General Multilevel Models (MLMs) for Two-Level Nested Data: Level-2 and Level-1 Fixed Effects

- Topics:
  - From single-level to multilevel empty means models
  - Intraclass correlation (ICC) and design effects
  - Fixed effects of level-2 predictors
  - Fixed effects of level-1 predictors

The Two Sides of a Single-Level Model

$$y_i = \beta_0 + \beta_1 X_i + \beta_2 Z_i + \beta_3 X_i Z_i + e_i$$

Our focus now

- **Model for the Means (Predicted Values):**
  - Each person's expected (predicted) outcome is a weighted linear function of his/her values on X and Z (and here, their interaction), each measured once per person
  - Estimated parameters are called fixed effects (here, β_0, β_1, β_2, and β_3)
  - The number of fixed effects will show up in formulas as k (so k = 4 here)

- **Model for the Variance ("Piles" of Variance):**
  - $e_i \sim N(0, \sigma_e^2)$ → **ONE** source of residual (unexplained) deviation
  - e_i has a mean of 0 with some estimated constant residual variance σ_e^2, is normally distributed, is unrelated to X and Z, and is unrelated across people (across all observations more generally, but just people here)
  - **Contains a single residual variance only in above single-level model**

Review: Variances and Covariances

Variance:
Dispersion of y

$$\text{Variance}(y_{pg}) = \frac{\sum_{i=1}^{N_g}(y_{pg} - \hat{y}_{pg})^2}{N_g - k}$$

Covariance:
How y's go together, unstandardized

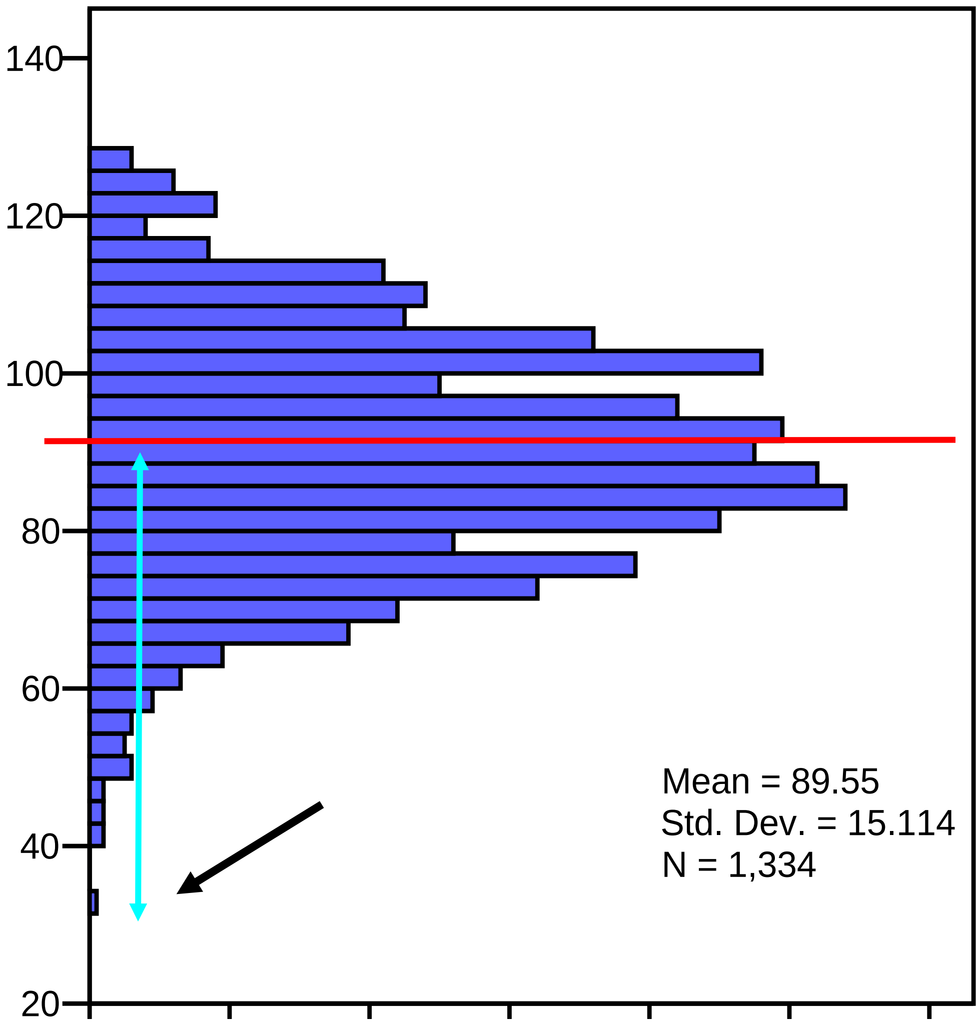
$$\text{Covariance}(y_1, y_2) = \frac{\sum_{i=1}^{N_g}(y_1 - \hat{y}_1)(y_2 - \hat{y}_2)}{N_g - k}$$

Correlation:
How y's go together, standardized (−1 to 1)

$$\text{Correlation}(y_1, y_2) = \frac{\text{Covariance}(y_1, y_2)}{\sqrt{\text{Variance}(y_1)} * \sqrt{\text{Variance}(y_2)}}$$

N_g = # Groups, p = person, g = groups
k = # fixed effects, $\hat{y}_{ti}$ = y predicted from fixed effects

An Empty Means, Single-Level Model



$$y_{pg} = \beta_0 + e_{pg}$$

Filling in values:

$$32 = \underbrace{90}_{\text{Y pred}} + -58$$

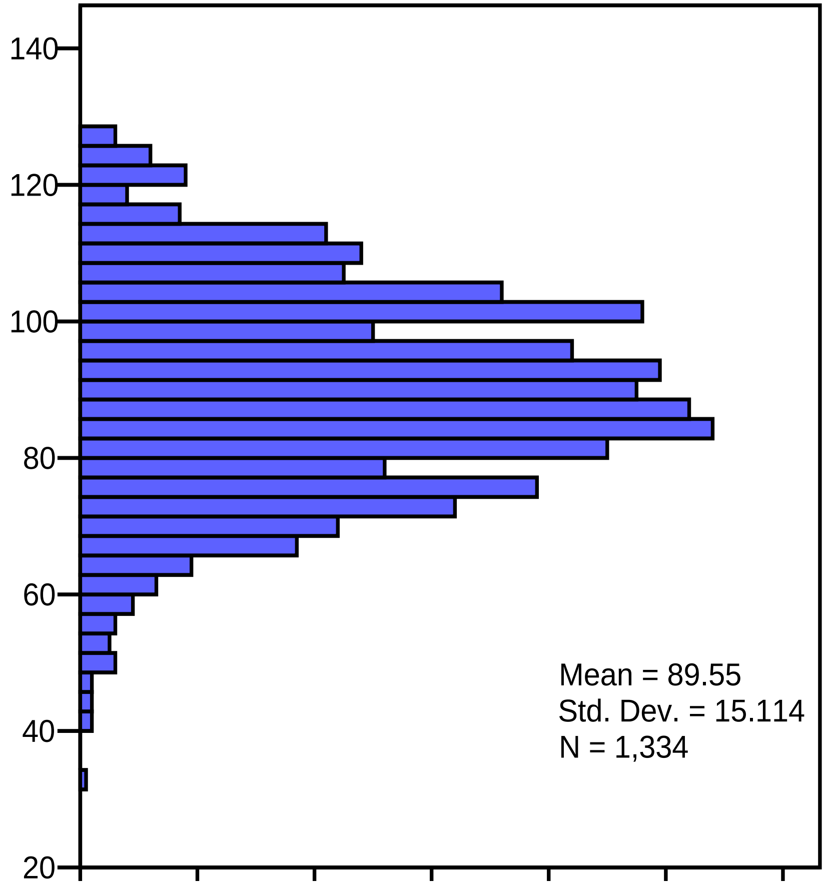
Model for the Means

Y Error Variance:

