Level-1 Predictors (in General Multilevel Models for Two-Level Nested Data)

- Topics:
  - Fixed slopes of level-1 person predictors
    - #1 rule of MLM club: No smushing allowed!
    - Level-1 within, level-2 between, and level-2 contextual effects
  - Model specification methods
    - Cluster-mean-centering (= observed-variable-centering)
    - Grand-mean-centering (= constant-centering) + cluster mean
    - Latent centering (= latent-variable-centering)
  - Complications and alternatives

MLMs for Clustered Data: Review

- Multilevel models (MLMs) are used to quantify and predict how much of an outcome's total variation is due to each dimension of sampling
- Empty means, two-level model for level-1 person $\boldsymbol{p}$ in level-2 cluster $\boldsymbol{c}$:

Level 1: $\boldsymbol{y_{pc} = \beta_{0c} + e_{pc}}$

Level 2: $\boldsymbol{\beta_{0c} = \gamma_{00} + U_{0c}}$

$\boldsymbol{\gamma_{00}}$ = fixed intercept (mean of cluster means)
$\boldsymbol{U_{0c}}$ = level-2 random intercept (with variance $\boldsymbol{\tau^2_{U_0}}$)
$\boldsymbol{e_{pc}}$ = level-1 residual (with variance $\boldsymbol{\sigma^2_e}$)

- **Total** outcome variation is partitioned into **two uncorrelated sources**:
  - **Level-2 between**-cluster (BC) mean differences → random intercept $\boldsymbol{\tau^2_{U_0}}$
  - **Level-1 within**-cluster (WC) cluster differences → residual $\boldsymbol{\sigma^2_e}$
  - Dependency effect size via Intraclass Correlation 1: $\boldsymbol{ICC_1 = \tau^2_{U_0} / (\tau^2_{U_0} + \sigma^2_e)}$
    - ICC_1 = proportion of total variance due to **cluster mean differences**
    - ICC_1 = average correlation of persons from same cluster

- Fixed slopes of level-2 predictors explain cluster mean differences, thereby reducing the level-2 random intercept variance $\boldsymbol{\tau^2_{U_0}}$

Level-1 Predictors: What **Not** to Do!

- Level-2 predictors ($\boldsymbol{L2x_c}$ below) are **cluster** characteristics
- Level-1 predictors ($\boldsymbol{L1x_{pc}}$ below) are **person** characteristics
  - *What if we added a L1 predictor directly (as we did before at L2)?*

Level 1: $\boldsymbol{y_{pc} = \beta_{0c} + \beta_{1c}(L1x_{pc} - C_1) + e_{pc}}$

Level 2: $\boldsymbol{\beta_{0c} = \gamma_{00} + \gamma_{01}(L2x_c - C_2) + U_{0c}}$

$\boldsymbol{\beta_{1c} = \gamma_{10}}$

$\boldsymbol{\gamma_{00}}$ = fixed intercept (at predictors=0)
$\boldsymbol{\gamma_{01}}$ = fixed slope of $L2x_c$
$\boldsymbol{\gamma_{10}}$ = fixed slope of $L1x_{pc}$
$\boldsymbol{U_{0c}}$ = level−2 random intercept
$\boldsymbol{e_{pc}}$ = level−1 residual
C_1 = level-1 centering constant
C_2 = level-2 centering constant

  - First subscript = which beta in level-1 model
    Second subscript = order of predictor in level-2 model

- All good, right? Many researchers mistakenly think so, but this model is **VERY LIKELY to be mis-specified**…
  - … For the **exact same reasons** we need MLM in the first place!

Level-1 (Person-Level) Predictors

- Modeling level-1 predictors is complicated (and often done incorrectly) because **each level-1 predictor is usually really 2 predictor variables** (each with their own slope), **not 1**

- Textbook example(s): Student Socioeconomic Status (SES)
  - Some **kids** have higher SES than others in their school:
    - **L1 WC variation in SES** *(represented directly as deviation from school mean)*
  - Some **schools** have more high-SES students than other schools:
    - **L2 BC variation in SES** *(represented as school mean or via external info)*

- Can quantify sources of predictor variance with an empty model ICC_1
  - ICC_1 = (L2 between variance) / (L2 between variance + L1 within variance)
  - $\boldsymbol{ICC_1 < 1?}$ L1 predictor has **WC** variation (so it *could* have a **L1 WC** slope)
  - $\boldsymbol{ICC_1 > 0?}$ L1 predictor has **BC** variation (so it *could* have a **L2 BC** slope)

Between- vs. Within-Cluster Effects

- Between- and within-cluster slopes in SAME direction
  - SES → Achievement in students
    - **WC: Kids with more money than other kids in their school may have greater achievement than other kids in their school (regardless of school mean SES)**
    - **BC: Schools with more money than other schools may have greater mean achievement than schools with less money**

- Between- and within-cluster slopes in OPPOSITE directions
  - Body mass → life expectancy in animals (Curran and Bauer, 2011)
    - **WC: Within a species, relatively bigger animals have shorter life expectancy (e.g., over-weight ducks die sooner than healthy-weight ducks)**
    - **BC: Larger species tend to have longer life expectancies than smaller species (e.g., whales live longer than cows, cows live longer than ducks)**

- L1 within-cluster and L2 between-cluster slopes usually differ
  - Why? Because variables have different **meanings** at each level!
  - Why? Because variables have different **scales** at each level!

What **Not** to Do: Smushed Effects!

Level-1: $y_{pc} = \beta_{0c} + \beta_{1c}(L1x_{pc} - C_1) + e_{pc}$

Level-2: $\beta_{0c} = \gamma_{00} + U_{0c}$

$\beta_{1c} = \gamma_{10}$

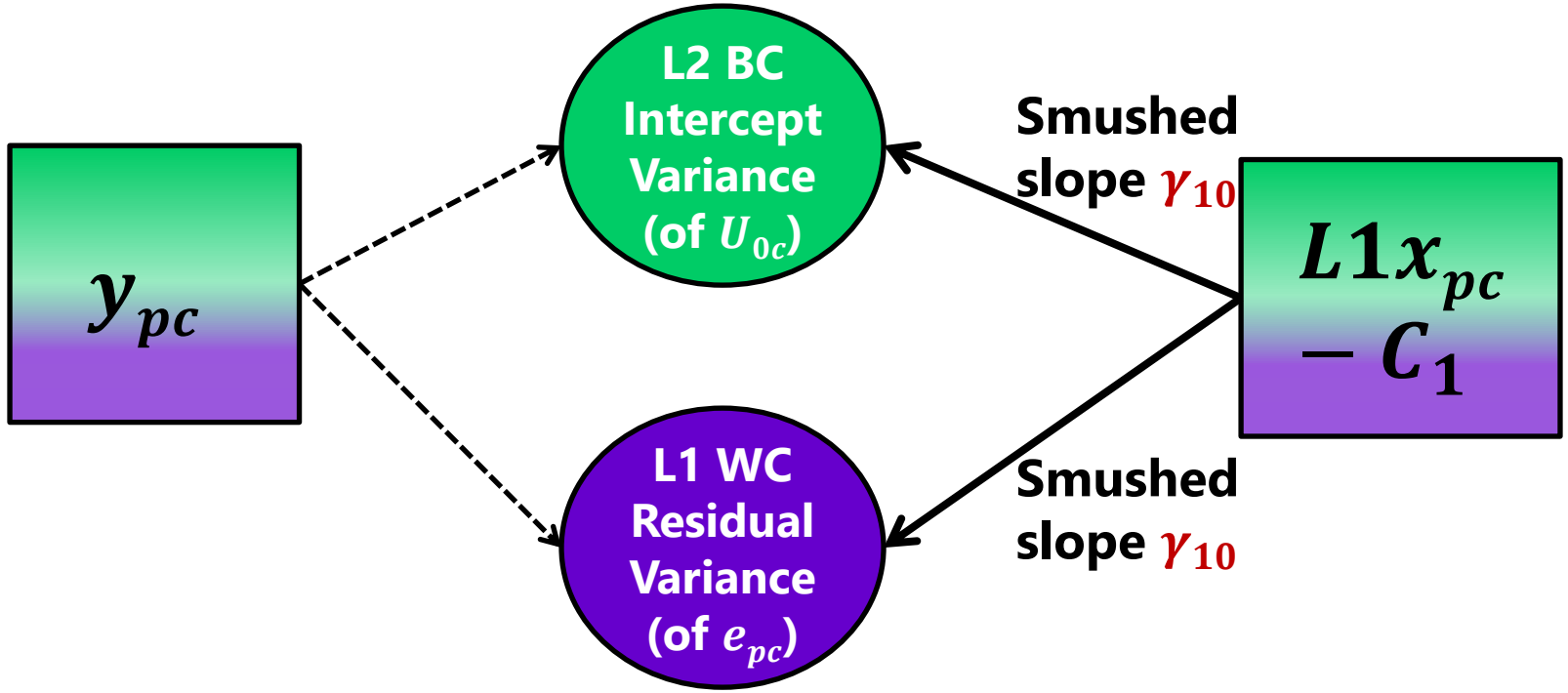
γ_{10} = **smushed effect** (see also *conflated*, *convergence*, or *composite* effect) that assumes equal within- and between-cluster slopes

- If level-1 predictor has both level-2 between and level-1 within variation, then its **one fixed slope has to do the work of two predictors**!
- A **smushed effect** is a **weighted combination of the L1 within and L2 between slopes,** usually **much closer to the L1 within slope** (due to larger $L1n$), and thus the L2 between model will be more affected by smushing
- Btw, **smushing** is seen in econometrics (aka, the "**endogeneity**" problem) in the context of when to model cluster dependency using fixed effects (i.e., turn cluster ID into a categorical predictor) instead of a random intercept (see Hamaker & Muthén, 2020)
  - A **smushed effect creates a correlation** between the L1 predictor and the L2 random intercept (because the **predictor's L2 effect is modeled wrong**)
  - Smushing for a level-1 main effect is solved when using **fixed effects for cluster ID,** such that the L2 effect of the L1 predictor is then controlled for in "common" variance
  - But we can still **avoid smushed effects** when using a cluster random intercept….

