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

- Topics:
  - Fixed slopes of level-1 person predictors
    - #1 rule of MLM: 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: $y_{pc} = \beta_{0c} + e_{pc}$

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

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

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

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

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

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

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

γ_{00} = fixed intercept (at pred=0)
γ_{01} = fixed slope of $L2x_c$
γ_{10} = fixed slope of $L1x_{pc}$
U_{0c} = level-2 random intercept
e_{pc} = level-1 residual

- 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 each source of variance with an empty model ICC
  - ICC = (L2 between variance) / (L2 between variance + L1 within variance)
  - **ICC < 1?** L1 predictor has **WC** variation (so it *could* have a **L1 WC** slope)
  - **ICC > 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: $\beta_{0c} = \gamma_{00} + U_{0c}$

$\beta_{1c} = \gamma_{10}$ ←

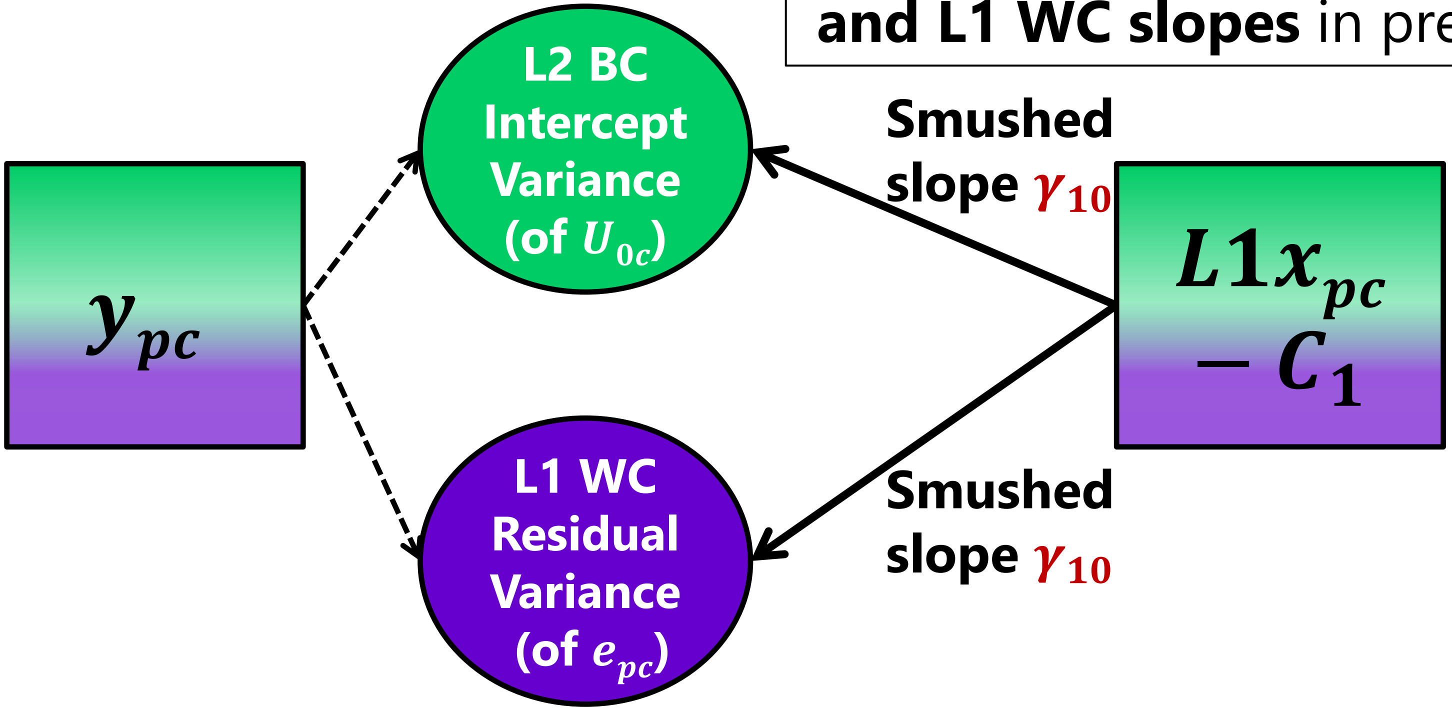
γ_{10} = **smushed effect** (see also *conflated*, *convergence*, or *composite* effect) that assumes equal within- and between-cluster slopes

- 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 **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, "**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
  - 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 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…. Next are the 3 main ways to do so!

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

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

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

- **The smushed effect will often be closer to the L1 within-cluster effect** (due to larger L1 sample size and corresponding smaller SE), and thus the L2 between-cluster model will be much more affected by smushing
- **It is the rule, not the exception, that between-cluster and within-cluster effects differ** (Snijders & Bosker, 2012, p. 60, and personal experience!)
- Btw, this same issue is known in the econometrics literature as the problem of "endogeneity" and is directly related to controversies of when one should use fixed instead of random effects to fully control for higher-level dependency → the use of fixed effects solves the problem of smushing (for main effects!)

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 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-2 Between predictor uses cluster mean of $L1x_{pc}$** $(=\overline{\boldsymbol{L1x}}_{\mathbf{c}})$
  - $\boldsymbol{CMx_c} = \overline{\boldsymbol{L1x}}_{\mathbf{c}} - \boldsymbol{C_2}$
  - 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"

- **Level-1 Within predictor = deviation from cluster mean of $L1x_{pc}$**
  - $\boldsymbol{WCx_{pc}} = \boldsymbol{L1x_{pc}} - \overline{\boldsymbol{L1x}}_{\boldsymbol{c}}$ *(uncentered cluster mean $\overline{\boldsymbol{L1x}}_{\boldsymbol{c}}$ 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"

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

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

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

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

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

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

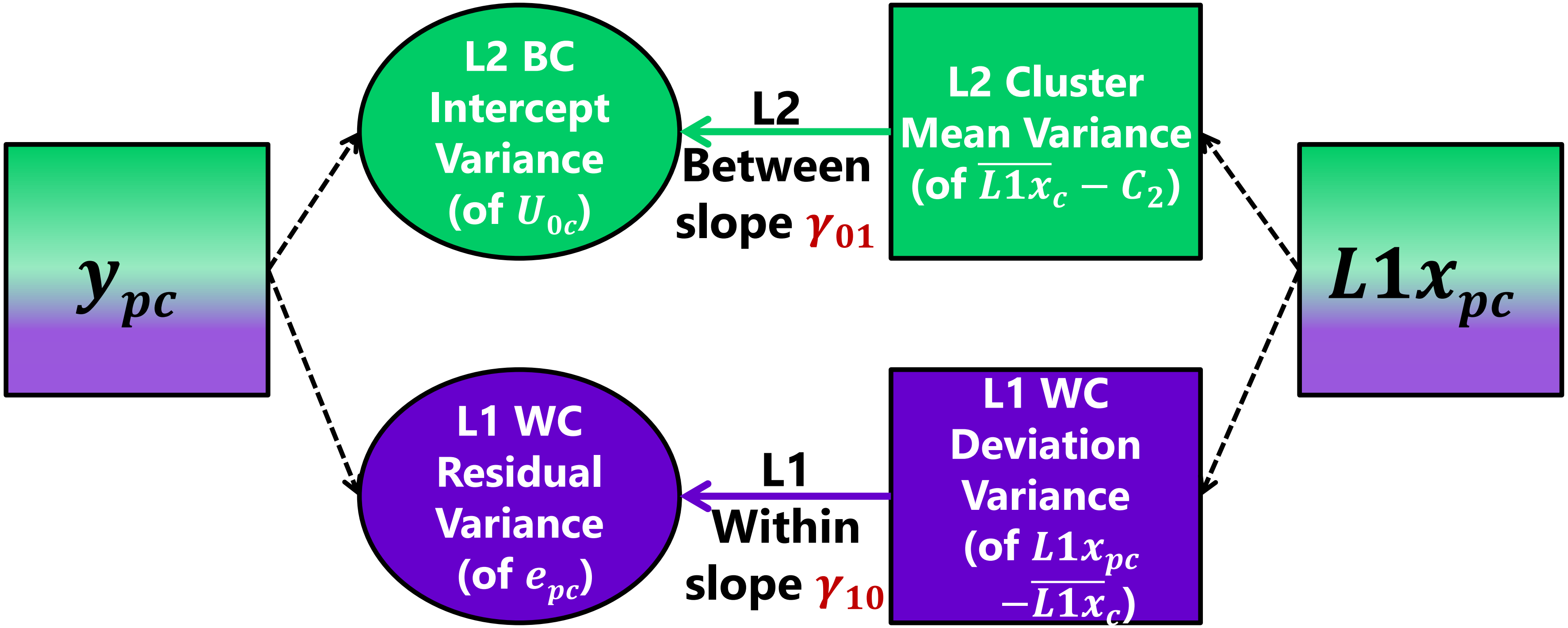
γ_{01} = between effect of having more *$\overline{L1x}_c$ than other clusters*