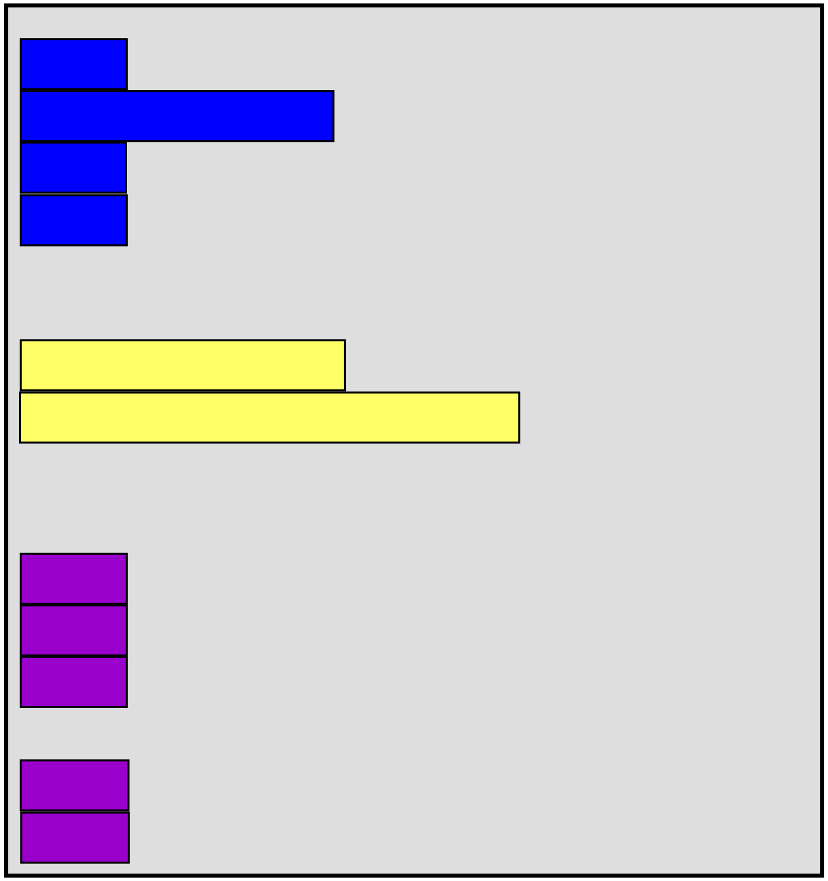
$$\frac{\Sigma\,(y - y_{pred})^2}{N_{total} - 1}$$

Adding Group-Level Information… (i.e., to become a Multilevel Model)

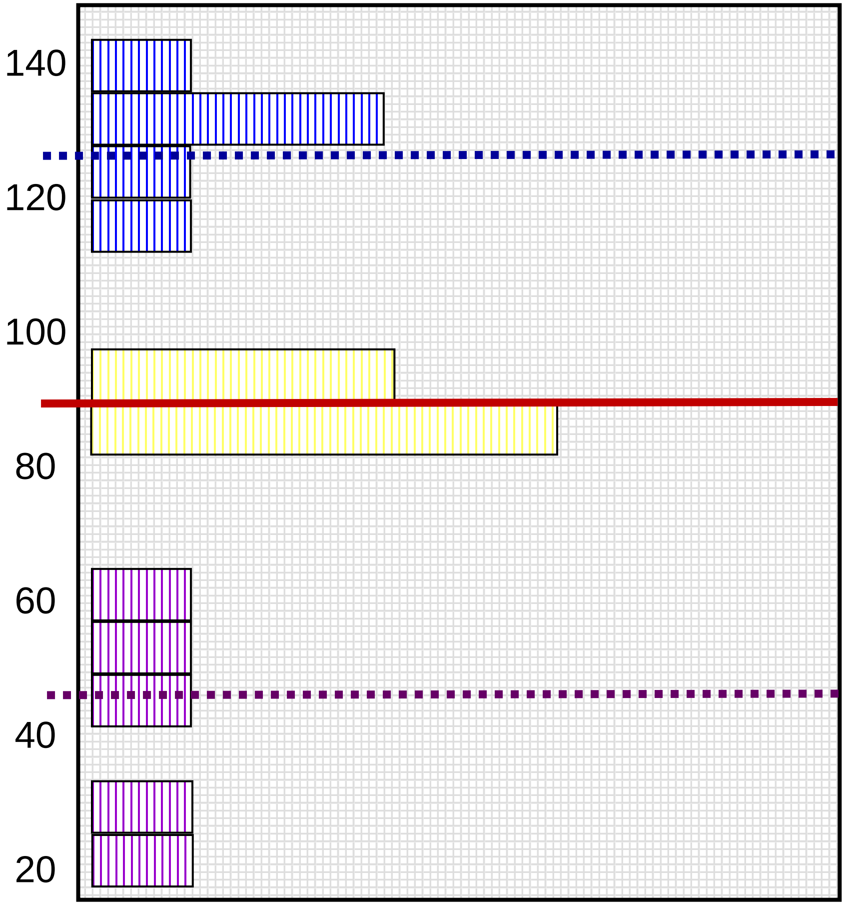
Full Sample Distribution



3 Groups, 5 Persons each



Empty Means Multilevel Model



Start off with Mean of Y as "best guess" for any value:

= Grand Mean

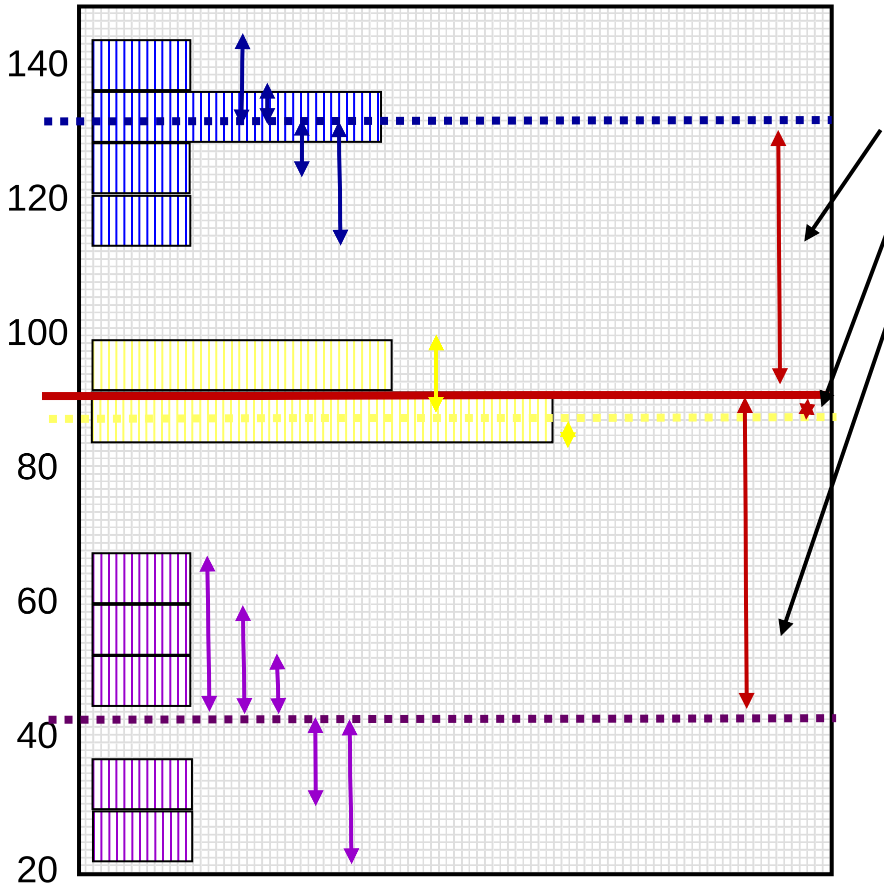
= Fixed Intercept

Can make better guess by taking advantage of group information:

= Group Mean

→ Random Intercept

Empty Means Multilevel Model



Variance of Y → 2 sources:

Between-Group (BG) Variance:

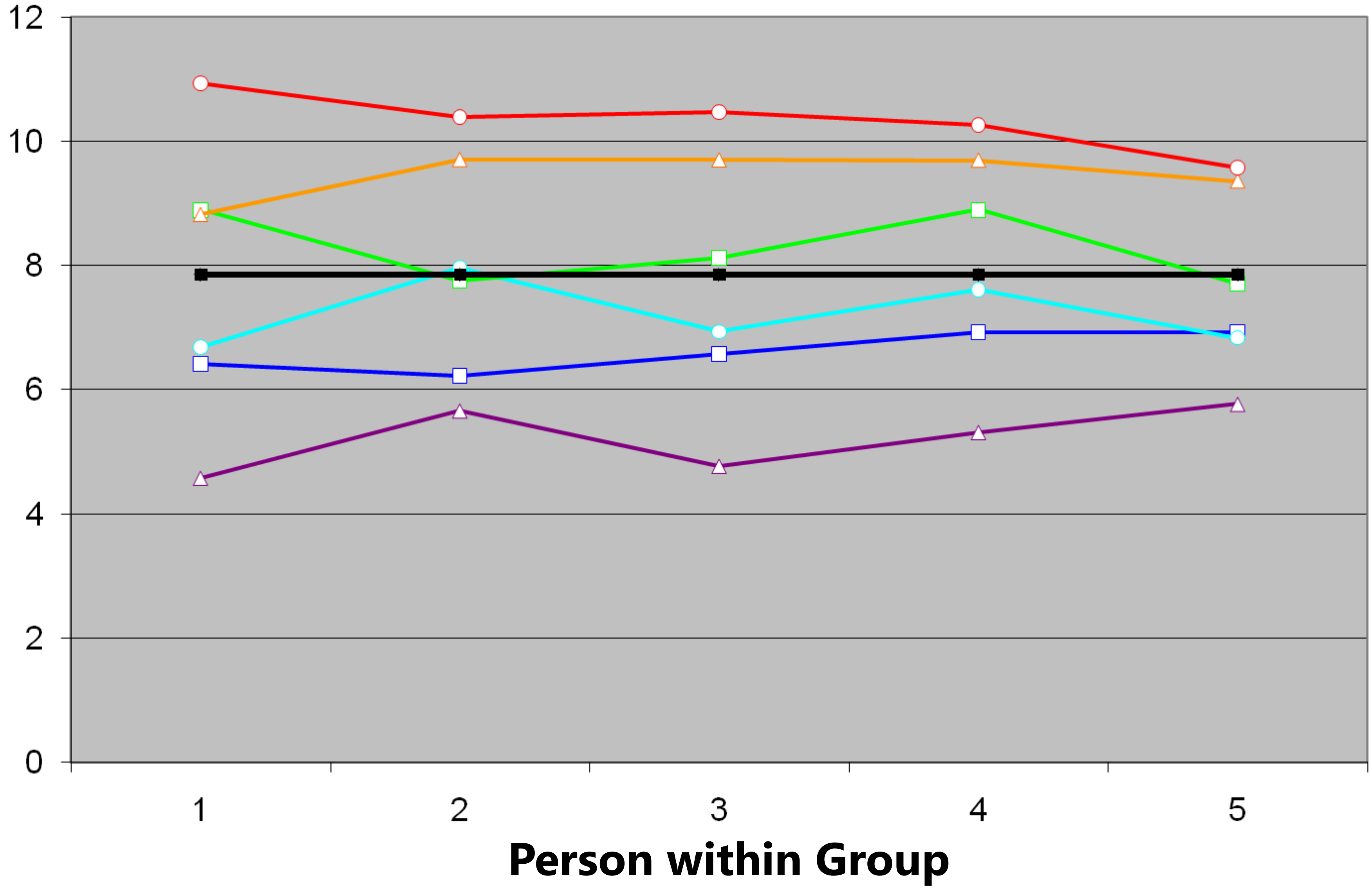
- → Differences from **GRAND** mean
- → **INTER**-group differences

Within-Group (WG) Variance:

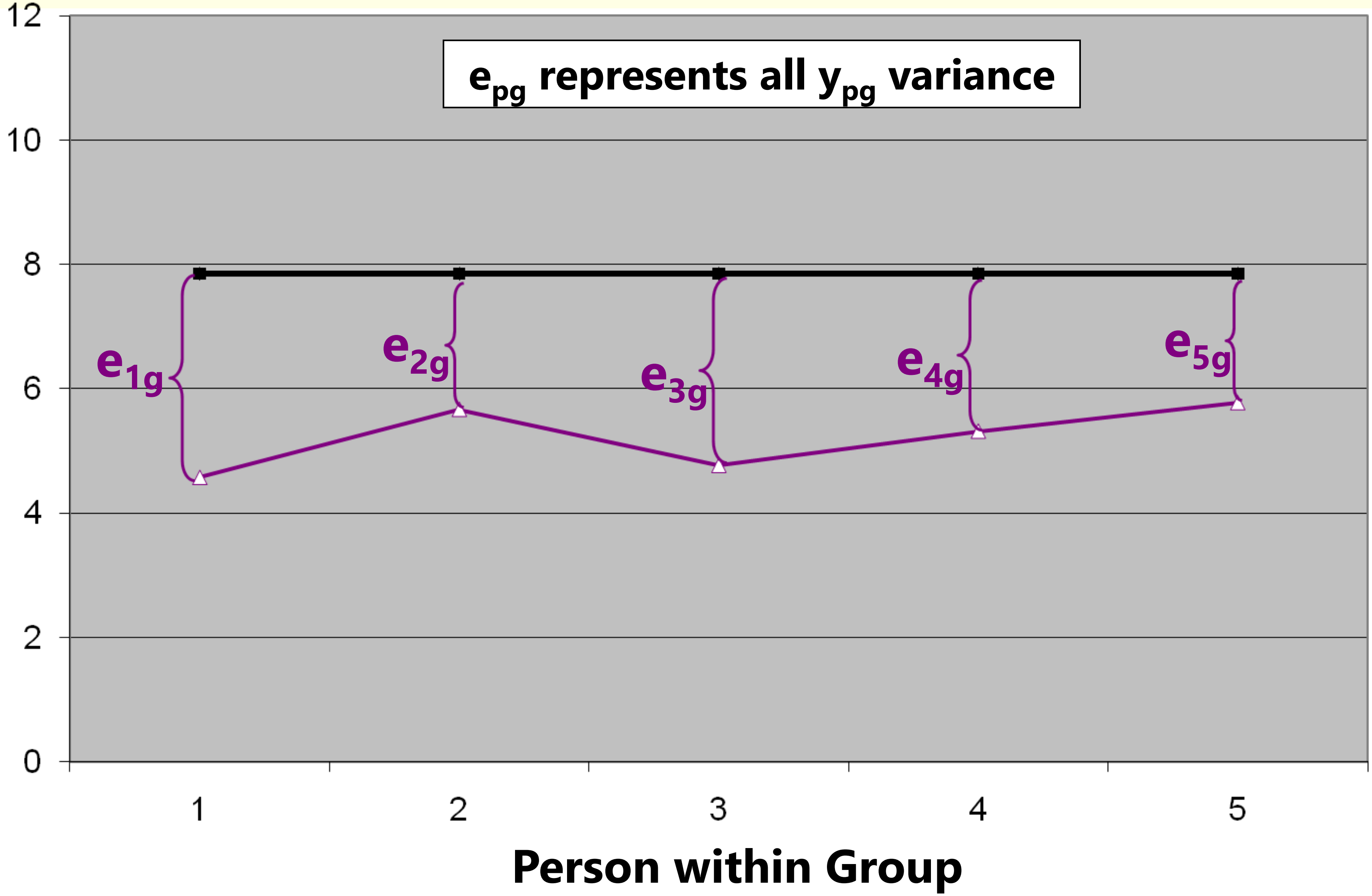
- → Differences from **GROUP** mean
- → **INTRA**-group differences

Now we have 2 piles of variance in Y to predict.