Univariate MLM: Adding a Level-1 Predictor Without Addressing Level-2 Part = Smushing

BC and WC variance in the **observed level-1 y_{pc} outcome** is partitioned by the **model** into estimated **variance components**

Observed level-1 $L1x_{pc}$ predictor still has both BC and WC variance. AND given that $L1x_{pc}$ has only **one fixed slope**, it captures a smushed effect that presumes **equal L2 BC and L1 WC slopes** in predicting y_{pc}!



Anticipating the Coefficient for the Smushed Effect of a Level-1 Predictor

$$\text{Smushed Effect: } \gamma_{\text{smushed}} \approx \frac{\frac{\gamma_{\text{BC}}}{\text{SE}_{\text{BC}}^2} + \frac{\gamma_{\text{WC}}}{\text{SE}_{\text{WC}}^2}}{\frac{1}{\text{SE}_{\text{BC}}^2} + \frac{1}{\text{SE}_{\text{WC}}^2}}$$

Adapted from Raudenbush & Bryk (2002, p. 138)

- **It is the rule, not the exception, that between-cluster and within-cluster effects differ** (Snijders & Bosker, 2012, p. 60, and personal experience!)
- **The smushed effect will often be closer to the level-1 within-cluster effect** (due to larger level-1 sample size and corresponding smaller SE), and thus the level-2 between-cluster model will be much more affected by smushing
- But **recovery of level-2 between cluster effects** will be less certain with **fewer level-2 units and a lower ICC_1** (for cluster mean differences) in the level-1 predictor!

3 Kinds of Fixed Slopes for Level-1 Predictors

- **Is there a Level-1 Within-Cluster (WC) slope?**
  - If you have a higher $\boldsymbol{L1x_{pc}}$ predictor value *than others in your cluster*, do you also have a higher (or lower) $\boldsymbol{y_{pc}}$ 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 ($\boldsymbol{\sigma_e^2}$) of the $\boldsymbol{y_{pc}}$ outcome

- **Is there a Level-2 Between-Cluster (BC) slope?**
  - Do clusters with higher average $\boldsymbol{L1x_{pc}}$ predictor values *than other clusters* also have higher (or lower) average $\boldsymbol{y_{pc}}$ outcomes *than other clusters*?
  - If so, the **level-2 between-cluster *part* of the level-1 predictor** will reduce level-2 random intercept variance ($\boldsymbol{\tau_{U_0}^2}$) of the $\boldsymbol{y_{pc}}$ outcome

- **Is there a Level-2 Contextual slope: Do the Level-2 BC and Level-1 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 $\boldsymbol{L1x_{pc}}$ 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, I call this "**variable-centering**" because you are **subtracting a variable** (e.g., the cluster mean here; could use other higher-unit variables as needed)
     - 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 the L1 predictor's level-2 mean to distinguish level-specific slopes
     - More generally, I call this "**constant-centering**" because you are **subtracting a constant** while still keeping all levels 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)

Option 1. Cluster-Mean-Centering (C-MC)

- We partition the L1 predictor $\boldsymbol{L1x_{pc}}$ into two variables that directly represent its **L2 between**-cluster (BC) and **L1 within**-cluster (WC) sources of variation, and **include these variables as the predictors:**

- **Level-1 Within predictor = deviation from cluster mean of $\boldsymbol{L1x_{pc}}$**
  - $\boldsymbol{WCx_{pc} = L1x_{pc} - \overline{L1x_c}}$ *(uncentered cluster mean $\overline{L1x_c}$ is used)*
  - WCx_{pc} is NOT centered at a constant – **we subtract a VARIABLE instead**
  - WCx_{pc} is positive? Above your cluster mean → "more than my cluster"
  - WCx_{pc} is negative? Below your cluster mean → "less than my cluster"

- **Level-2 Between predictor uses cluster mean of $\boldsymbol{L1x_{pc}}$ $(= \overline{\boldsymbol{L1x_c}})$**
  - $\boldsymbol{CMx_c = \overline{L1x_c} - C_2}$
  - CMx_c is centered at constant C_2, chosen for meaningful 0 (e.g., sample mean)
  - CMx_c is positive? Above sample mean → "more than other clusters"
  - CMx_c is negative? Below sample mean → "less than other clusters"

Cluster-MC L1 Predictor + Cluster Mean

→ WC and BC effects directly through **separate** parameters

$L1x_{pc}$ is cluster-mean-centered into WCx_{pc}, with CMx_c at L2:

Level-1: $y_{pc} = \beta_{0c} + \beta_{1c}(WCx_{pc}) + e_{pc}$

$WCx_{pc} = L1x_{pc} - \overline{L1x}_c$ → only has L1 within variation

Level-2: $\beta_{0c} = \gamma_{00} + \gamma_{01}(CMx_c) + U_{0c}$

$\beta_{1c} = \gamma_{10}$

$CMx_c = \overline{L1x}_c - C_2$ → only has L2 between variation

γ_{10} = within effect of having more $L1x_{pc}$ *than others in your cluster*

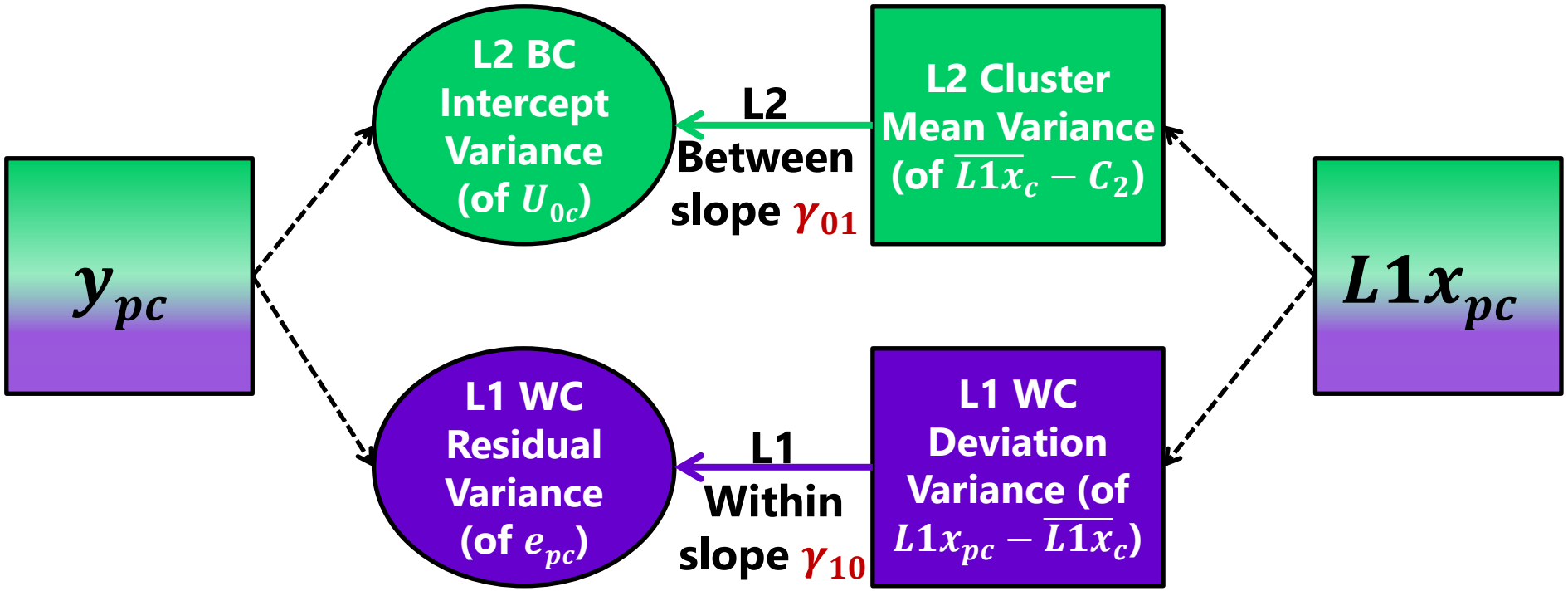
γ_{01} = between effect of having more $\overline{L1x}_c$ *than other clusters*

Because WCx_{pc} and CMx_c are uncorrelated, each gets the total effect for its level (L1 = within, L2 = between)