Because WCx_{pc} and CMx_c are uncorrelated, each gets the <u>total</u> 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**



y_{pc}

L2 BC Intercept Variance (of U_{0c})

L2 Between slope γ_{01}

L2 Cluster Mean Variance (of $\overline{L1x_c} - C_2$)

$L1x_{pc}$

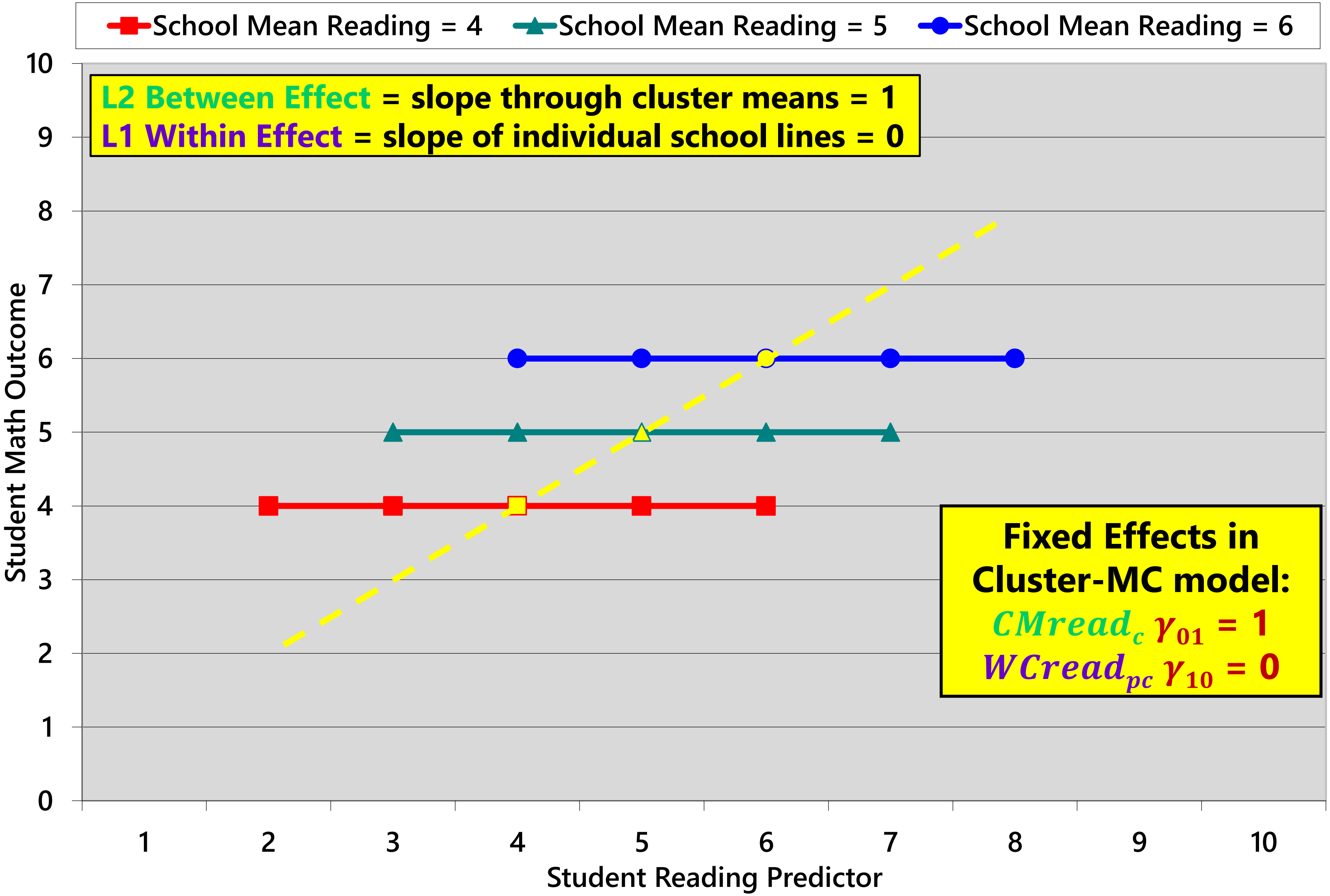
L1 WC Residual Variance (of e_{pc})

L1 Within slope γ_{10}

L1 WC Deviation Variance (of $L1x_{pc} - \overline{L1x_c}$)

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

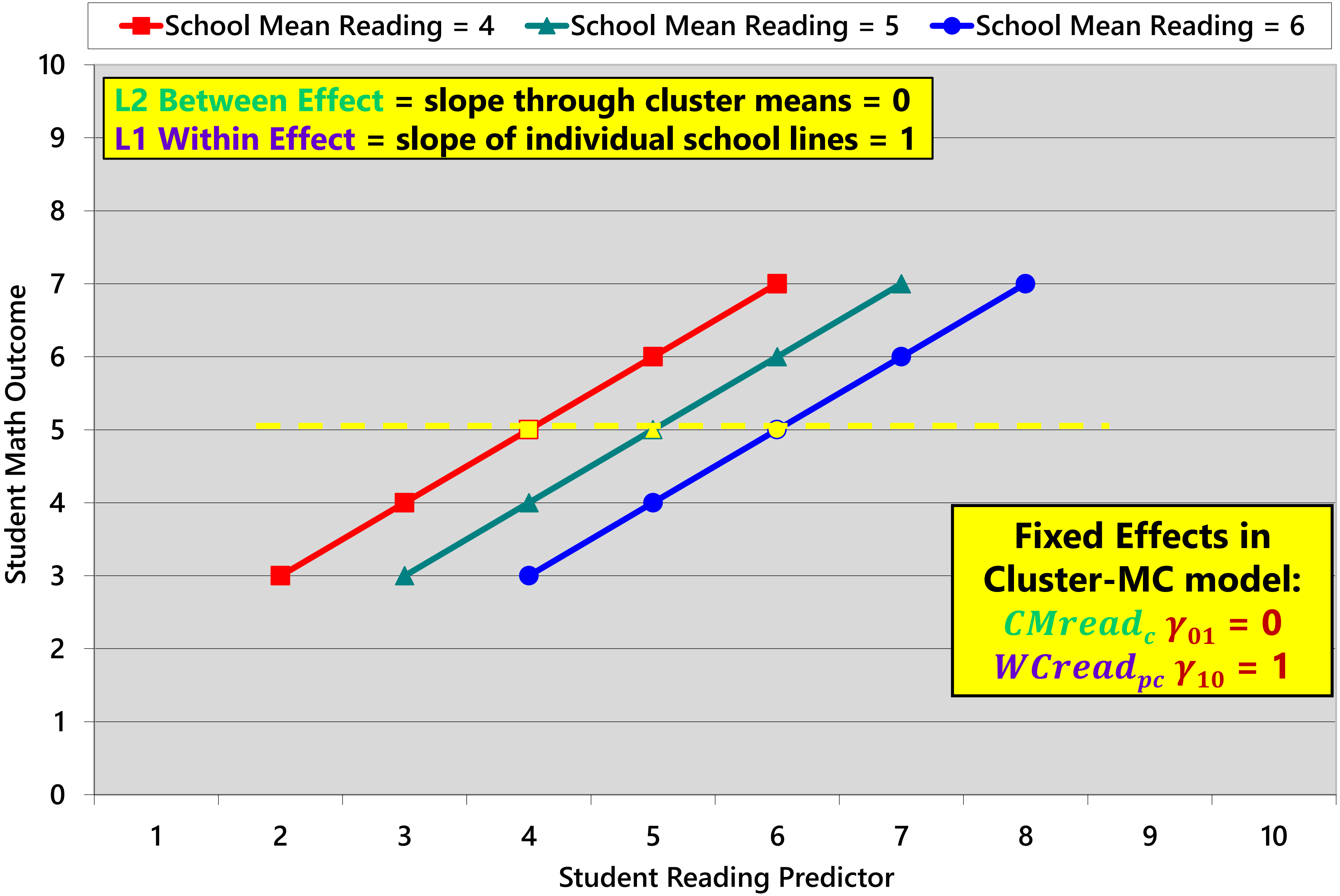
ALL Between Effect, **NO** Within Effect