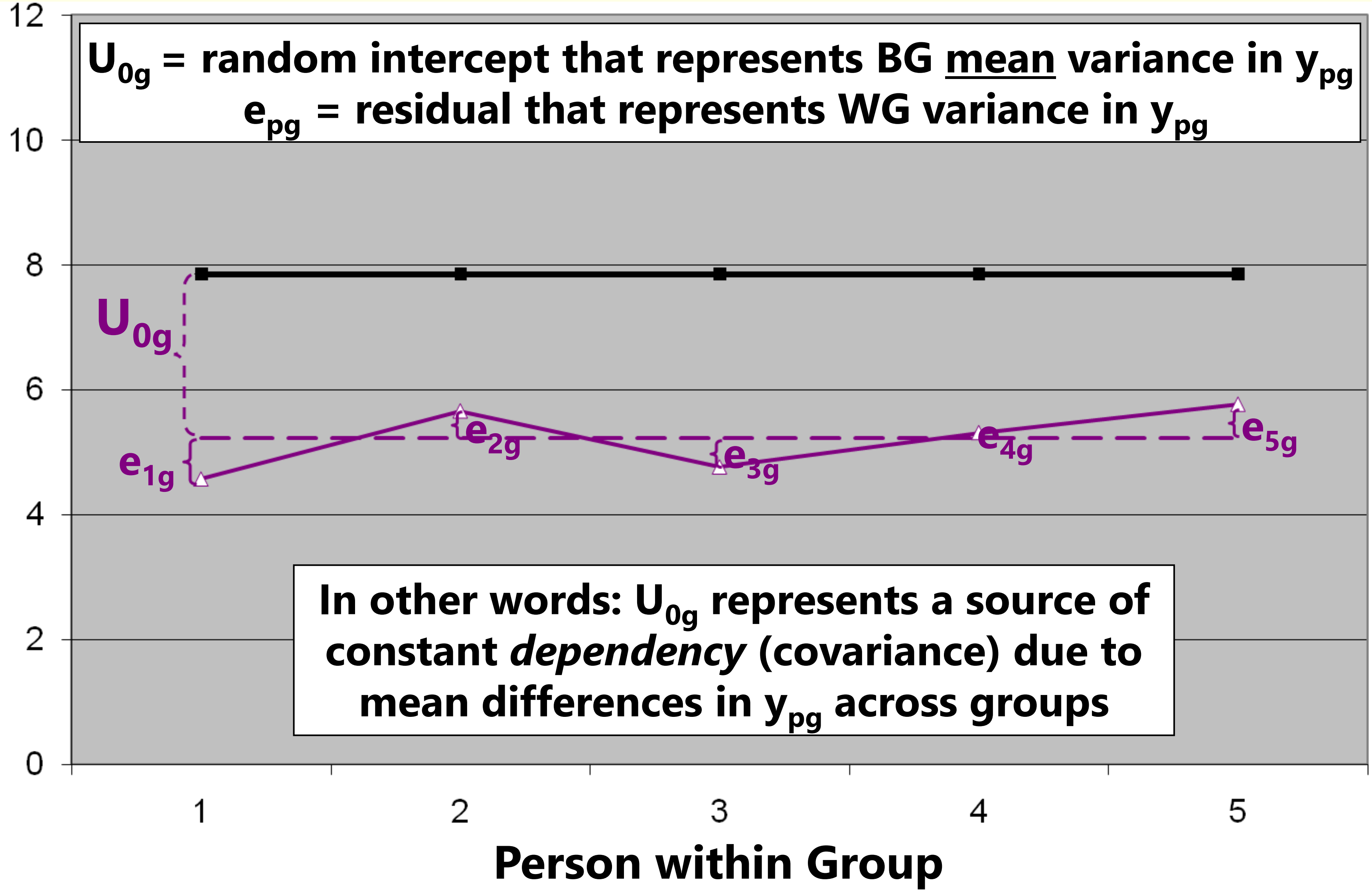
Hypothetical Two-Level Nested Data



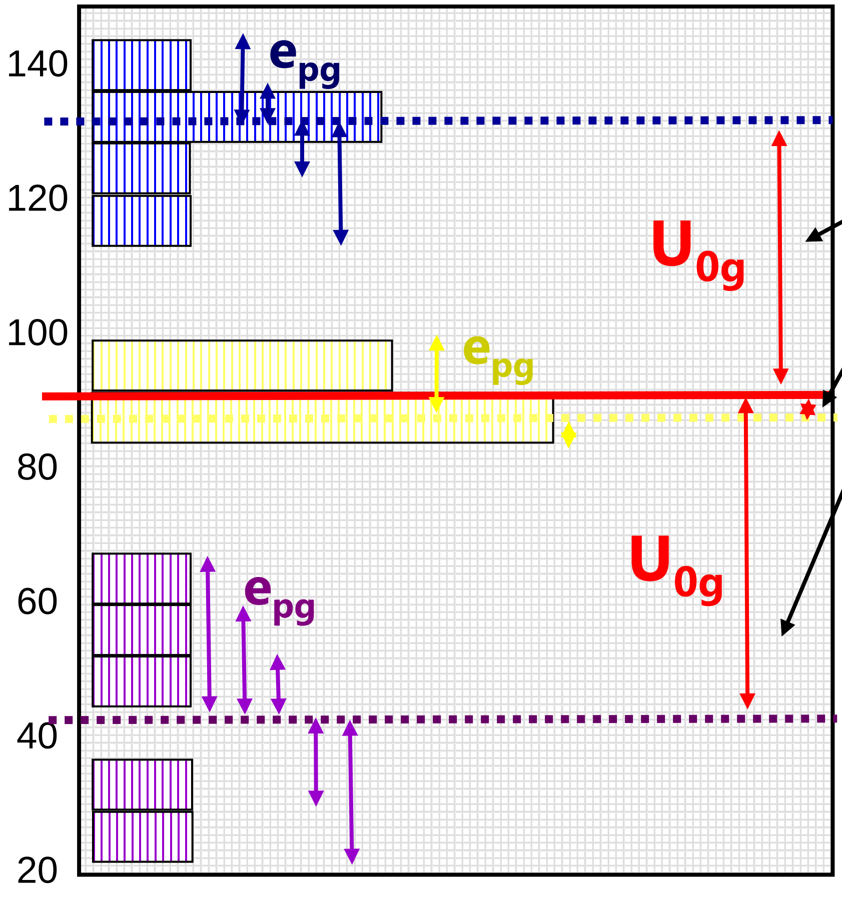
"Error" in a Single-Level Model for the Variance



Partitioning of "Error" in a Multilevel Model for the Variance



Empty Means Multilevel Model



Variance of Y → 2 sources

Level 2 Random Intercept Variance (of U_{0g}, as $\tau^2_{U_0}$):

→ **Between**-group variance

→ Differences from **GRAND** mean

→ **INTER**-group differences

Level 1 Residual Variance (of e_{pg}, as σ^2_e):