Univariate MLM: Cluster-Mean-Centering

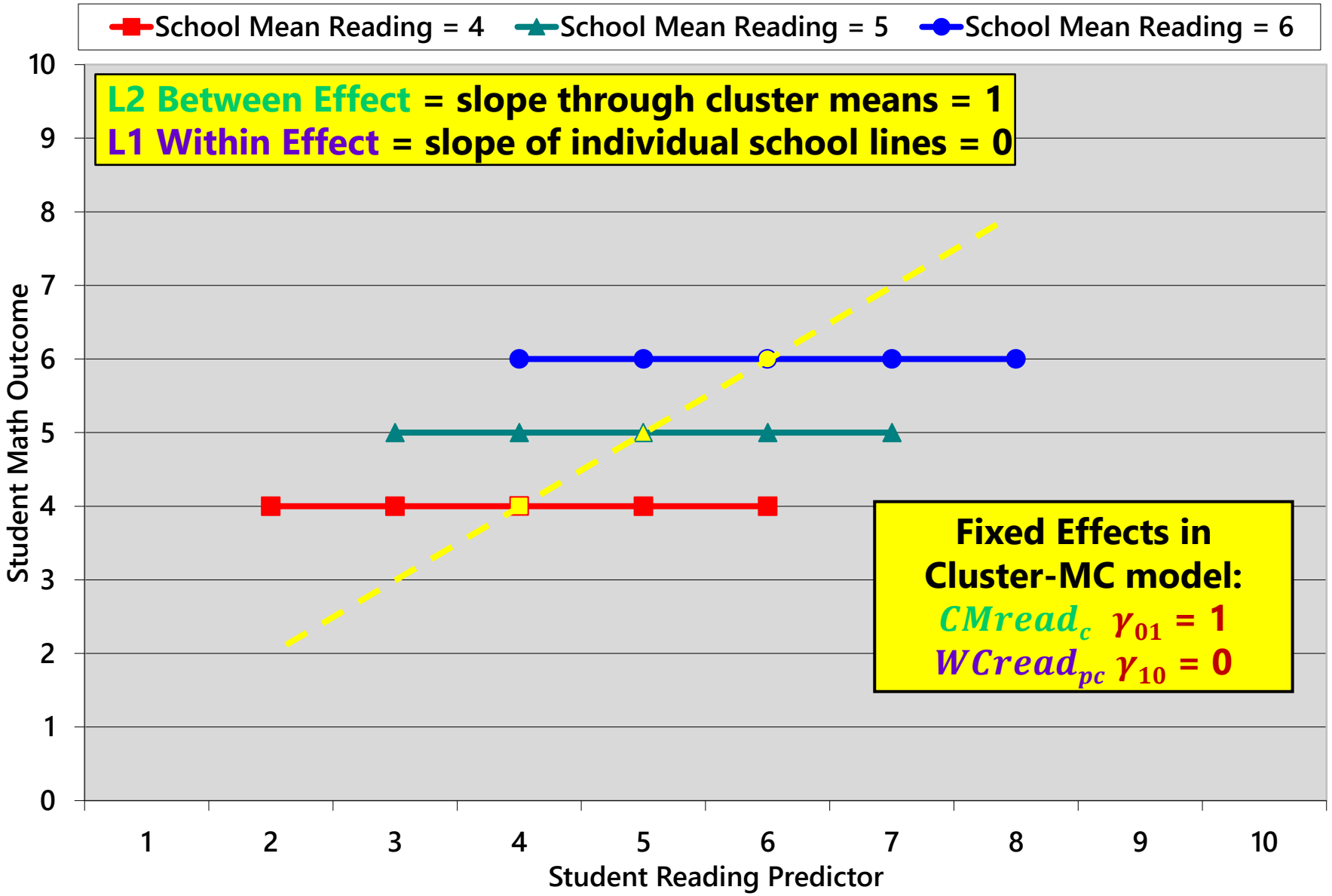
Model-based partitioning of level-1 y_{pc} outcome into level-specific **latent variables**

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

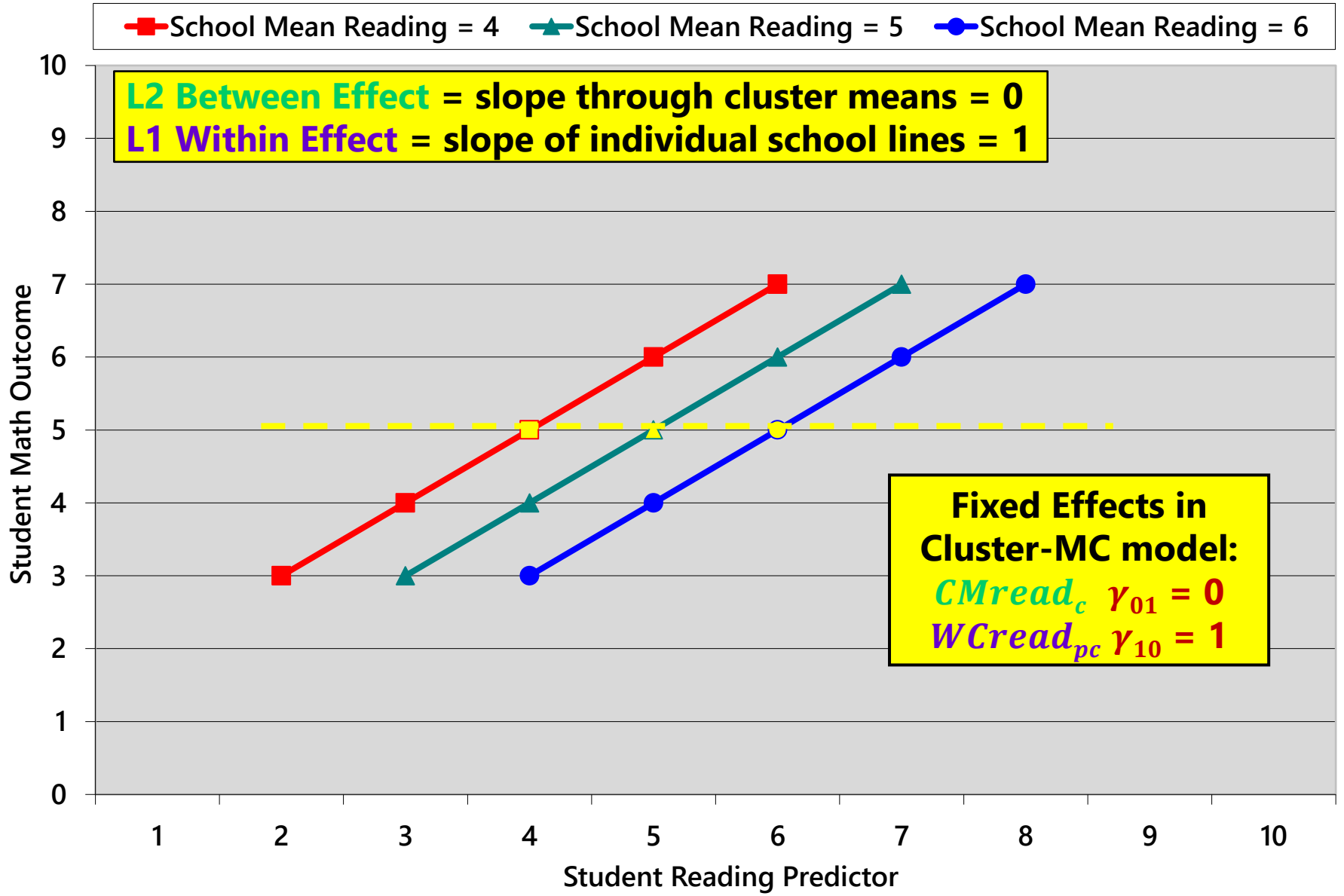


Why not let the model make variance components for $L1x_{pc}$, too? That is option 3, multivariate MLM (or "multilevel SEM"): stay tuned…

ALL Between Effect, **NO** Within Effect



NO Between Effect, **ALL** Within Effect



Example: Cluster-MC Level-1 Predictor

Example from Snijders & Bosker (2012) ch. 9: Predicting language outcomes for 3,566 students (p) from 191 schools (c) → **adding student verbal ability**

Level 1: $Lang_{pc} = \beta_{0c} + \beta_{1c}(Verbal_{pc} - \overline{Verbal_c}) + e_{pc}$

Level 2: $\beta_{0c} = \gamma_{00} + \gamma_{01}(HW_c - 2) + \gamma_{02}(MixGrd_c) + \gamma_{03}(\overline{Verbal_c} - 10) + U_{0c}$

$\beta_{1c} = \gamma_{10}$

```
print("Add L2-CM and Cluster-MC-L1 Student Verbal")
Model5 = lmer(data=Lecture3, REML=TRUE,
              formula=langpost~1+hw2+mixgrd+CMverb10+WCverb+(1|schoolID))
lmerSummary(Model5)
con1DSummary(contest1D(Model5, L=c(0,0,0,1,-1)), name="L2 Contextual")
PseudoR2(baseModel=Model2, altModel1=Model5)
```

L1 WCverb = $Verbal_{pc} - \overline{Verbal_c}$
L2 CMverb10 = $\overline{Verbal_c} - 10$

```
Fixed Effects:
 Term          Est    SE       DF t.value p.value
 Intercept  41.579 0.362  172.098 114.721   <.001
 hw2        -0.053 0.459  178.450  -0.115   0.909
 mixgrd     -1.121 0.516  197.041  -2.173   0.031
 CMverb10    3.660 0.271  207.383  13.508   <.001
 WCverb      2.423 0.057 3372.684  42.370   <.001
```

Btw, L2 Contextual = 1.237, SE = 0.277, $p < .0001$

```
Pseudo-R2
 Term       Model  Variance Pseudo.R2
 Intercept  Model2   17.815
 Intercept  Model5    8.396     0.529
 Residual   Model2   62.229
 Residual   Model5   40.550     0.348
```

Example: Cluster-MC Level-1 Predictor

Model for the Means (relevant new parameters only):

- γ_{00} = **41.58** = fixed **intercept**: expected language for students in a school with homework=2 (~mean), mixgrd=0 (= not mixed), and school mean verbal=10; for student verbal=10
- γ_{03} = **3.66*** = fixed **BC slope** of **school verbal**: difference in **school mean** language per unit higher mean verbal ability *than other schools*
- γ_{10} = **2.42*** = fixed **WC slope** of **student verbal**: difference in **student** language per unit higher verbal ability *than their school mean*

