd} - C_2$
  → Directly tests if within-school and within-district effects ≠, but only if L3 is there!
  → **Contextual** effect of more average exp ***than other schools in district*** ≠ 0?

- **L3 Districts:** *District mean experience (relative to constant)*
  → $BDexp_{d} = DistrictMeanExp_{sd} - C_3$
  → Directly tests if within-district and between-district effects ≠ ?
  → **Contextual** effect of more average experience ***than other districts*** ≠ 0?

Option 2: Contextual Effects Per Level Using **Constant-Centering**

Notation: *t* = L1 teacher, *s* = L2 school, *d* = L3 district
SM = school mean, DM = district mean, C = centering constants

Only ok if L1 will have not have a random slope

L1: $y_{tsd} = \beta_{0sd} + \beta_{1sd}(Exp_{tsd} - C_1) + e_{tsd}$

L2: $\beta_{0sd} = \delta_{00d} + \delta_{01d}(SMexp_{sd} - C_2) + U_{0sd}$

$\beta_{1sd} = \delta_{10d} + U_{1sd}$

Fixed intercept, Contextual between-district exp main effect

Level 3: $\delta_{00d} = \gamma_{000} + \gamma_{001}(DMexp_d - C_3) + V_{00d}$

$\delta_{01d} = \gamma_{010} + V_{01d}$

Contextual within-district exp main effect

$\delta_{10d} = \gamma_{100} + V_{10d}$

Within-school exp main effect

What does it mean to omit higher-level effects under each centering method?

- **Variable-Centering**: Omitting a fixed slope assumes that the effect at that level **does not exist** (= 0 like usual for no slope)
  - Remove L3 effect? Assume L3 between-district effect = 0
    - *L1 effect = within-school effect, L2 effect = within-district effect*
  - Then remove L2 effect? Assume L2 within-district effect = 0
    - *L1 effect = within-school effect*

- **Constant-Centering**: Omitting a fixed slope means the effect at that level **is equivalent to** the effect at **the level below**
  - Remove L3 effect? Assume L3 between-district = L2 within-district effect
    - *L1 effect = within-school effect, L2 effect = smushed L2-WD and L3-BD effects*
  - Then remove L2 effect? Assume L2 between-school effect = L1 effect
    - *L1 smushed = within-school, within-district, and between-district effects*
  - The same problems exist for cross-level interactions, too!

Pseudo-R^2 in Three-Level Models

- Although it may not work this neatly in real data, here is the logic for how each type of fixed slope should explain variance

- **Main effects** and purely **same-level interactions** are straightforward—they target their **own level**:
  - L1 main effects and L1 interactions → L1 residual variance
  - L2 main effects and L2 interactions → L2 random intercept variance
  - L3 main effects and L3 interactions → L3 random intercept variance

- For **cross-level interactions**, which variance gets explained **depends** on if **random slopes** are included at each level…
  - L3 * L1 → L3 random variance in that L1 slope if included, or L2 random variance in that L1 slope if included, or L1 residual otherwise
  - L3 * L2 → L3 random variance in that L2 slope if included, or L1 residual otherwise
  - L2 * L1 → L2 random variance in that L1 slope if included, or L1 residual otherwise

Summary: Three-Level Random Effects Models

- Estimating 3-level models requires no new concepts, but everything is an order of complexity higher:
  - **Partitioning variance** over 3 levels instead of 2 → **many possible ICCs**
  - **Random slope variance will come from the variance directly below:**
    - Level-2 random slope variance comes from level-1 residual
    - Level-3 random slope variance comes from level-2 random slope (or residual)
  - **Level-1 effects can be random over level 2, level 3, or both at once**
    - ICCs can be computed for level-1 slopes that are random over both level-2 and level-3 (assuming the L2 and L3 variance models match)
    - Smushing of level-1 fixed effects should be tested over levels 2 AND 3
  - **Level-2 effects can be random over level 3**
    - Smushing of level-2 fixed effects should be tested over level 3
  - **Level-3 effects cannot be random**; no worries about smushing
  - **Pseudo-R^2** follows similar patterns as for two-level models
  - Phew….!