→ **Within**-group variance

→ Differences from **GROUP** mean

→ **INTRA**-group differences

Empty Means, Random Intercept Model

GLM Empty Model:

- $y_i = \beta_0 + e_i$

MLM Empty Model:

- Level 1:

$$y_{pg} = \beta_{0g} + e_{pg}$$

- Level 2:

$$\beta_{0g} = \gamma_{00} + U_{0g}$$

Fixed Intercept =grand mean of group means (because no predictors yet)

Random Intercept = group-specific deviation from predicted intercept

Residual = person-specific deviation from group's predicted outcome

3 Total Parameters:

Model for the Means (1):

- Fixed Intercept γ_{00}

Model for the Variance (2):

- Level-1 Variance of e_{pg} → σ^2_e
- Level-2 Variance of U_{0g} → $\tau^2_{U_0}$

Composite equation:

$$y_{pg} = (\gamma_{00} + U_{0g}) + e_{pg}$$

Intraclass Correlation (ICC)

$$\text{ICC} = \frac{\text{BG}}{\text{BG} + \text{WG}} = \frac{\text{Intercept Variance}}{\text{Intercept Variance} + \text{Residual Variance}}$$

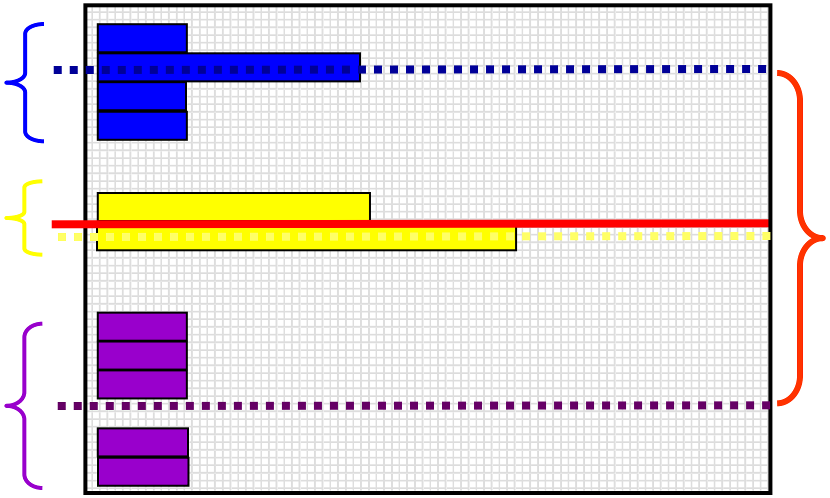
$$= \frac{\boldsymbol{\tau^2_{U_0}}}{\boldsymbol{\tau^2_{U_0}} + \boldsymbol{\sigma^2_e}}$$

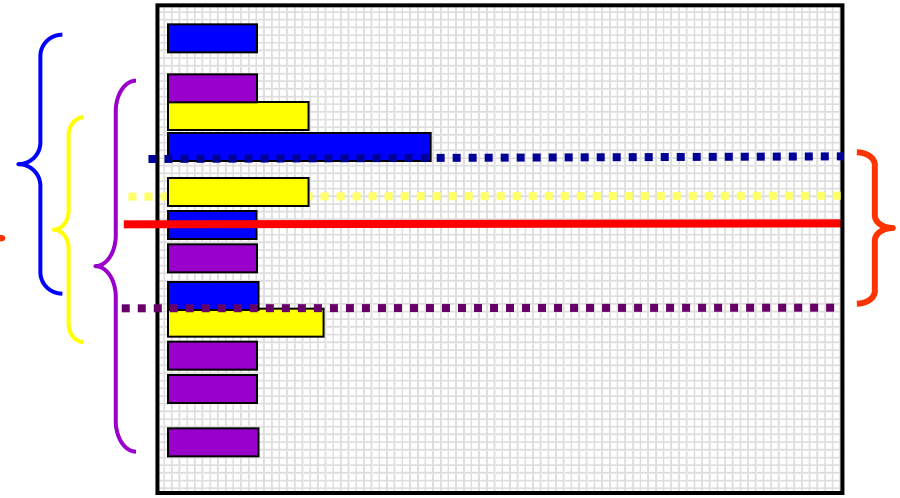
$\boldsymbol{\tau^2_{U_0}}$ → Why don't all groups have the same mean?
$\boldsymbol{\sigma^2_e}$ → Why don't all people from the same group have the same outcome?

- ICC = Proportion of total variance that is <u>between groups</u>
- ICC = Average correlation among persons from same <u>group</u>
- ICC is a standardized way of expressing how much we need to worry about *dependency due to <u>group mean differences</u>* **(i.e., ICC is an effect size for *constant* <u>group</u> dependency)**
  - Dependency of other kinds can still be created by differences between groups in the effects of predictors (stay tuned)

$$\text{ICC} = \frac{\text{BetweenGroup}}{\text{BetweenGroup} + \text{WithinGroup}}$$

Counter-Intuitive: Between-Group Variance is the numerator, but the ICC is the correlation within groups—what??





ICC = BTW / BTW + within

→ Large ICC

→ Large correlation within groups

ICC = btw / btw + WITHIN

→ Small ICC

→ Small correlation within groups

Effects of Clustering on Effective N

- **Design Effect** expresses how much effective sample size needs to be adjusted due to clustering/grouping
- **Design Effect** = ratio of the variance using a given sampling design to the variance using a simple random sample from the same population, given the same total sample size either way

- Design Effect = $1 + \left[(n-1) * \text{ICC}\right]$ n = # level-1 units
- Effective sample size → $N_{\text{effective}} = \dfrac{\text{\# Total Observations}}{\text{Design Effect}}$

- As ICC goes UP and cluster size goes UP, the effective sample size goes DOWN
  - See Snijders & Bosker (2012) for more info and for a modified formula that takes unequal group sizes into account

Design Effects in 2-Level Nesting

- Design Effect = $1 + \left[(n-1) * \text{ICC}\right]$
- Effective sample size → $N_{\text{effective}} = \frac{\text{\# Total Observations}}{\text{Design Effect}}$

- $n = 5$ patients from each of 100 doctors, ICC = .30?
  - Patients Design Effect = 1 + (4 * .30) = 2.20
  - $\boldsymbol{N_{\text{effective}}}$ = 500 / 2.20 = **227** (not 500)

- $n = 20$ students from each of 50 schools, ICC = .05?
  - Students Design Effect = 1 + (19 * .05) = 1.95
  - $\boldsymbol{N_{\text{effective}}}$ = 1000 / 1.95 = **513** (not 1000)

Does a non-significant ICC mean you can ignore groups and just do a regression?

- Effective sample size depends on BOTH the ICC and the number of people per group: As ICC goes UP and group size goes UP, the effective sample size goes DOWN
  - So there is NO VALUE OF ICC that is "safe" to ignore, not even ~0!
  - An ICC=0 in an *empty means (unconditional)* model can become ICC>0 after adding level-1 predictors because reducing the residual variance will then increase the random intercept variance (→ *conditional* ICC > 0)
  - Design effects can become much higher given good level-1 predictors!