Model for the Variance:

- U_{0c} = level-2 random intercept → remaining deviation from fixed-effect-predicted school mean language for school c (with variance $\tau^2_{U_0} = 8.39$)
  - Pseudo-$R^2_{U0} = \frac{17.815-8.396}{17.815}$ = **.529 → 52.9%** explained (of original 22.3% L2 BC)
- e_{pc} = level-1 residual = deviation of the observed outcome for student p from their outcome predicted by β_{0c} and β_{1c} (with variance $\sigma^2_e = 40.55$)
  - Pseudo-$R^2_e = \frac{62.229-40.550}{62.229}$ = **.348 → 34.8%** explained (of original 77.7% L1 WC)

L1 Predictors: Intermediate Summary

- **2 kinds of slopes Cluster-Mean-Centering tells us *directly*:**

- **Is there a Level-1 Within-Cluster (WC) slope?**
  - If you have higher predictor values than the rest of your cluster, do you also have higher outcomes values than the rest of your cluster, such that the within-cluster deviation of the L1 predictor accounts for L1 residual variance ($\boldsymbol{\sigma_e^2}$)?
  - This is all that the L1 part of the predictor should logically be able to tell us!
  - **Given directly by fixed slope of $\boldsymbol{WCx_{pc}}$ regardless of whether $\boldsymbol{CMx_c}$ is there**
  - Note: L1 WC slope multiplies the **relative** value of $\boldsymbol{L1x_{pc}}$, NOT the **original** $\boldsymbol{L1x_{pc}}$

- **Is there a Level-2 Between-Cluster (BC) slope?**
  - Do clusters with higher predictor values than other clusters (*on average*) also have higher outcomes than other clusters (*on average*), such that the cluster mean of the L1 predictor accounts for level-2 random intercept variance ($\boldsymbol{\tau_{U_0}^2}$)?
  - **Given directly by fixed slope of $\boldsymbol{CMx_c}$ regardless of whether $\boldsymbol{WCx_{pc}}$ is there**
  - Note: L2 BC slope is NOT controlling for the original $\boldsymbol{L1x_{pc}}$ for each person

Formative Assessment Questions

1. In a two-level model for persons nested in clusters: What is the purpose of level-2 predictors, and what is the purpose of level-1 predictors? Which kind of variance should each kind of predictor explain?

2. "No smushing allowed" is the lecture's #1 rule of MLM club (slide 1). In your own words, explain why a raw level-1 predictor almost always contains two distinct sources of variance, and why giving it a single fixed slope creates a "smushed" effect.

3. Describe how to use cluster-mean-centering (or variable-centering more generally) for a level-1 predictor to directly estimate its within-cluster and between-cluster effects. How does one interpret each slope? Do these interpretations change if either predictor is not included?

4. Give an example of within-cluster and between-cluster effects from your research, work, or other experiences. Can you think of an example in which they'd be of different magnitudes or even in opposite directions?

3rd Kind of Slope for Level-1 Predictors

- **What Cluster-Mean-Centering DOES NOT tell us *directly*:**

- **Is there a Level-2 Contextual effect: Do the BC and WC slopes differ?**
  - After controlling for the original L1 predictor per person, is there still **an incremental contribution from having a higher cluster mean** of the L1 predictor (i.e., does a cluster's general tendency for the predictor explain more $\tau^2_{U_0}$ above and beyond just the person-specific value of the L1 predictor)?
  - If you **moved a student (and their characteristics) to a new school**, how would their outcome change?
  - If there is no contextual effect, then the L1 predictor's **L2 BC** and **L1 WC** slopes show **convergence**, which means their effects are of equivalent magnitude

- **To answer this question about the Level-2 Contextual effect for the incremental contribution of the cluster mean, we have two options:**
  - Still use Cluster-MC, and ask for the **contextual slope = between − within** (via SAS ESTIMATE, R contest1D or glht, SPSS TEST, STATA LINCOM, Mplus NEW…)
  - Use **"grand-mean-centering" (aka "constant-centering")** for the L1 predictor: $L1x_{pc} = L1x_{pc} - C_1$ → **centered at CONSTANT C_1, NOT A LEVEL-2 VARIABLE**
    - Which constant only matters for the reference point; it could be the grand mean or any (even 0)

Why the Difference in the Level-2 Effect? Remember Regular Old Regression…

- In this model: $y_p = \beta_0 + \beta_1(x1_p) + \beta_2(x2_p) + e_p$

Hang onto this idea…

- If $x1_p$ and $x2_p$ **ARE NOT** correlated:
  - β_1 carries **ALL the relationship** between $x1_p$ and y_p
  - β_2 carries **ALL the relationship** between $x2_p$ and y_p

- If $x1_p$ and $x2_p$ **ARE** correlated:
  - β_1 is **different than** the bivariate relationship between $x1_i$ and y_i
    - "Unique" effect of $x1_p$ *controlling for* $x2_p$ (i.e., *holding* $x2_p$ *constant)*
  - β_2 is **different than** the bivariate relationship between $x2_i$ and y_i
    - "Unique" effect of $x2_p$ *controlling for* $x1_p$ (i.e., *holding* $x1_p$ *constant)*

Reporting Descriptives and Correlations in MLM

- Just as we need to distinguish level-1-within from level-2-between cluster relations in fixed (and random!) slopes, we need to report **level-specific descriptive statistics and correlations using cluster-MC** (and ICC_1 for *every* level-1 variable!)
- Example from Snijders & Bosker (2012) ch. 9: 3,566 students from 191 schools
  - Language (outcome) $ICC_1 = .223$; verbal (predictor) $ICC_1 = .120$
  - School- and student-level correlations differ for level-2 school means (.72 vs .68)
  - Raw level-1 correlation between language and verbal is smushed (slightly, in this case)

School-Level-2 Correlations

Variable	hw	mixgrd	CMverb	CMlang
hw	1.00	-.14	.10	.09
mixgrd	-.14	1.00	-.16	-.23
CMverb	.10	-.16	1.00	.72
CMlang	.09	-.23	**.72**	1.00

What happens when the level-1 and level-2 predictors are correlated???

Student-Level-1 Correlations

Variable	verbal	langpost	WCverb	WClang	CMverb	CMlang
verbal	1.00	.60	.92	.54	.40	.27
langpost	**.60**	1.00	.51	.87	.33	.49
WCverb	.92	.51	1.00	.59	.00	.00
WClang	.54	.87	**.59**	1.00	.00	.00
CMverb	.40	.33	.00	.00	1.00	.68
CMlang	.27	.49	.00	.00	**.68**	1.00

Cluster-Mean-Centering vs. Grand-Mean-Centering (Variable- vs. Constant-Centering) for Level-1 Predictors

Level 2		Original	Cluster-MC Level 1	Grand-MC Level 1
$\overline{L1x_c}$	$CMx_c = \overline{L1x_c} - 5$	$L1x_{pc}$	$WCx_{pc} = L1x_{pc} - \overline{L1x_c}$	$L1x_{pc} = L1x_{pc} - 5$
3	−2	2	−1	−3
3	−2	4	1	−1
7	2	6	−1	1
7	2	8	1	3

Same L2 CMx_c goes into the model given either way of centering the L1 predictor $L1x_{pc}$

In **Cluster-MC** (**variable-centering**), the level-2 BC mean variation is gone from WCx_{pc}, so it is NOT CORRELATED with CMx_c

In **grand-MC** (**constant-centering)**, the level-2 BC mean variation is still inside $L1x_{pc}$, so it IS STILL CORRELATED with CMx_c

So the effects of CMx_c and $L1x_{pc}$ when included together under grand-MC (constant-centering) will be different than if either predictor were included by itself…

Option 2. Level-1 Predictor + Cluster Mean
→ Model tests contextual = difference of WC vs. BC effects

$L1x_{pc}$ is constant-centered, but <u>WITH</u> CMx_c at Level 2:

Level 1: $y_{pc} = \beta_{0c} + \beta_{1c}(L1x_{pc}) + e_{pc}$

$L1x_{pc} = L1x_{pc} - C_1$ → still has both L2 between and L1 within variation

Level 2: $\beta_{0c} = \gamma_{00} + \gamma_{01}(CMx_c) + U_{0c}$

$\beta_{1c} = \gamma_{10}$

$CMx_c = \overline{L1x_c} - C_2$ → only has L2 between variation

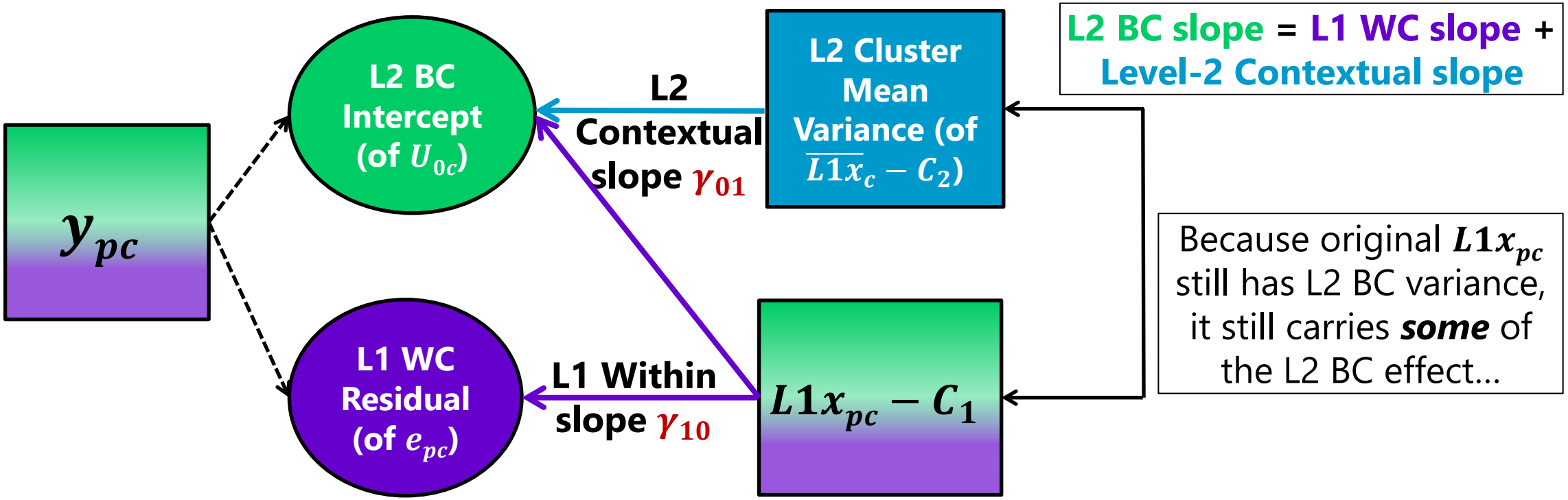
γ_{10} becomes the within effect → *unique* L1 effect after controlling for L2 CMx_c