L2 Between Effect = slope through cluster means = 1
L1 Within Effect = slope of individual school lines = 0

Fixed Effects in Cluster-MC model:
$CMread_c\ \gamma_{01} = 1$
$WCread_{pc}\ \gamma_{10} = 0$

NO Between Effect, **ALL** Within Effect



L2 Between Effect = slope through cluster means = 0
L1 Within Effect = slope of individual school lines = 1

Fixed Effects in Cluster-MC model:

$CMread_c\ \gamma_{01}$ = 0

$WCread_{pc}\ \gamma_{10}$ = 1

Adding L2 Between and L1 Within Predictors: (2a) Syntax by Univariate MLM Program

SAS:

```
PROC MIXED DATA=work.Example COVTEST NOCLPRINT IC METHOD=REML;
     CLASS schoolID;
     MODEL langpost = hw2 mixgrd CMverb10 WCverb / SOLUTION DDFM=Satterthwaite;
     RANDOM INTERCEPT / TYPE=UN SUBJECT=schoolID;
     ESTIMATE "L2 Contextual Effect of Verbal"  CMverb10 1 WCverb -1;
RUN;
```

R lmer from lme4 package—using lmerTest package to get Satterthwaite denominator DF and contest1D:

```
name = lmer(data=Example, REML=TRUE,
            formula=langpost~1+hw2+mixgrd+CMverb10+WCverb+(1+|schoolID))
summary(name, ddf="Satterthwaite")
contest1D(name, ddf="Satterthwaite", L=c(0,0,0,1,-1)) # L2 Contextual effect of verbal
```

STATA:

```
mixed langpost c.hw2 c.mixgrd c.CMverb10 c.WCverb, || schoolID:, ///
      reml dfmethod(satterthwaite) dftable(pvalue) nolog
lincom c.CMverb10*1 + c.WCverb*-1, small // L2 Contextual effect of verbal
```

SPSS:

```
MIXED langpost BY schoolID WITH hw2 mixgrd CMverb10 WCverb
      /METHOD    = REML
      /CRITERIA  = DFMETHOD(SATTERTHWAITE)
      /PRINT     = SOLUTION TESTCOV
      /FIXED     = hw2 mixgrd CMverb10 WCverb
      /RANDOM    = INTERCEPT | COVTYPE(UN) SUBJECT(schoolID)
      /TEST      = "L2 Contextual effect of verbal" CMverb10 1 WCverb -1.
```

Electronic materials for this example from my 2023 APA training sessions are **here**

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**

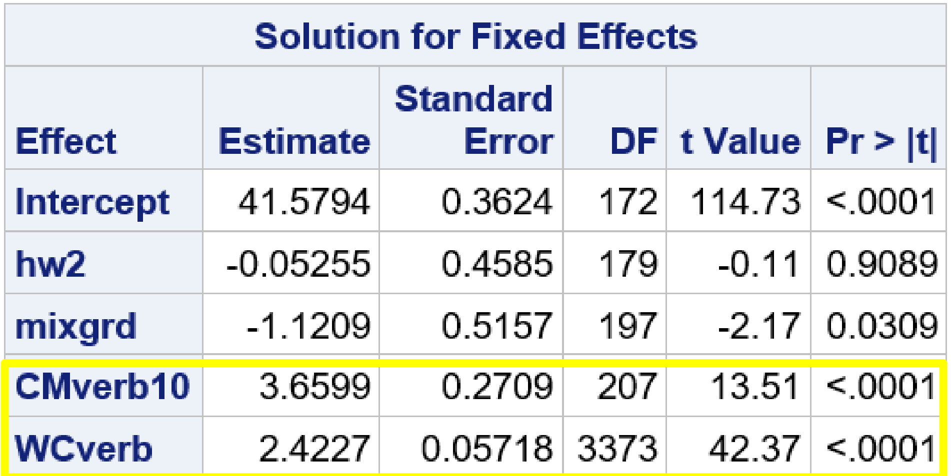
Level-1: $Lang_{pc} = \beta_{0c} + \beta_{1c}(Verbal_{pc} - \overline{Verbal}_c) + e_{pc}$

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

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

Results from SAS MIXED:

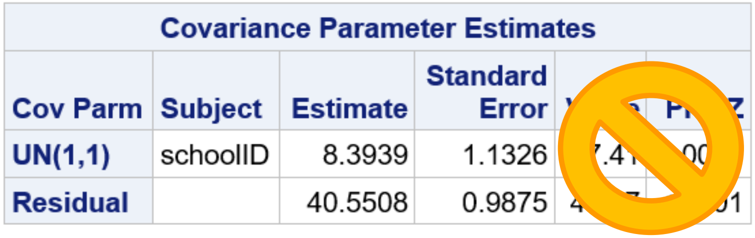
L1 WCverb = $Verbal_{pc} - \overline{Verbal}_c$
L2 CMverb10 = $\overline{Verbal}_c - 10$



Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	41.5794	0.3624	172	114.73	<.0001
hw2	-0.05255	0.4585	179	-0.11	0.9089
mixgrd	-1.1209	0.5157	197	-2.17	0.0309
CMverb10	3.6599	0.2709	207	13.51	<.0001
WCverb	2.4227	0.05718	3373	42.37	<.0001

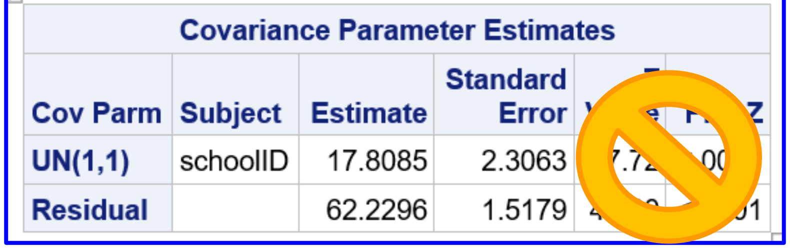
Btw, L2 Contextual = 1.237, SE = 0.277, p < .0001



Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	[illegible]	[illegible]
UN(1,1)	schoolID	8.3939	1.1326	[illegible]	[illegible]
Residual		40.5508	0.9875	[illegible]	[illegible]

From empty model to compare:



Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	[illegible]	[illegible]
UN(1,1)	schoolID	17.8085	2.3063	[illegible]	[illegible]
Residual		62.2296	1.5179	[illegible]	[illegible]

Example: Cluster-MC Level-1 Predictor

Model for the Means (relevant new parameters only):

- $\boldsymbol{\gamma_{00}}$ **=** **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 a student whose verbal = 10
- $\boldsymbol{\gamma_{03}}$ **=** **3.66*** = fixed **BC slope** of **school verbal**: difference in **school mean** language per unit higher mean verbal ability *than other schools*
- $\boldsymbol{\gamma_{10}}$ **=** **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:

- $\boldsymbol{U_{0c}}$ = level-2 random intercept = deviation of the original from predicted school mean language for school c (with variance $\boldsymbol{\tau^2_{U_0}}$ **= 8.39**), where "original" is from the empty means, random intercept model
  - Pseudo-$R^2_{U0} = \frac{\mathbf{17.809-8.394}}{\mathbf{17.809}}$ **= .529 → 52.9%** explained (of original 22.3% L2 BC)
- $\boldsymbol{e_{pc}}$ = level-1 residual = deviation of the observed outcome for student p from their outcome predicted by $\boldsymbol{\beta_{0c}}$ and $\boldsymbol{\beta_{1c}}$ (with variance $\boldsymbol{\sigma^2_e}$ **= 40.55**)
  - Pseudo-$R^2_e = \frac{\mathbf{62.230-40.551}}{\mathbf{62.230}}$ **= .348 → 34.8%** explained (of original 77.7% L1 WC)

3 Kinds of Fixed Slopes for L1 Predictors

- **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 <u>than the rest of your cluster</u>, do you also have higher outcomes values <u>than the rest of your cluster</u>, such that the within-cluster deviation of the L1 predictor accounts for L1 residual outcome 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 <u>than other clusters</u> (*on average*) also have higher outcomes <u>than other clusters</u> (*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