- So just do a multilevel analysis anyway…
  - Even if "that's not your question"… because people come from groups, you still have to model group dependency appropriately because of:
    - Effect of clustering on level-1 fixed effect SEs → **biased SEs**
    - Potential for **contextual effects** of level-1 predictors

2 Options for Differences Across Groups

Represent Group Differences as Fixed Effects

- Include (#groups-1) contrasts for group membership in the **model for the means** (via CLASS/i.)→ so group is NOT another "level"
- Permits inference about differences between specific groups, but you cannot include between-group predictors (group is saturated)
- Snijders & Bosker (2012, p. 48) recommend if #groups < 10ish
- Clustered-corrected SEs are analogous to this option

Represent Group Differences as a Random Effect

- Include a **random intercept variance in the model for the variance**, such that group differences become another "level"
- Permits inference about differences across groups more generally, for which you can test effects of between-group predictors
- Better if #groups > 10ish and you want to **predict** group differences

MLMs with Level-2 (Group-Level) Predictors

- **Level-2 predictors** are constant over persons from the same group—they are group-level characteristics
- For example: level-1 students (p) nested in level-2 schools (g) that are public (=0) or private (=1) and vary in size (0 = mean)

- Level-1 Model: $\mathbf{Math_{pg} = \beta_{0g} + e_{pg}}$

- "Unconditional" Level-2 Model (without predictors):
  - $\boldsymbol{\beta_{0g} = \gamma_{00} + U_{0g}}$

$\tau^2_{U_0}$ → **All possible variance** due to intercept (mean) differences across schools → baseline

- "Conditional" Level-2 Model (with predictors):
  - $\boldsymbol{\beta_{0g} = \gamma_{00} + \gamma_{01}(Private_g) + \gamma_{02}(Size_g) + \gamma_{03}(Private_g)(Size_g) + U_{0g}}$
  - First subscript = which beta in level-1 model
    Second subscript = order of predictor in level-2 model

$\tau^2_{U_0}$ → Intercept variance **leftover** after controlling for school type and size

Example Conditional Multilevel Model

- **Level-1 Model:** $\mathbf{Math_{pg} = \beta_{0g} + e_{pg}}$
- **Level-2 Model:** $\mathbf{\beta_{0g} = \gamma_{00} + \gamma_{01}(Private_g) + \gamma_{02}(Size_g) + \gamma_{03}(Private_g)(Size_g) + U_{0g}}$

- $\boldsymbol{\gamma_{00}}$ = fixed intercept: expected math for students from a school with private=0 (=public school) and size=0 (=mean size)
- $\boldsymbol{\gamma_{01}}$ = fixed *simple* slope of private: difference in math between public and private schools *of size=0*
- $\boldsymbol{\gamma_{02}}$ = fixed *simple* slope of size: difference in math per one-unit higher size *for public schools*
- $\boldsymbol{\gamma_{03}}$ = fixed *interaction* slope with two possible interpretations:
  - difference in *effect of private* per one-unit higher in size
  - difference in *effect of size* between public and private schools
- $\mathbf{U_{0g}}$ = level-2 random intercept = deviation between actual and predicted school mean math for school g → unknown reasons, so is "error"
- $\mathbf{e_{pg}}$ = level-1 residual = deviation of student p's math from school's mean

Effect Size for Level-2 Predictors

- Direct: convert t-statistic for fixed effect into d or r
  - $d = \frac{2t}{\sqrt{DF_{den}}}$, $r = \frac{t}{\sqrt{t^2 + DF_{den}}}$

Note: These formulas can be used with any model (multilevel or not).

- Indirect: explained variance of two complementary kinds
  - **Pseudo-R²**: amount of variance explained *per variance component*

$$\text{Pseudo } R^2_{U_0} = \frac{\text{random intercept variance}_{\text{fewer}} - \text{random intercept variance}_{\text{more}}}{\text{random intercept variance}_{\text{fewer}}}$$

    - It can go negative if adding "really" unhelpful predictors or if the level-1 model is mis-specified (stay tuned); these problems can be remedied by calculating it with model-implied total variance (see Rights & Sterba, 2018)

  - **Total-R²**: amount of total variance explained (across piles)
    - Generate model-predicted y's from fixed effects ONLY and correlate with observed outcome; square that correlation to get total R^2

Example Conditional MLM Effect Size

- Let's assume the following values:
  - Empty means model: Level-1 $\sigma_e^2 = 8$ and Level-2 $\tau_{U_0}^2 = 2$, so **ICC = .20**
  - Conditional model: Level-1 $\sigma_e^2 = 8$ (because no level-1 predictors yet), Level-2 $\tau_{U_0}^2 = 1$ after controlling for private, size, and their interaction

- Variance explained measures:
  - Pseudo-$R_{U_0}^2 = \frac{2-1}{2} = \mathbf{.50}$ → The three level-2 fixed effects of private and size explained *50% of the <u>school-level intercept variance</u> in math*
  - Total-R^2 approximation when there is only a random intercept: Total-R^2 = Pseudo-$R_{U_0}^2$ * ICC = $.50 * .20 = \mathbf{.10}$ → The three fixed effects of private and size explained *10% of the <u>total variance</u> in math*
  - Because these R^2 values mean very different things, it is essential to clearly describe how you calculated them and what they then mean

Intermediate Summary

- Modeling process begins with an empty means model to determine how much variance is attributable to each dimension of sampling:
  - Level-2 between-group mean differences: random intercept ($\tau^2_{U_0}$)
  - Level-1 within-group person differences: residual (σ^2_e)
  - Effects size via Intraclass correlation: $\text{ICC} = \tau^2_{U_0}$ / ($\tau^2_{U_0}$+ σ^2_e)
    - ICC = proportion of total variance due to group mean differences
    - ICC = average correlation of persons from same group
    - Higher ICC and level-1 sample size → larger design effect → smaller effective N

- Treating groups as a random effect (by including $\tau^2_{U_0}$) allows us to test the effects of level-2 between-group predictors
  - Significance tests via univariate and multivariate Wald tests as usual (and t and F can be converted to d or r effect sizes)
  - Level-2 predictors reduce level-2 random intercept variance
    - Reduction in *level-2 intercept variance* is quantified by Pseudo-$R^2_{U_0}$
    - Reduction in *total variance* is quantified by ICC* Pseudo-$R^2_{U_0}$

Fixed Effects of Level-1 (Person-Level) Predictors: DANGER AHEAD!!!!

- Level-1 predictors are **person-level** characteristics
- For example: level-1 students (p) nested in level-2 schools (g) level-1 x = amount of time parents read to them (0 = mean)

- **Level-1 Model:** $\mathbf{Math_{pg} = \beta_{0g} + \beta_{1g}(Read_{pg}) + e_{pg}}$
- **Level-2 Model:** $\mathbf{\beta_{0g} = \gamma_{00} + U_{0g}}$

  $\mathbf{\beta_{1g} = \gamma_{10}}$

- All good, right? Many researchers mistakenly think so, but this model is VERY VERY VERY LIKELY to be broken
  - And no, it does NOT depend on your "theory" whether this is ok

Level-1 (Person-Level) Predictors

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

- Example: Amount of time parents spend reading to students
  - Some kids spend more time being read to than others in their school:
    - **WG variation in Read** *(represented directly as deviation from school mean)*
  - Some schools have more reading-focused parents than other schools:
    - **BG variation in Read** *(represented as school mean or via external info)*

- Can quantify each source of variance with an ICC
  - ICC = (BG variance) / (BG variance + WG variance)
  - ICC > 0? Level-1 predictor has BG variation (so it *could* have BG effect)
  - ICC < 1? Level-1 predictor has WG variation (so it *could* have WG effect)

Between-Group vs. Within-Group Effects

- Between-group and within-group effects in SAME direction
  - SES→ Achievement?
    - **BG: Schools with more money than other schools may have greater mean achievement than schools with less money**
    - **WG: 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)**
- Between-group and within-group effects in OPPOSITE directions
  - Body mass → life expectancy in animals (Curran and Bauer, 2011)?
    - **BG: Larger species tend to have longer life expectancies than smaller species (e.g., whales live longer than cows, cows live longer than ducks)**
    - **WG: Within a species, relatively bigger animals have shorter life expectancy (e.g., over-weight ducks die sooner than healthy-weight ducks)**
- Variables have **different meanings** and **different scales** across levels (so "one-unit" effects will rarely be the same across levels)!