γ_{01} *becomes* the L2 Contextual slope that indicates <u>how the L2 BC effect differs from the L1 WC effect</u>
→ *unique* level-2 slope after controlling for $L1x_{pc}$
→ does cluster mean matter beyond person value?
→ outcome difference if a person moved to a new cluster (but otherwise was the same person)

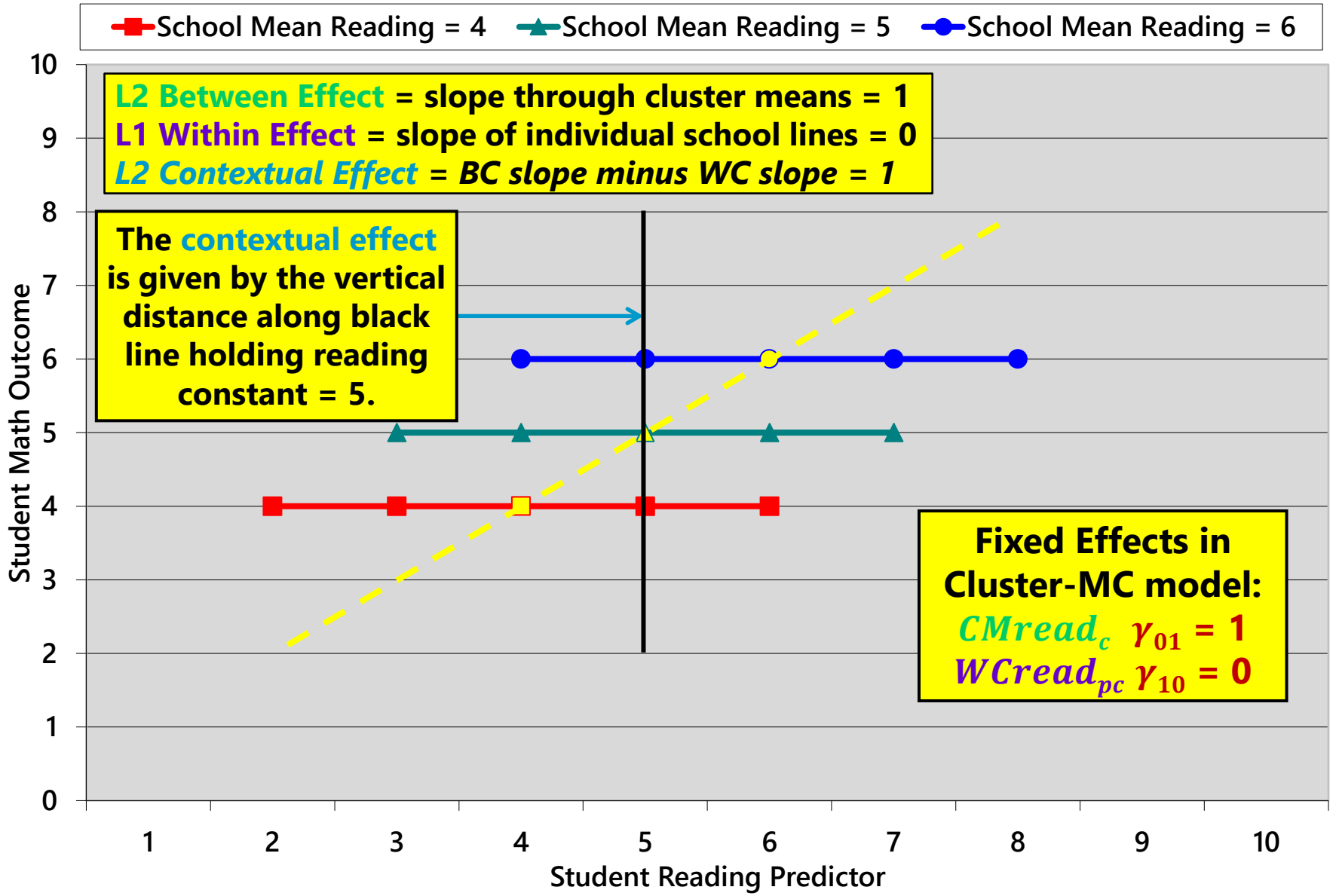
Constant-Centering + Cluster Mean

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

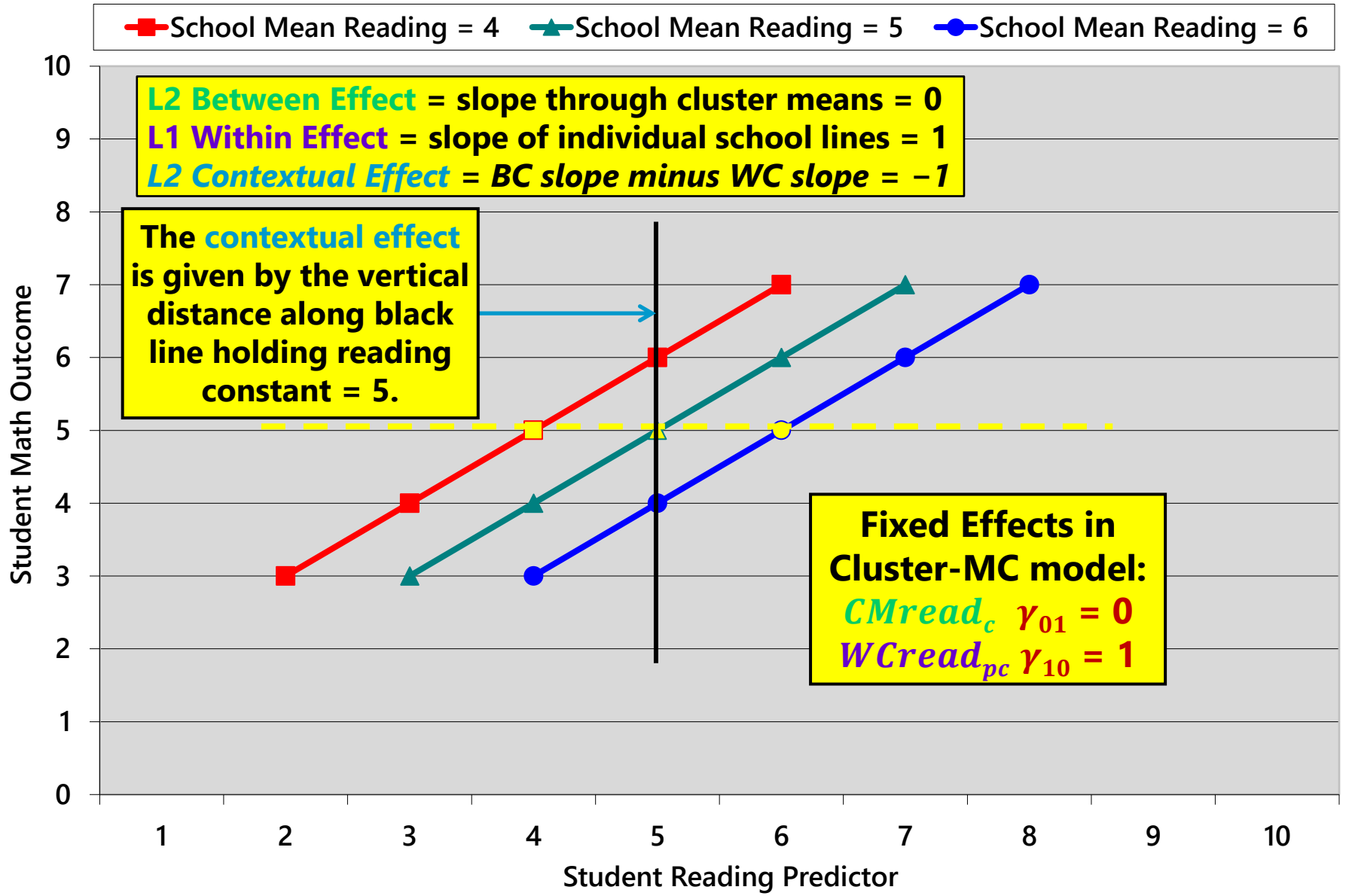
$L1x_{pc}$ **is still NOT partitioned**, but cluster mean $\overline{L1x_c} - C_2$ is added to allow an **incremental "contextual" L2 effect**