3rd Kind of Slope for L1 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 value of the 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 $\boldsymbol{\tau^2_{U_0}}$ above and beyond just the person-specific value of the L1 predictor)?
  - 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, SPSS TEST, STATA LINCOM, Mplus NEW…)
  - Use **"grand-mean-centering"** for the L1 predictor: $\boldsymbol{L1x_{pc} = L1x_{pc} - C_1}$
    → **centered at CONSTANT $\boldsymbol{C_1}$, 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: $y_p = \beta_0 + \beta_1(x1_p) + \beta_2(x2_p) + e_p$
- 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)*

- **Hang onto that idea…**

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

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

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

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

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

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

γ_{01} becomes the within effect → *unique* L1 effect after controlling for L2 CMx_c

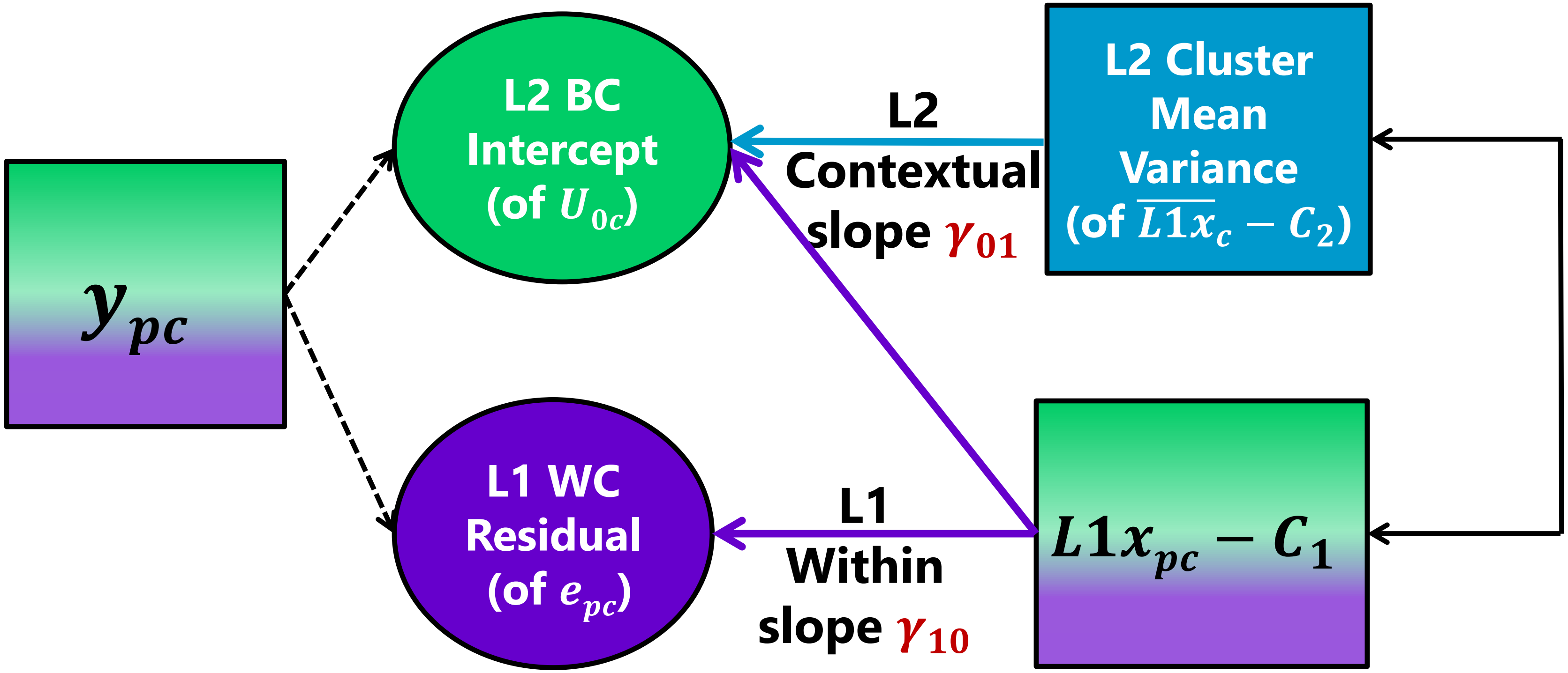
γ_{01} *becomes* the L2 Contextual slope that indicates <u>how the L2 BC effect differs from the L1 WC effect</u>

- **→ *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 L2 effect**



y_{pc}

L2 BC Intercept (of U_{0c})

L1 WC Residual (of e_{pc})

L2 Cluster Mean Variance (of $\overline{L1x_c} - C_2$)

$L1x_{pc} - C_1$

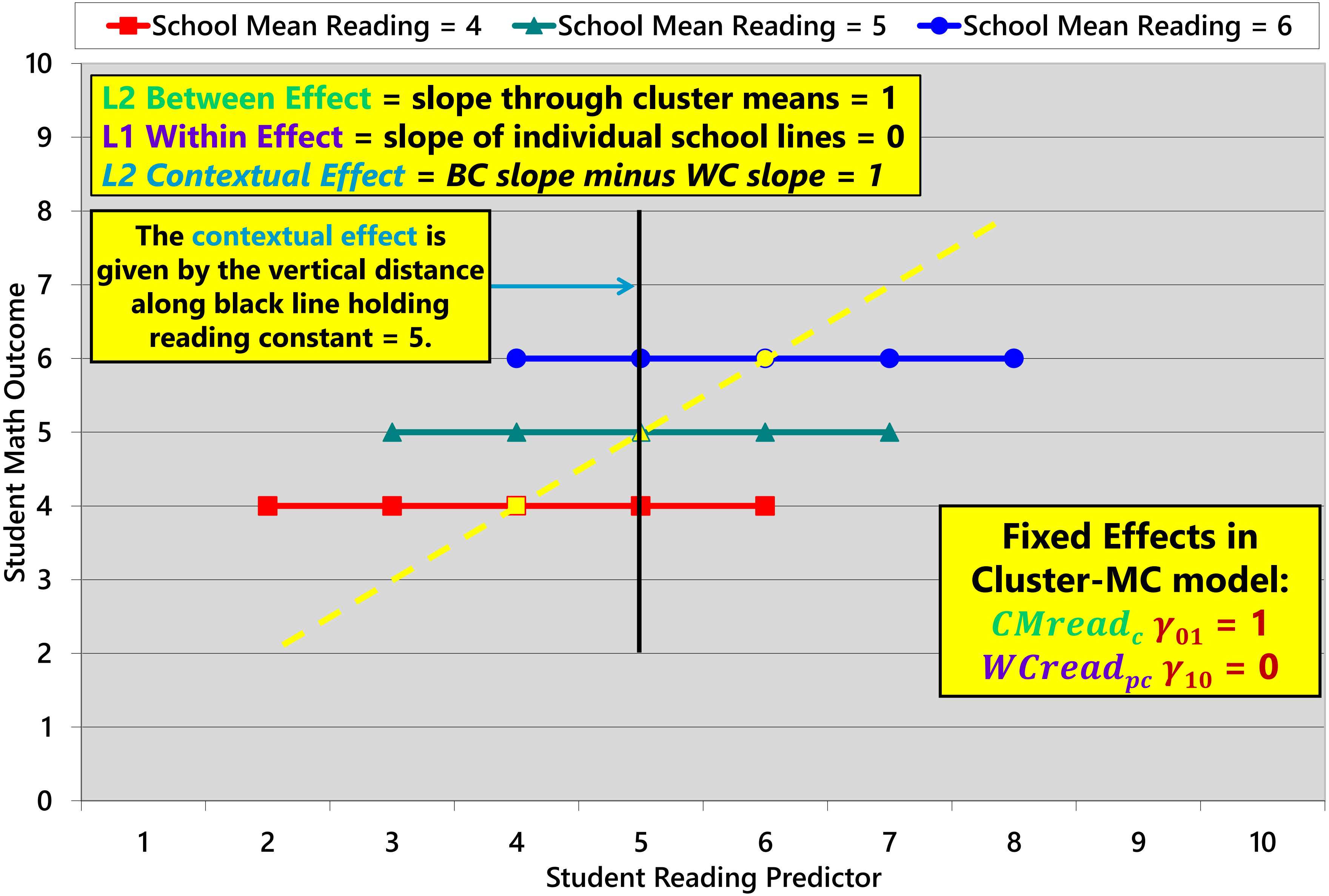
L2 Contextual slope γ_{01}

L1 Within slope γ_{10}

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…

ALL Between Effect, **NO** Within Effect

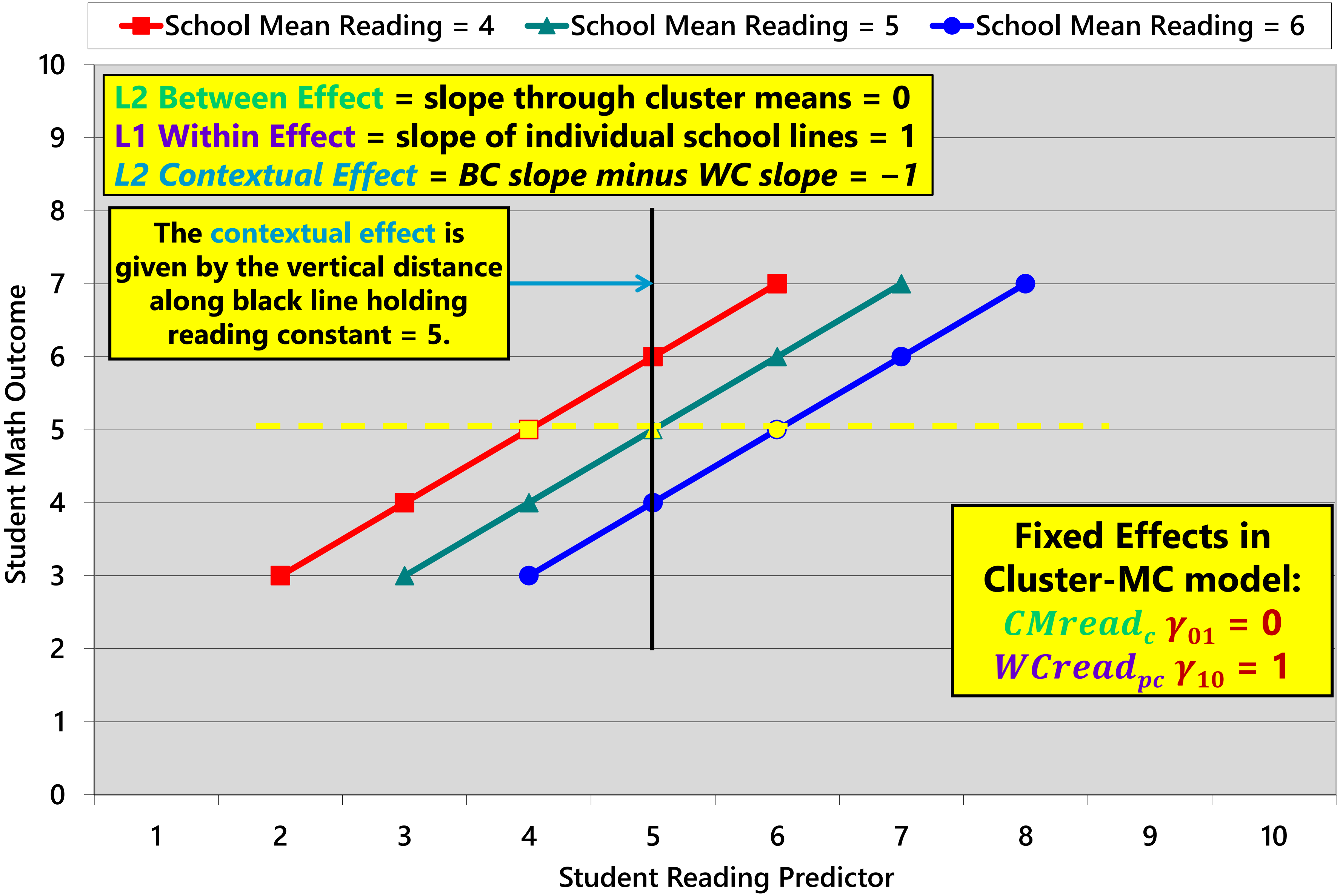


L2 Between Effect = slope through cluster means = 1
L1 Within Effect = slope of individual school lines = 0
L2 Contextual Effect = BC slope minus WC slope = 1

The contextual effect is given by the vertical distance along black line holding reading constant = 5.

Fixed Effects in Cluster-MC model:
$CMread_c\ \gamma_{01} = 1$
$WCread_{pc}\ \gamma_{10} = 0$

NO Between Effect, **ALL** Within Effect



L2 Between Effect = slope through cluster means = 0
L1 Within Effect = slope of individual school lines = 1
L2 Contextual Effect = BC slope minus WC slope = −1

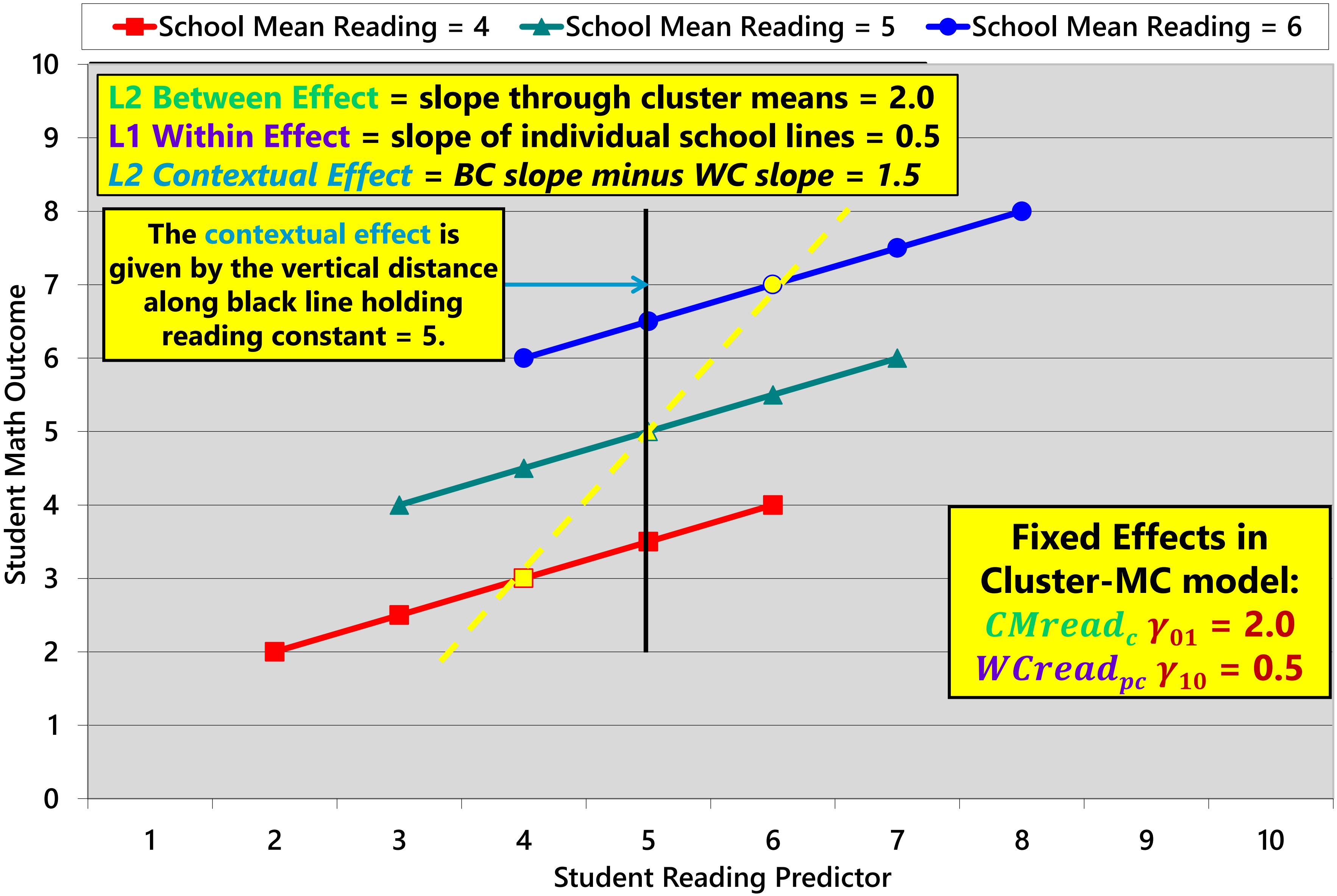
The contextual effect is given by the vertical distance along black line holding reading constant = 5.