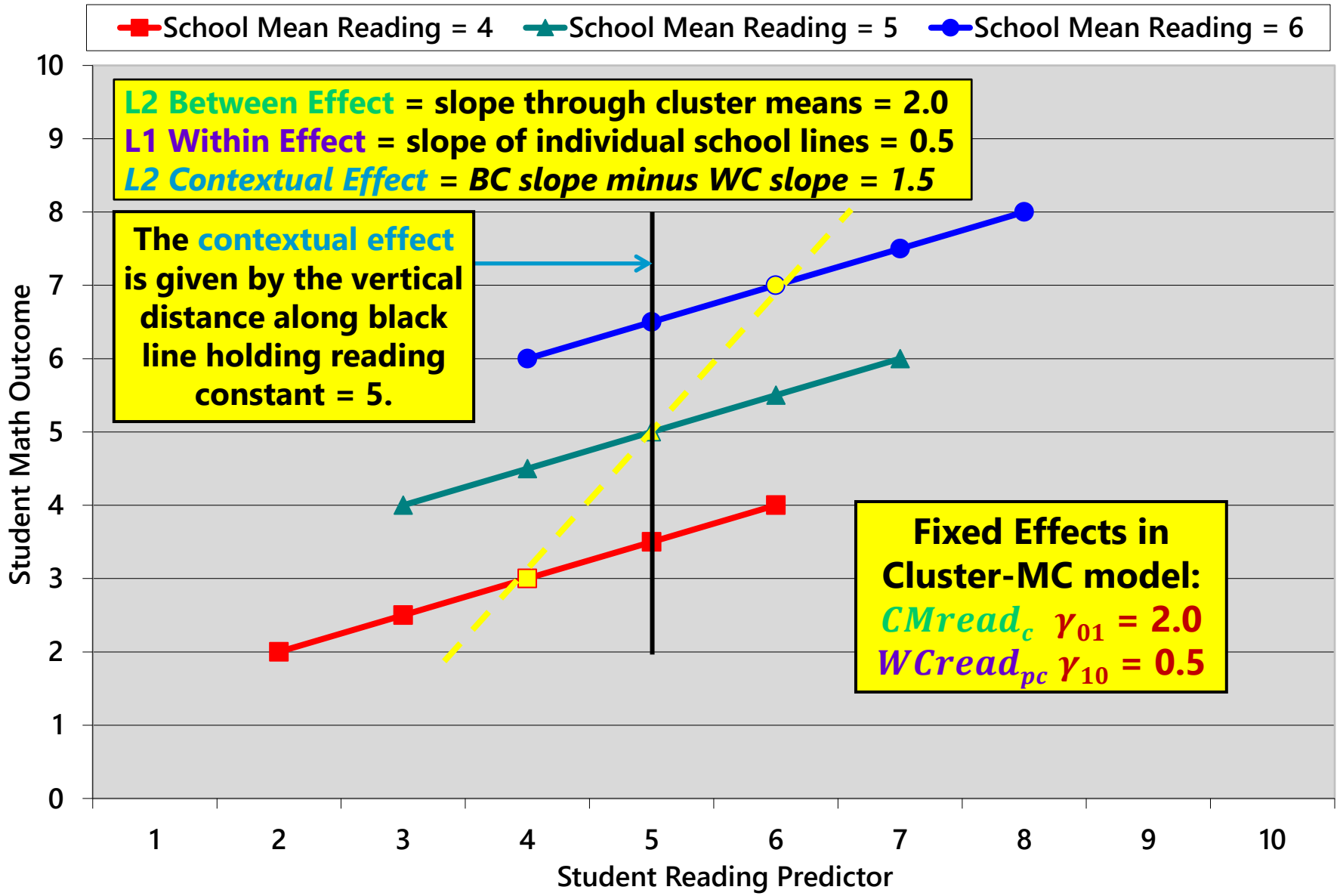
ALL Between Effect, **NO** Within Effect



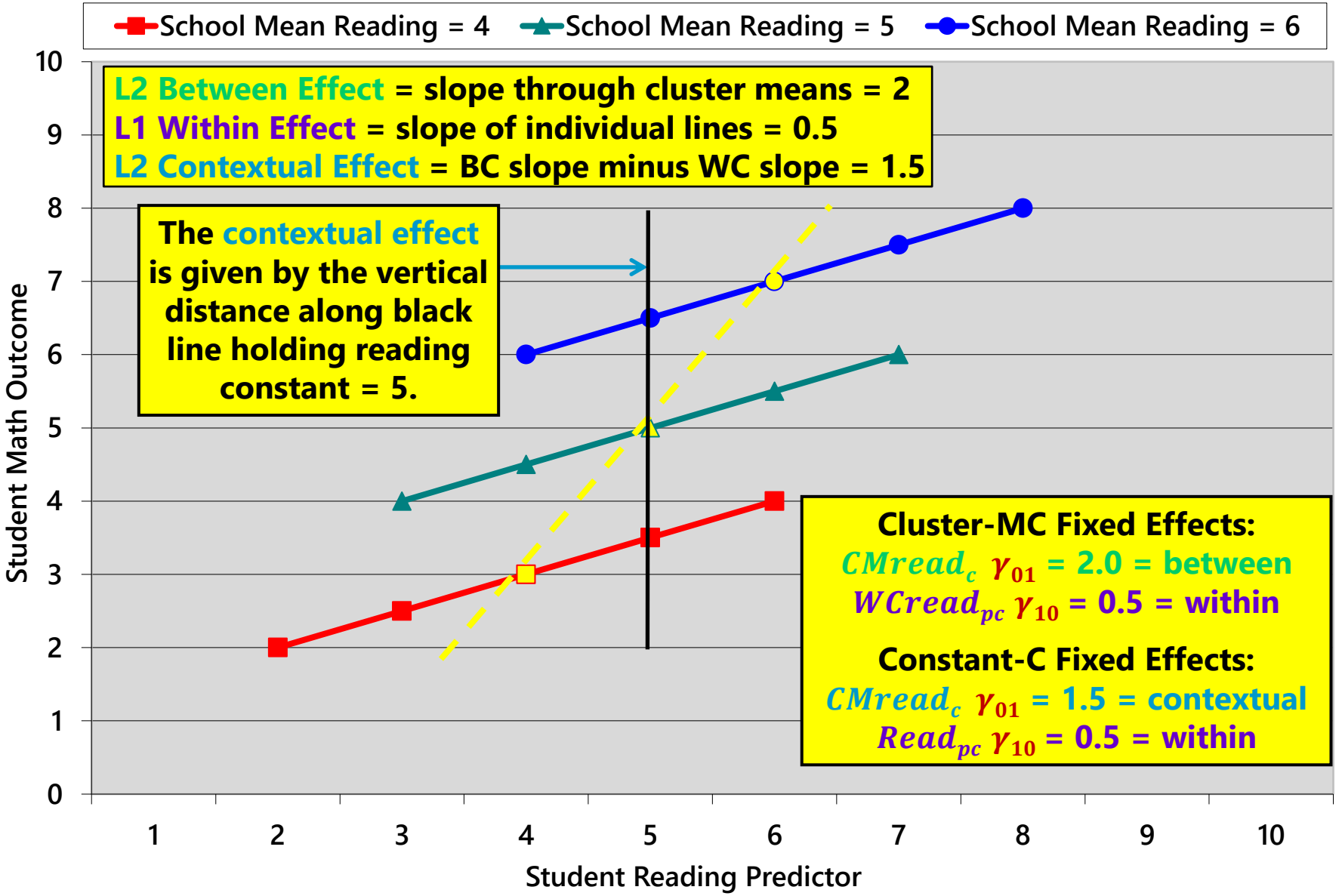
NO Between Effect, **ALL** Within Effect



Between Effect > Within Effect



Between, Within, and Contextual Effects



Example: Grand-MC Level-1 Predictor

Level 1: $Lang_{pc} = \beta_{0c} + \beta_{1c}(Verbal_{pc} - 10) + e_{pc}$

Level 2: $\beta_{0c} = \gamma_{00} + \gamma_{01}(HW_c - 2) + \gamma_{02}(MixGrd_c) + \gamma_{03}(\overline{Verbal_c} - 10) + U_{0c}$

$\beta_{1c} = \gamma_{10}$

Constant-C from above:
L1 verb10 = $Verbal_{pc} - 10$ (**differs**)
L2 CMverb10 = $\overline{Verbal_c} - 10$ (**same**)

Compared to Cluster-MC from before:
L1 WCverb = $Verbal_{pc} - \overline{Verbal_c}$ (**differs**)
L2 CMverb10 = $\overline{Verbal_c} - 10$ (**same**)

Fixed Effects:

Term	Est	SE	DF	t.value	p.value
Intercept	41.579	0.362	172.098	114.721	<.001
hw2	-0.053	0.459	178.450	-0.115	0.909
mixgrd	-1.121	0.516	197.041	-2.173	0.031
CMverb10	1.237	0.277	226.180	4.468	<.001
verb10	2.423	0.057	3372.684	42.370	<.001

Btw, L2 Between = 3.660, SE = 0.271, $p < .0001$

← **L2?** →
L1 WC

Fixed Effects:

Term	Est	SE	DF	t.value	p.value
Intercept	41.579	0.362	172.098	114.721	<.001
hw2	-0.053	0.459	178.450	-0.115	0.909
mixgrd	-1.121	0.516	197.041	-2.173	0.031
CMverb10	3.660	0.271	207.383	13.508	<.001
WCverb	2.423	0.057	3372.684	42.370	<.001

Btw, L2 Contextual = 1.237, SE = 0.277, $p < .0001$

- **L2 Contextual** slope = **1.237** using constant-C L1 (or **Contextual** = **Between** − **Within**)
- **L2 Between** slope = **3.660** using cluster-MC L1 (or **Between** = **Contextual** + **Within**)
- Btw, the **smushed** slope would have been **2.472** = Within (close) = Between (too small)!