Fixed Effects in Cluster-MC model:

$CMread_c$ $\gamma_{01} = 0$

$WCread_{pc}$ $\gamma_{10} = 1$

Between Effect > Within Effect



L2 Between Effect = slope through cluster means = 2.0

L1 Within Effect = slope of individual school lines = 0.5

L2 Contextual Effect = BC slope minus WC slope = 1.5

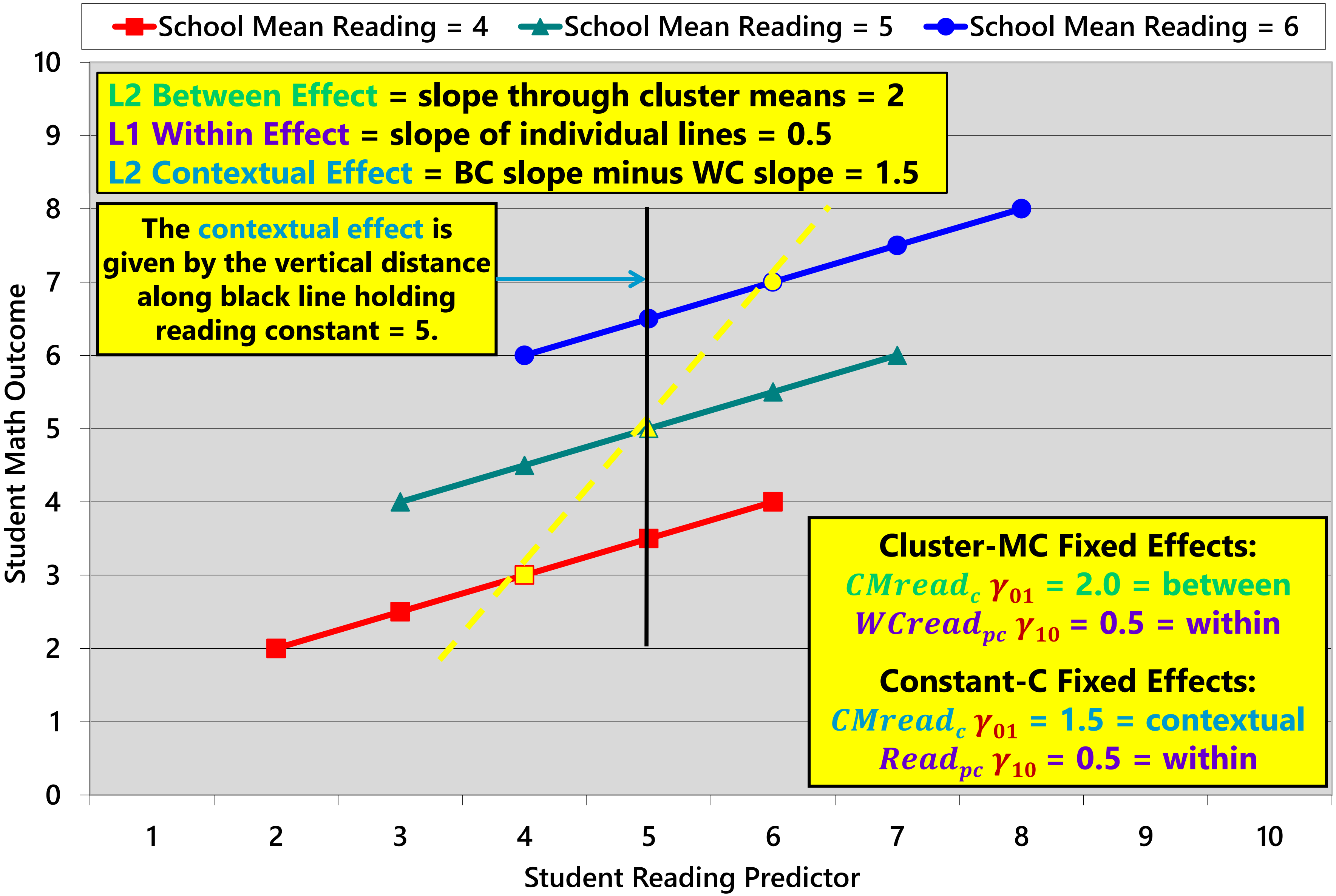
The contextual effect is given by the vertical distance along black line holding reading constant = 5.

Fixed Effects in Cluster-MC model:

$CMread_c\ \gamma_{01} = 2.0$

$WCread_{pc}\ \gamma_{10} = 0.5$

Between, Within, and Contextual Effects



L2 Between Effect = slope through cluster means = 2
L1 Within Effect = slope of individual lines = 0.5
L2 Contextual Effect = BC slope minus WC slope = 1.5

The contextual effect is given by the vertical distance along black line holding reading constant = 5.

Cluster-MC Fixed Effects:

$CMread_c\ \gamma_{01}$ = 2.0 = between

$WCread_{pc}\ \gamma_{10}$ = 0.5 = within

Constant-C Fixed Effects:

$CMread_c\ \gamma_{01}$ = 1.5 = contextual

$Read_{pc}\ \gamma_{10}$ = 0.5 = within

Adding L2 Contextual and L1 Within Predictors: (3a) Syntax by Univariate MLM Program

SAS:

```
PROC MIXED DATA=work.Example COVTEST NOCLPRINT IC METHOD=REML;
     CLASS schoolID;
     MODEL langpost = hw2 mixgrd CMverb10 verb10 / SOLUTION DDFM=Satterthwaite;
     RANDOM INTERCEPT / TYPE=UN SUBJECT=schoolID;
     ESTIMATE "L2 Between Effect of Verbal"  CMverb10 1 verb10 1;
RUN;
```

R lmer from lme4 package—using lmerTest package to get Satterthwaite denominator DF and contest1D:

```
name = lmer(data=Example, REML=TRUE,
            formula=langpost~1+hw2+mixgrd+CMverb10+verb10+(1+|schoolID))
summary(name, ddf="Satterthwaite")
contest1D(name, ddf="Satterthwaite", L=c(0,0,0,1,1)) # L2 Between effect of verbal
```

STATA:

```
mixed langpost c.hw2 c.mixgrd c.CMverb10 c.verb10, || schoolID:, ///
      reml dfmethod(satterthwaite) dftable(pvalue) nolog
lincom c.CMverb10*1 + c.verb10*1, small // L2 Between effect of verbal
```

SPSS:

```
MIXED langpost BY schoolID WITH hw2 mixgrd CMverb10 verb10
      /METHOD    = REML
      /CRITERIA  = DFMETHOD(SATTERTHWAITE)
      /PRINT     = SOLUTION TESTCOV
      /FIXED     = hw2 mixgrd CMverb10 verb10
      /RANDOM    = INTERCEPT | COVTYPE(UN) SUBJECT(schoolID)
      /TEST      = "L2 Between effect of verbal" CMverb10 1 verb10 1.
```

Electronic materials for this example from my 2023 APA training sessions are **here**

Example: Grand-MC Level-1 Predictor

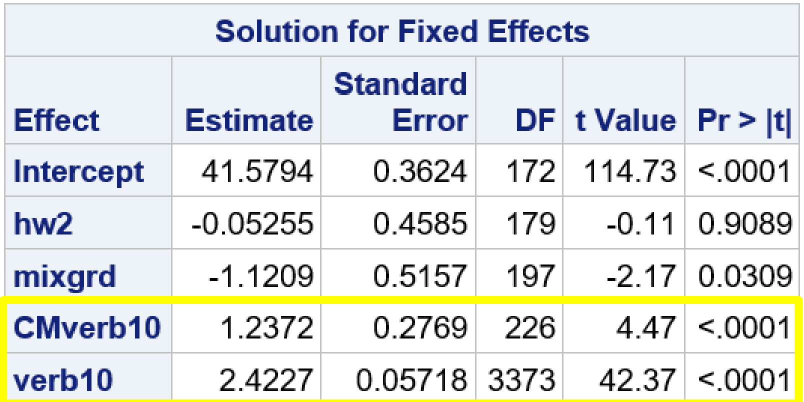
Level-1: $Lang_{pc} = \beta_{0c} + \beta_{1c}(Verbal_{pc} - 10) + e_{pc}$

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

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

Fixed Effects from SAS MIXED (model for variance is same):