Cluster-MC and Grand-MC Models are Equivalent Given a Fixed Level-1 Main Effect Slope Only

Cluster-MC: $WCx_{pc} = L1x_{pc} - CMx_c$

Level-1: $y_{pc} = \beta_{0c} + \beta_{1c}(L1x_{pc} - CMx_c) + e_{pc}$

Level-2: $\beta_{0c} = \gamma_{00} + \gamma_{01}(CMx_c) + U_{0c}$

$\beta_{1c} = \gamma_{10}$

$\rightarrow y_{pc} = \gamma_{00} + \gamma_{01}(CMx_c) + \gamma_{10}(L1x_{pc} - CMx_c) + U_{0c} + e_{pc}$

$\rightarrow y_{pc} = \gamma_{00} + (\gamma_{01} - \gamma_{10})(CMx_c) + \gamma_{10}(L1x_{pc}) + U_{0c} + e_{pc}$

Grand-MC:

Level-1: $y_{pc} = \beta_{0c} + \beta_{1c}(L1x_{pc}) + e_{pc}$

Level-2: $\beta_{0c} = \gamma_{00} + \gamma_{01}(CMx_c) + U_{0c}$

$\beta_{1c} = \gamma_{10}$

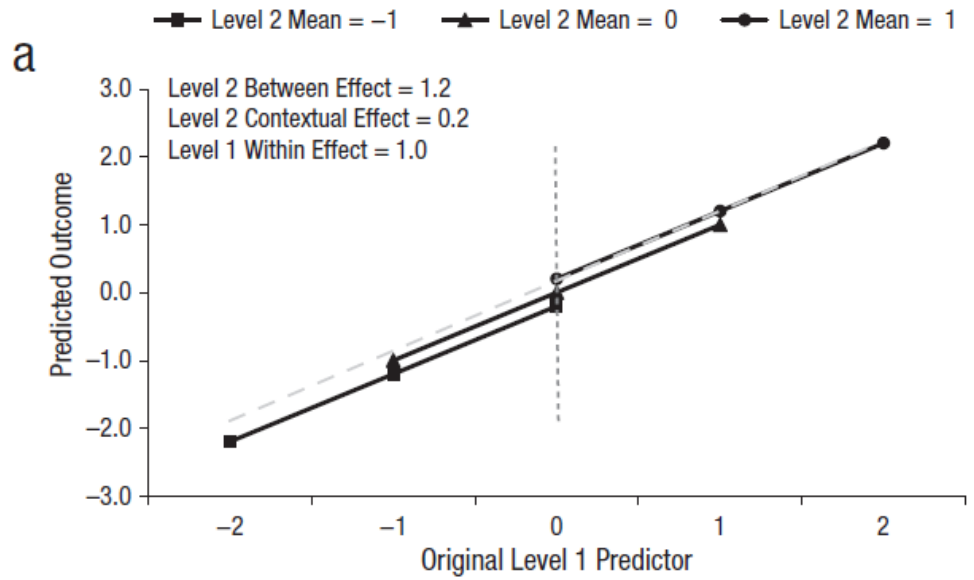
$\rightarrow y_{pc} = \gamma_{00} + \gamma_{01}(CMx_c) + \gamma_{10}(L1x_{pc}) + U_{0c} + e_{pc}$

Btw, I am using a centering constant = 0 at both levels to simplify the notation so that $\overline{L1x_c} = CMx_c$

Composite Model:
← As Cluster-MC
← As Grand-MC

Effect	Cluster-MC	Grand-MC
Intercept	γ_{00}	γ_{00}
WC Effect	γ_{10}	γ_{10}
Contextual	$\gamma_{01} - \gamma_{10}$	γ_{01}
BC Effect	γ_{01}	$\gamma_{01} + \gamma_{10}$

More on Between vs. Contextual Effects



- Image from Hoffman (2019) example using student SES

- *Top:* Contextual effect is minimal—there is no added benefit to going to a high-SES school when comparing across schools *at same level of student SES* (L1 predictor)

Option 3. Latent-Centering (C-MC)

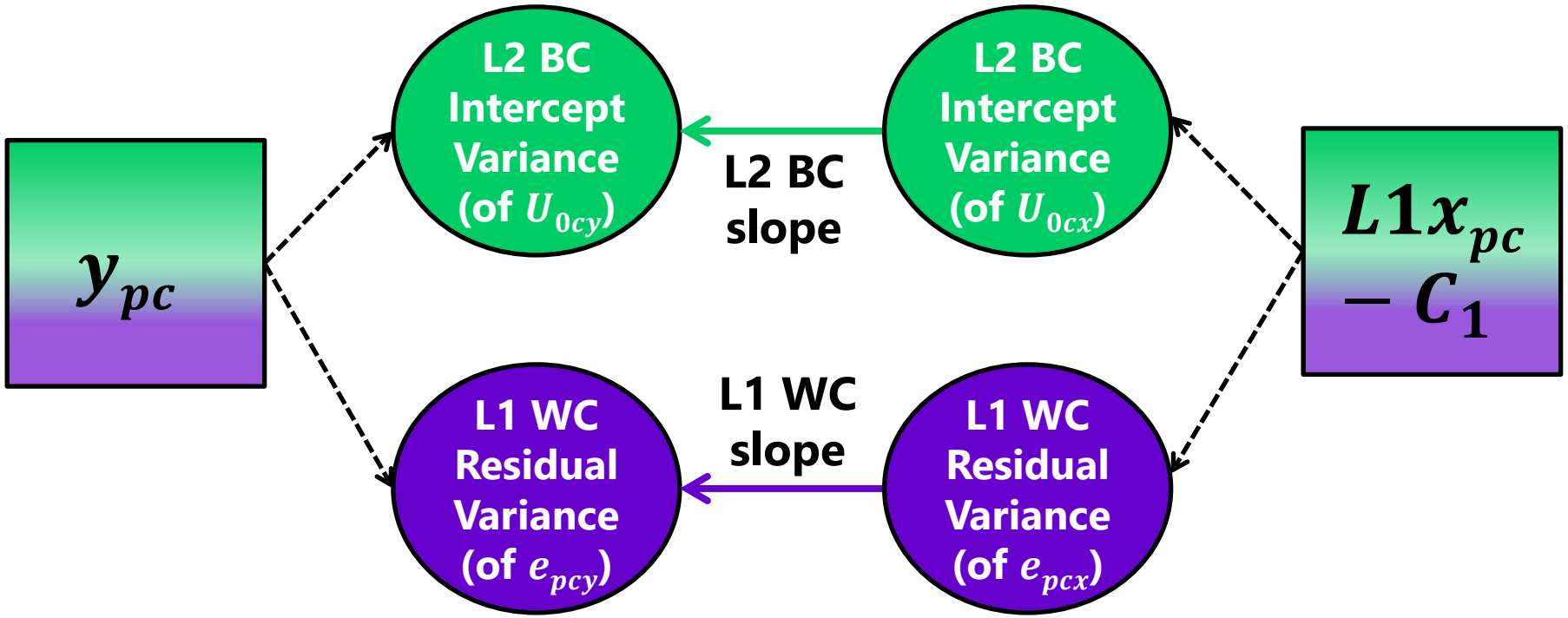
- We let the model partition the L1 predictor $\boldsymbol{L1x_{pc}}$ into two latent variables that directly represent its **L2 between**-cluster (BC) and **L1 within**-cluster (WC) sources of variation, just as we did for y_{pc}!
  - At a minimum: Fit an empty means, random intercept model for $L1x_{pc}$ (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, $\boldsymbol{\overline{L1x_c} - C_2}$, as in cluster-MC)*
  - Level-1 WC differences are represented by L1 residual for $L1x_{pc}$ *(instead of observed cluster mean deviation, $\boldsymbol{L1x_{pc} - \overline{L1x_c}}$, as in cluster-MC)*

- Requires multivariate software that can predict more than one column (either single-level SEM or multilevel-SEM, aka M-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)
  - Next 2 slides have a quick example for now (see 2023 class unit 8 for more)

Option 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**



Some 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)

Mplus M-SEM: Latent Centering of L1 Verbal

```
TITLE:   Model2a: Latent Centering of Student Verbal Ability Predicting Language
         Specifying L1 effect in WITHIN model directly
DATA:    FILE = ExampleData.csv;  ! Can just list file if syntax in same folder
         TYPE = INDIVIDUAL; FORMAT = FREE;   ! Defaults
VARIABLE:
! List of ALL variables in stacked data file, in order (up to 8 characters)
  NAMES = schoolID studID lang verbal homework mixgrd;
! List of ALL variables used in model (DEFINED variables go last)
  USEVARIABLES = lang mixgrd hw2 verb10;
! Missing data codes (here, -999)
  MISSING = ALL (-999);
! Identify level-2 ID
  CLUSTER = schoolID;
! Predictor variables with variation ONLY at level 1 -- none here
  WITHIN = ;
! Predictor variables with variation ONLY at level 2 (DEFINED last)
  BETWEEN = mixgrd hw2;

DEFINE:      hw2 = homework-2;         ! Center L2 homework at 2
             verb10 = verbal - 10;     ! Center L1 verbal at 10
ANALYSIS:    TYPE = TWOLEVEL RANDOM;   ! 2-level model with random slopes
             ESTIMATOR = ML;           ! Can also use MLR for non-normality
MODEL:
! Level-1, Within-Cluster (WC) Model
%WITHIN%
  lang;                        ! L1 residual variance in lang
  verb10;                      ! L1 residual variance in verbal (new)
  lang ON verb10 (within);     ! NO Placeholder, L1 Within verbal -> lang

! Level-2, Between-Cluster Model
%BETWEEN%
 [lang];                       ! Fixed intercept for lang
  lang;                        ! L2 random intercept variance in lang
 [verb10];                     ! Fixed intercept for verbal (new)
  verb10;                      ! L2 random intercept variance in verbal (new)
  lang ON hw2 mixgrd;          ! Between fixed slopes of L2 preds -> lang
  lang ON verb10 (between);    ! Between fixed slope of verbal -> lang

MODEL CONSTRAINT:              ! Linear combinations of fixed effects
 NEW(context);                 ! Name each new created fixed effect
 context = between - within;   ! L2 Contextual fixed slope of verbal -> lang
```