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

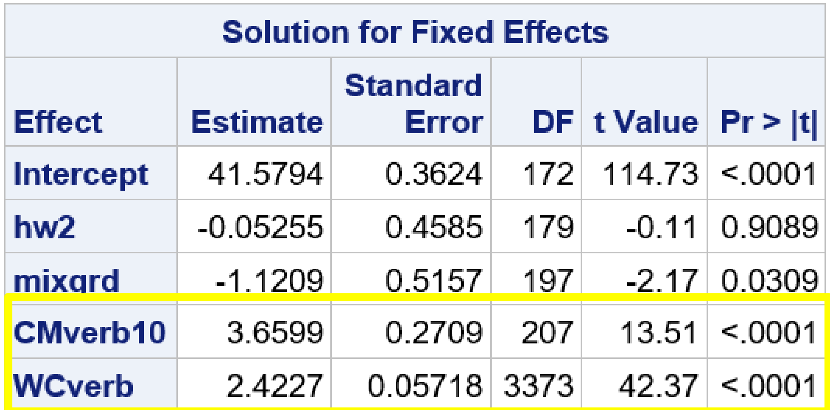


Solution for Fixed Effects					
Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	41.5794	0.3624	172	114.73	<.0001
hw2	-0.05255	0.4585	179	-0.11	0.9089
mixgrd	-1.1209	0.5157	197	-2.17	0.0309
CMverb10	1.2372	0.2769	226	4.47	<.0001
verb10	2.4227	0.05718	3373	42.37	<.0001

← L2? →

L1 WC

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



Solution for Fixed Effects					
Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	41.5794	0.3624	172	114.73	<.0001
hw2	-0.05255	0.4585	179	-0.11	0.9089
mixgrd	-1.1209	0.5157	197	-2.17	0.0309
CMverb10	3.6599	0.2709	207	13.51	<.0001
WCverb	2.4227	0.05718	3373	42.37	<.0001

- L2 Contextual slope = **1.24** using constant-C L1 (or Contextual = Between − Within)
- L2 Between slope = **3.66** 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

Btw, I am using a centering constant = 0 at both levels to simplify the notation so that $\overline{L1x}_c = CMx_c$

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

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

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

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

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

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

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

Grand-MC:

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

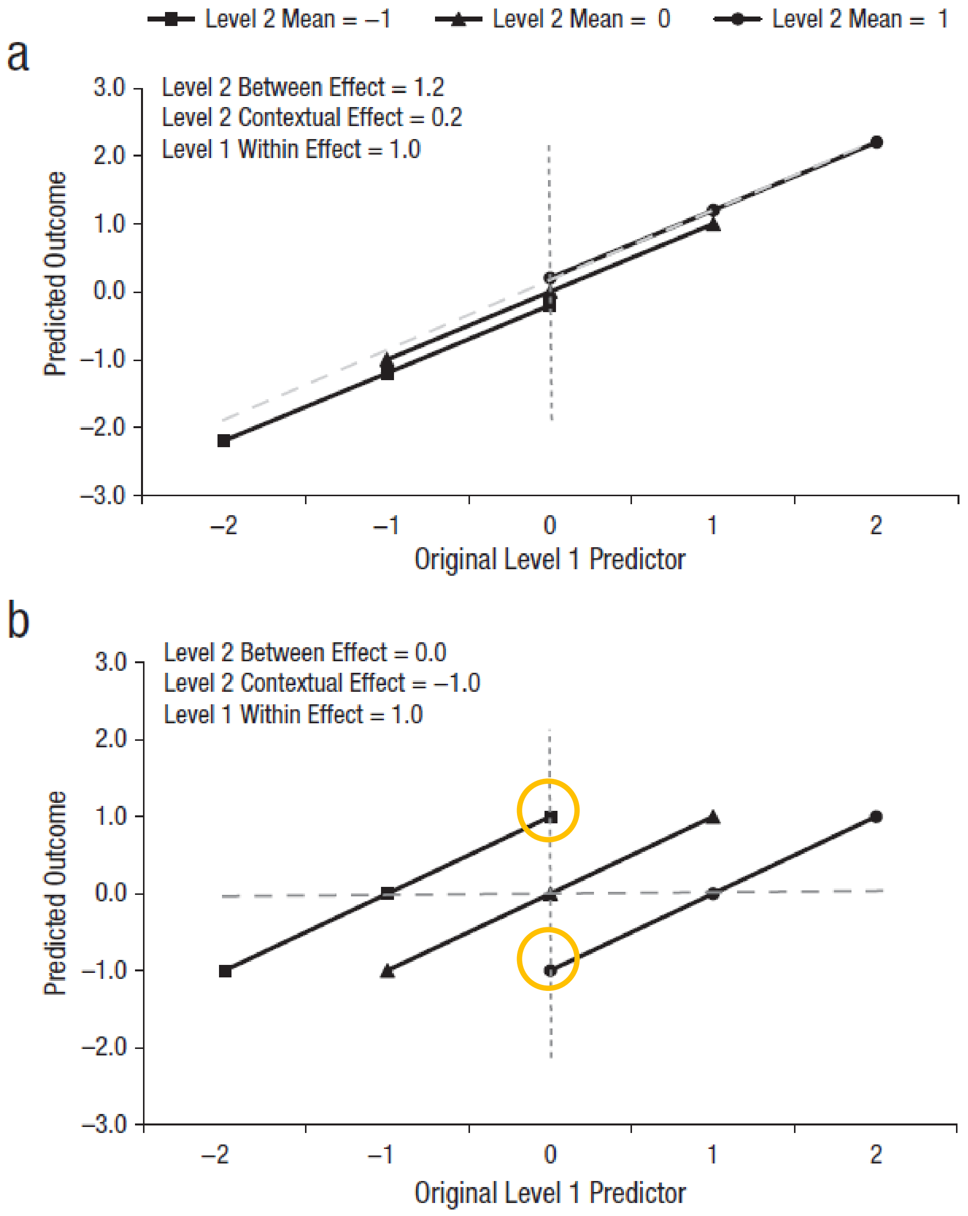
Level-2: $\beta_{0c} = \gamma_{00} + \gamma_{01}(CMx_c) + U_{0c}$

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

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

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)
- *Bottom:* Contextual effect is negative—at the same student SES level, relatively high students from low-SES schools do better than relatively low students from high-SES schools

Option 3. Latent-Centering (C-MC)

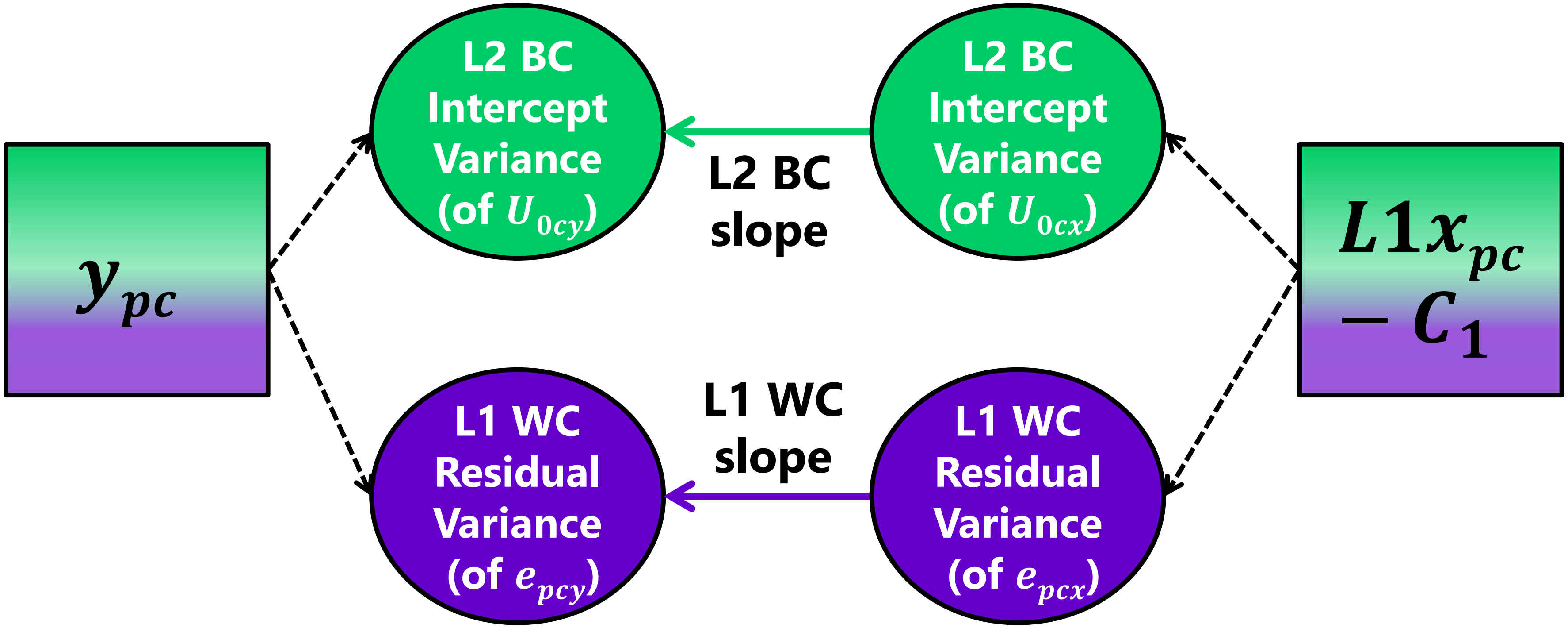
- We let the model partition the L1 predictor $\boldsymbol{L1x_{pc}}$ into two latent variables that directly represent its **L2 between**-cluster (BC) and **L1 within**-cluster (WC) sources of variation, just as we did for y_{pc} !
  - At a minimum: Fit an empty means, random intercept model for $L1x_{pc}$ (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, $\boldsymbol{\overline{L1x}_c} - \boldsymbol{C_2}$, 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 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)
  - The next 2 slides have a quick example for now, but we will explore this in more detail later in the semester in the context of multilevel mediation

Option 3: Latent-Centering in Multivariate MLM

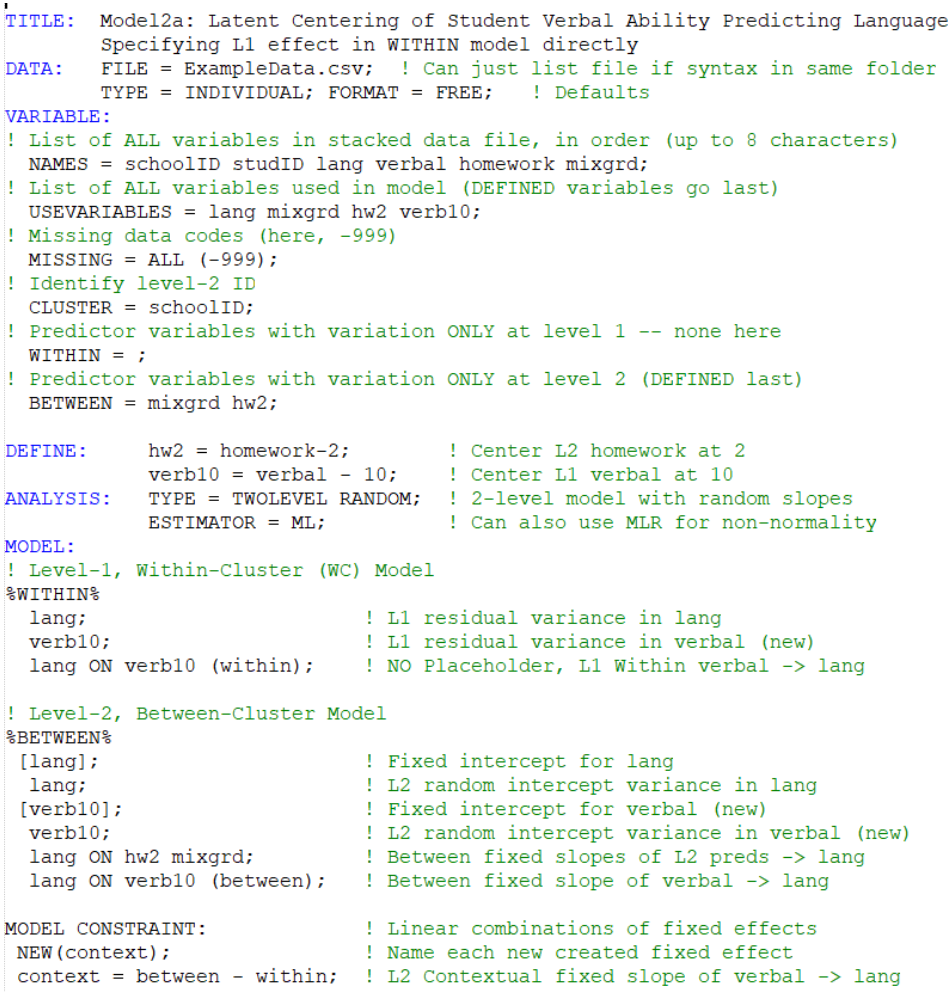
Model-based partitioning of level-1 y_{pc} outcome into level-specific **latent variables**

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



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, L., & Walters, R. W. (2022). Catching up on multilevel modeling. *Annual Review of Psychology, 73*, 629-658.

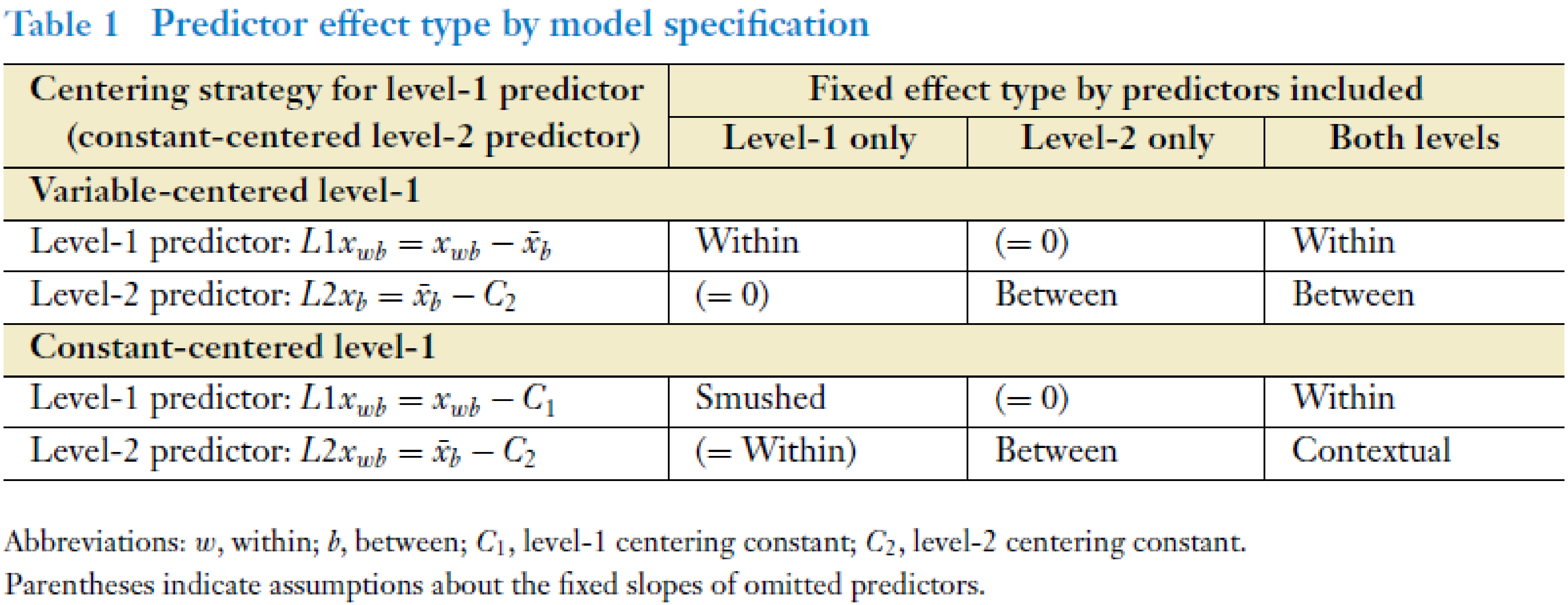


Table 1 Predictor effect type by model specification

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 $\boldsymbol{\tau_{U_0}^2}$ = Observed $\boldsymbol{\tau_{U_0}^2}$ − $\boldsymbol{(\sigma_e^2/L1n)}$**

- **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, although there is some debate about the "right" order of operations when using cluster-MC L1 predictors (e.g., 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 what 0 means is inconsistent across the main effects and interaction, which messes up the fixed intercept and simple slopes!
  - 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
- **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 $\boldsymbol{L1x_{pc}}$: $\overline{\boldsymbol{L1x}_{c}}$ = 0.5? Then:
    $\boldsymbol{WCx_{pc}} = \boldsymbol{L1x_{pc}} - \overline{\boldsymbol{L1x}_{c}}$ → $\boldsymbol{WCx_{pc}}$ **= +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** (such as is common in STATA or Mplus)
  - Although they control fixed effect standard errors for dependency due to cluster mean differences, **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**: Manually carve up into L2 BC (cluster mean → **L2 Between slope**) and L1 WC deviation (→ **L1 Within slope**)
  2. **Grand-Mean-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, so we'll save this one for later in the semester

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