	Estimate	S.E.	P-Value
Within Level			
LANGPOST ON			
VERB10	2.425	0.057	0.000
Variances → NEW!			
VERB10	3.688	0.090	0.000
Residual Variances			
LANGPOST	40.536	0.987	0.000
Between Level			
LANGPOST ON			
HW2	-0.076	0.456	0.867
MIXGRD	-1.193	0.513	0.020
VERB10	4.239	0.421	0.000
Means → NEW!			
VERB10	-0.017	0.062	0.786
Intercepts			
LANGPOST	41.597	0.360	0.000
Variances → NEW!			
VERB10	0.510	0.080	0.000
Residual Variances			
LANGPOST	7.765	1.126	0.000
New/Additional Parameters			
CONTEXT	1.814	0.429	0.000

Relative to the cluster-MC univariate MLM (using REML estimation), in the latent-centered multivariate MLM (using ML estimation), the **L2 Between effect is larger** (4.24 vs. 3.66), a phenomenon known as "Lüdtke's bias"

I *Usually* Prefer Variable-Centering (using observed or latent variables)…

- **…because constant-centering is much easier to screw up! ☺**
- Table 1 below from: Hoffman & Walters (2022). Catching up on multilevel modeling

Table 1 Predictor effect type by model specification

Centering strategy for level-1 predictor (constant-centered level-2 predictor)	Fixed effect type by predictors included		
	Level-1 only	Level-2 only	Both levels
Variable-centered level-1			
Level-1 predictor: $L1x_{wb} = x_{wb} - \bar{x}_b$	Within	(= 0)	Within
Level-2 predictor: $L2x_b = \bar{x}_b - C_2$	(= 0)	Between	Between
Constant-centered level-1			
Level-1 predictor: $L1x_{wb} = x_{wb} - C_1$	Smushed	(= 0)	Within
Level-2 predictor: $L2x_{wb} = \bar{x}_b - C_2$	(= Within)	Between	Contextual

Abbreviations: w, within; b, between; C_1, level-1 centering constant; C_2, level-2 centering constant.
Parentheses indicate assumptions about the fixed slopes of omitted predictors.

Explained Variance by Fixed Slopes

- **Fixed slopes of level-2 cluster predictors *by themselves*:**
  - L2 BC main effects or interactions reduce L2 random intercept variance

- **Fixed slopes of level-1 person predictors *without L2 variance*:**
  - L1 WC main effects or interactions among L1 predictors reduce L1 residual variance, which can make L2 random intercept variance increase (→ smaller correction factor)
    - Remember: **True $\tau^2_{U_0}$ = Observed $\tau^2_{U_0}$ – ($\sigma^2_e/L1n$)**

- **Fixed slopes of level-1 person predictors *with L2 variance*:**
  - L1 part of L1 WC main effects reduce L1 residual variance
    - L2 part of L1 WC main effects *can* reduce L2 random intercept variance, OR it could increase it instead if the smushed effect is really wrong
    - Need to add corresponding L2 main effects order to prevent smushing!
  - For interactions among L1 predictors, interactions of corresponding L2 predictors are needed to un-smush the L1 interaction (stay tuned!)
    - See Hoffman & Walters (2022) and Hoffman (2019) for elaboration

Complication: Level-1 Interactions

- Interactions among L1 predictors can also be examined, with debate about the "right" order of operations when using cluster-MC L1 predictors (as $\boldsymbol{WCx_{pc} = L1x_{pc} - \overline{L1x_c}}$)
  - If the cluster-MC L1 predictors $\boldsymbol{WCx_{pc}}$ and $\boldsymbol{WCz_{pc}}$ are correlated and/or not normally distributed, then their product may still have L2 variance!

- Two possible methods to compute L1 interaction (Loeys et al., 2018), in which $\boldsymbol{WCx_{pc}}$ and $\boldsymbol{WCz_{pc}}$ are used for main effects either way
  - P1C2: (1) compute product of original → $\boldsymbol{L1x_{pc} * L1z_{pc} = L1xz_{pc}}$
    (2) center the L1 product using the L2 mean of the new L1 product
    $\boldsymbol{L1\ Interact = L1xz_{pc} - \overline{L1xz_c}}$
    - Shown to be unbiased and most efficient, but only in a model without any level-2 fixed effects (so it's not clear whether this result generalizes to more complex models)
  - C1P2: (1) center $L1x_{pc}$ and $L1z_{pc}$ using L2 means → $\boldsymbol{WCx_{pc} = L1x_{pc} - \overline{L1x_c}}$
    (2) compute product of centered versions → $\boldsymbol{WCx_{pc} * WCz_{pc}}$
    - Can be biased unless 3 additional cross-variable interactions are included: 1 for the L2 means ($\boldsymbol{\overline{L1x_c} * \overline{L1z_c}}$) and 2 "cross-level" interactions ($\boldsymbol{WCx_{pc} * \overline{L1z_c}}$) and ($\boldsymbol{WCz_{pc} * \overline{L1x_c}}$)

Constant-Centering for L1 predictors (+Cluster Mean!) may be preferable when:

- **You really do want level-2 contextual effects**
  - Directly model the incremental contribution of the cluster mean after controlling for a person's actual (not relative) predictor value (i.e., what would happen at a new school)
- **For *categorical* level-1 predictors** (see Yaremych et al., 2023)
  - e.g., 0/1 predictors when cluster-MC → impossible values
  - e.g., cluster mean of binary $L1x_{pc}$: $\overline{L1x}_c$ = 0.5? Then:
    $WCx_{pc} = L1x_{pc} - \overline{L1x}_c$ → WCx_{pc} = **+0.5 or −0.5 → weird**
- **When the cluster mean is not a reliable cluster-level predictor**
  - When the sample of persons within clusters is not complete enough to form a useful cluster mean, using externally-provided info may do a better job of representing the cluster (in which case cluster-MC doesn't really make sense without the cluster mean to go in with it)

- But cluster-MC or latent-centering is needed instead to prevent a L1 predictor's **random slope** from being smushed… stay tuned!

Do I *have* to use MLM?

- Although **MLMs** (which have at least a **random intercept** by definition) are an optimal way to model clustered data, they are not the only possibility
  - See McNeish (2023) on your reading list for an extended discussion!

- Another acceptable option is to **use fixed effects** to control for cluster mean differences: use Cluster ID as a categorical predictor *(i.e., include L2n – 1 dummy-coded predictors to saturate all cluster mean differences)*
  - Controls fixed effect standard errors for dependency due to cluster mean differences AND prevents smushed L1 main effects and L1 interactions
  - Better option than a random intercept in MLM for small L2 sample sizes
  - Analogously, you could also cluster-MC your outcome and all predictors (remove ALL cluster mean differences ahead of time) → single-level model

- A **partially acceptable** option is to use **cluster-corrected standard errors**
- Although they control fixed effect SEs for dependency due to cluster mean differences (instead of using random effects), **THEY DO NOT PREVENT SMUSHED EFFECTS!**
  - Thus, you **still need to use one of the three options** to do so shown here (cluster-mean-centering, grand-MC +cluster mean, or latent-centering)

Summary

- **Level-1 predictors** are person characteristics, but they **almost always contain cluster mean differences** (level-2 variance) as well
  - **Variance** at each level → **different prediction** at each level!
  - Yes, you need to care about the cluster variance in your L1 predictors!

- **3 options** for specifying fixed slopes of a L1 predictor in order to distinguish its level-specific effects (i.e., **avoid smushed effects**):
  1. **Cluster-Mean-Centering (variable-centering)**: Manually carve up L1 predictor into L2 BC (cluster mean → **L2 Between slope**) and L1 WC deviation (→ **L1 Within slope**)
  2. **Grand-Mean-Centering (constant-centering)**: Add cluster mean to become **L2 Contextual slope**, then L1 predictor's unique effect is **L1 Within slope**
  3. **Latent-Centering**: Let multivariate MLM estimate L2 and L1 variance components, same as for the outcome → analogous to Cluster-MC
     - Requires multivariate software (as in M*plus*, gsem in STATA, or OpenMx in R)

- Next up: **random slopes** and **cross-level interactions!**

Formative Assessment Questions

1. Describe the difference between centering using the cluster mean or just a constant for level-1 predictors. What type of variability does each type of level-1 predictor then contain?

2. When using constant-centering for a level-1 predictor, how does one interpret its fixed effect when included by itself? How does this effect change after including the level-2 cluster mean of the predictor? How does one then interpret the fixed effect for the level-2 cluster mean?

3. Give an example of how one would interpret the within, between, and contextual effects of a level-1 predictor from your own area of research (or an area you are otherwise interested in).

4. In reading results sections for multilevel modeling analyses, what are two indicators that a smushed version of a level-1 slope has been estimated?