Fixed Effects of Level-1 (Person-Level) Predictors: HERE IS THE DANGER

- For example: level-1 students (p) nested in level-2 schools (g) level-1 x = amount of time parents read to them (0 = mean)

- **Level-1:** $\mathbf{Math_{pg} = \beta_{0g} + \beta_{1g}(Read_{pg}) + e_{pg}}$
- **Level-2:** $\mathbf{\beta_{0g} = \gamma_{00} + U_{0g}}$

  $\mathbf{\beta_{1g} = \gamma_{10}}$

γ_{10} = *smushed* WG and BG effects

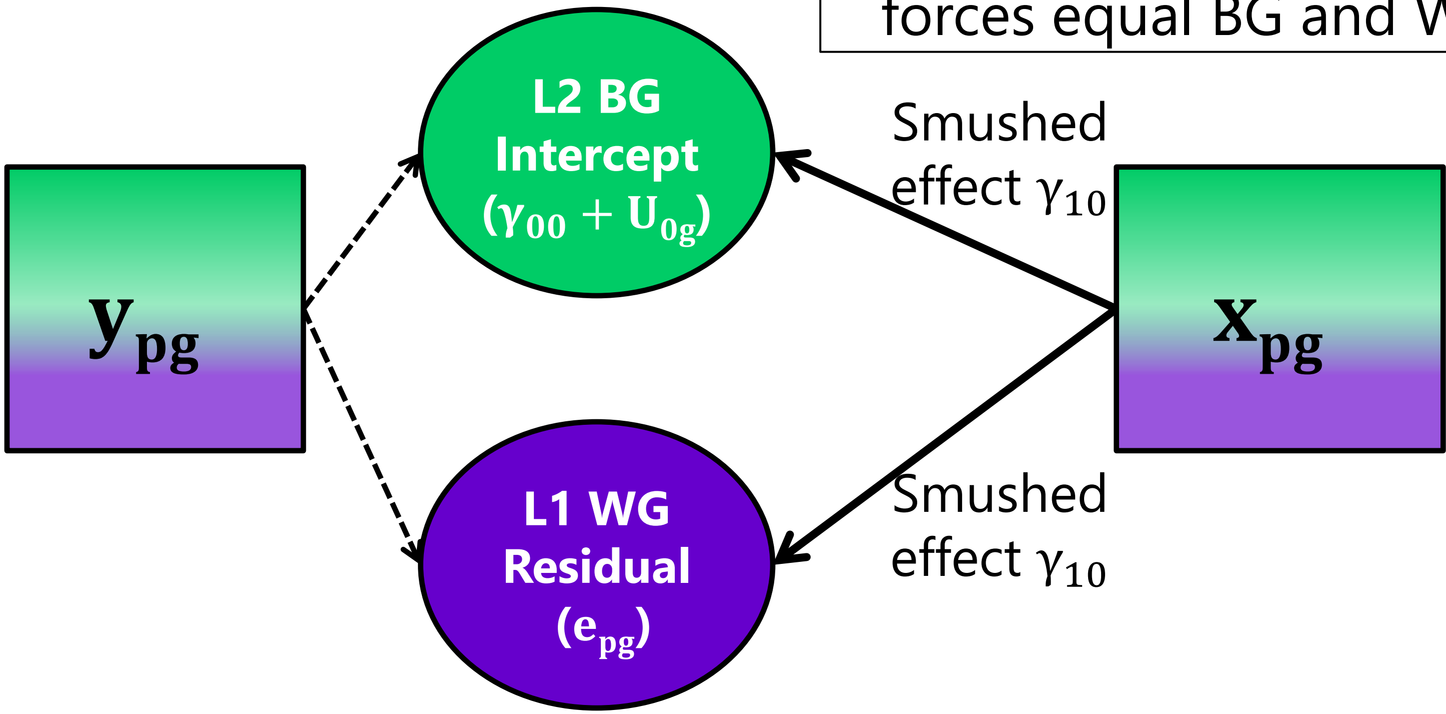
A *smushed* effect is also referred to as the *convergence*, *conflated*, or *composite* effect

- If the level-1 x has both Level-2 BG and Level-1 WG variation, then its one fixed effect has to do the work of two predictors

Grand-Mean-Centering: Smushing

Model-based partitioning of y_{pg} outcome variance into **variance components**:

Original x_{pg} has not been partitioned AND it has only **one fixed effect** coefficient in the model. Thus, that smushed effect forces equal BG and WG effects.



Smushed Effects of Level-1 Predictors

$$\text{Convergence Effect: } \gamma_{\text{conv}} \approx \frac{\dfrac{\gamma_{BG}}{SE_{BG}^2} + \dfrac{\gamma_{WG}}{SE_{WG}^2}}{\dfrac{1}{SE_{BG}^2} + \dfrac{1}{SE_{WG}^2}}$$

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

- **The smushed effect will often be closer to the within-group effect** (due to larger level-1 sample size and thus smaller SE), and thus the level-2 BG model will be much more affected by smushing
- **It is the rule, not the exception, that between and within effects differ** (Snijders & Bosker, 2012, p. 60, and personal experience!)
- This same problem 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

So How Do We Fix It Using Random Effects Instead? 3 Strategies

- Within univariate MLMs (like in SAS or STATA MIXED)
  1. Create new level-specific predictor variables: this is known as "**group-mean-centering**" (or variable-centering, to me more generally) → less likely to be screwed up, so we start here
  2. Rely on statistical partialing to unsmush the level-1 effects: this is known as "**grand-mean-centering**" (or constant-centering, to me more generally) → much more likely to be screwed up

- Within multivariate MLMs (estimated as single-level or multilevel structural equation models like in Mplus)
  3. Treat level-1 predictors the same way as level-1 outcomes: allow the model to partition their variance to become "latent" variables (i.e., level-2 random intercepts and level-1 residuals)

Group-Mean-Centering (Group-MC)

- In Group-MC, we partition the level-1 predictor x_{pg} into 2 variables that directly represent its BG (level-2) and WG (level-1) sources of variation, and **include these variables as the predictors instead**:

- **Level-2, Group Mean (GM) predictor = group mean of x_{pg}**
  - $\mathbf{GMx_g} = \bar{\mathbf{X}}_\mathbf{g} - \boldsymbol{C_2}$
  - $\mathrm{GMx_g}$ is centered at constant C_2, chosen for meaningful 0 (e.g., sample mean)
  - $\mathrm{GMx_g}$ is positive? Above sample mean → "more than other groups"
  - $\mathrm{GMx_g}$ is negative? Below sample mean → "less than other groups"

- **Level-1, WG predictor = deviation from group mean of x_{pg}**
  - $\mathbf{WGx_{pg}} = \mathbf{x_{pg}} - \bar{\mathbf{X}}_\mathbf{g}$ *(uncentered person mean $\bar{\boldsymbol{X}}_{\boldsymbol{g}}$ is used to center x_{pg})*
  - $\mathrm{WGx_{pg}}$ is NOT centered at a constant; it is centered at a VARIABLE
  - $\mathrm{WGx_{pg}}$ is positive? Above your group mean → "more than my others"
  - $\mathrm{WGx_{pg}}$ is negative? Below your group mean → "less than my others"

Clustered Data Model with **Group-Mean-Centered Level-1 x_{pg}**

→ WG and BG Effects directly through **separate** parameters

x_{pg} is group-mean-centered into WGx_{pg}, with GMx_g at L2:

Level 1: $y_{pg} = \beta_{0g} + \beta_{1g}(WGx_{pg}) + e_{pg}$

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

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

$WGx_{pg} = x_{pg} - \bar{X}_g$ → it has only Level-1 WG variation

$GMx_g = \bar{X}_g - C_2$ → it has only Level-2 BG variation

γ_{10} = WG main effect of having more x_{pg} than others in your group

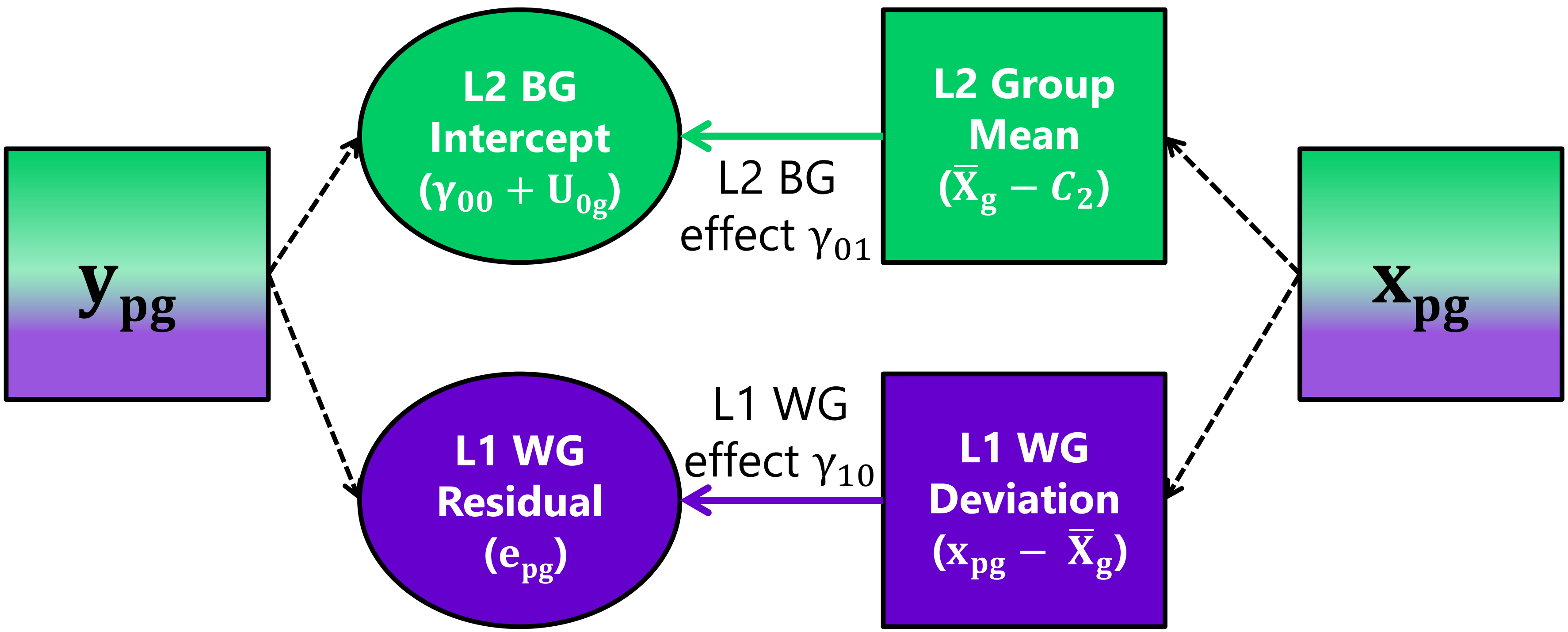
γ_{01} = BG main effect of having more $\bar{X}_g$ than other groups

Because WGx_{pg} and GMx_g are uncorrelated, each gets the total effect for its level (WG=L1, BG=L2)

Group-Mean-Centering

Model-based partitioning of y_{pg} outcome variance into **variance components**:

Brute-force partitioning of x_{pg} predictor variance into **observed variables**:



Why not let the model make variance components for x_{pg}, too?
This is the basis of multivariate MLM (or "multilevel SEM"): stay tuned…

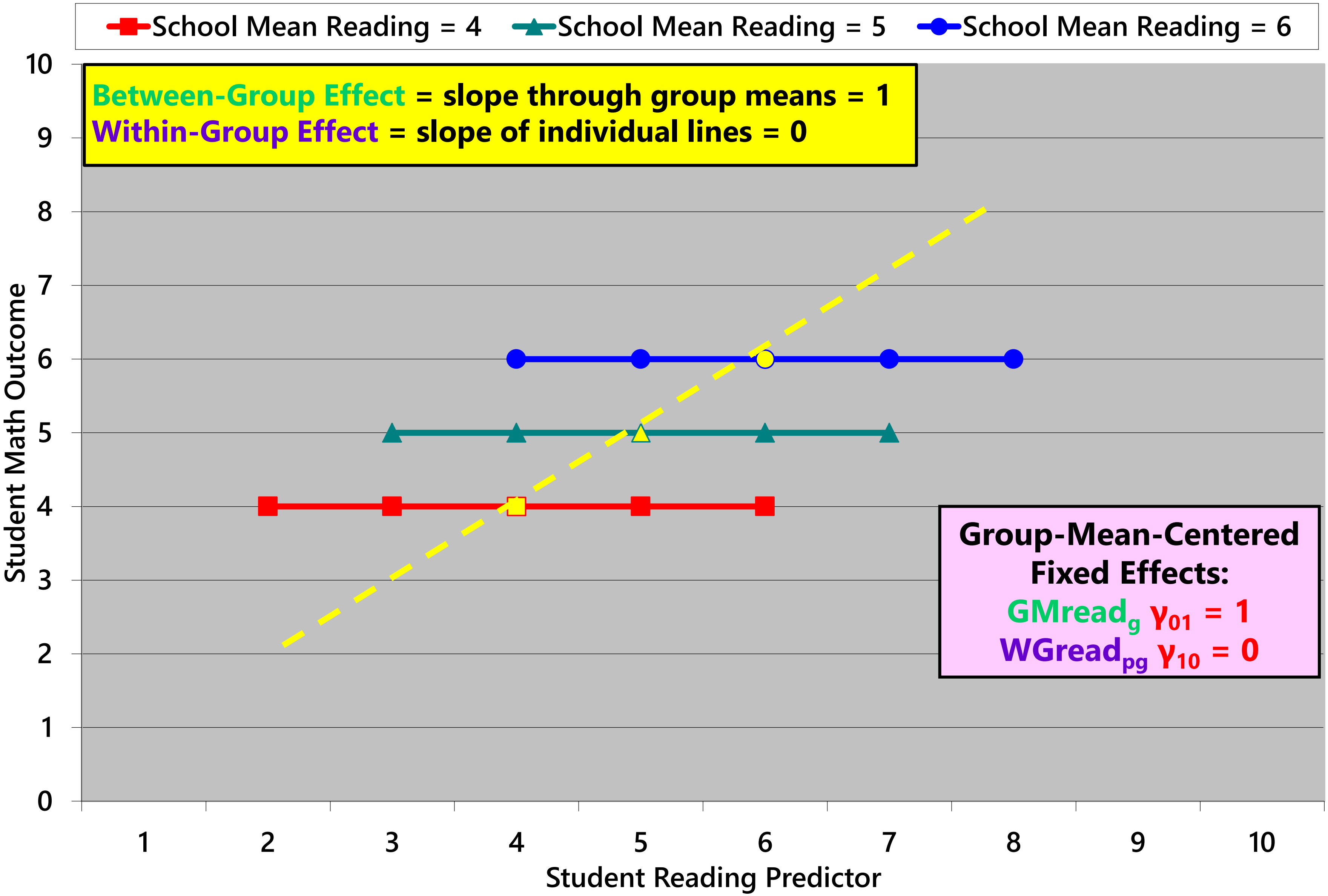
3 Kinds of Effects for Level-1 Predictors

- **What Group-Mean-Centering tells us directly:**

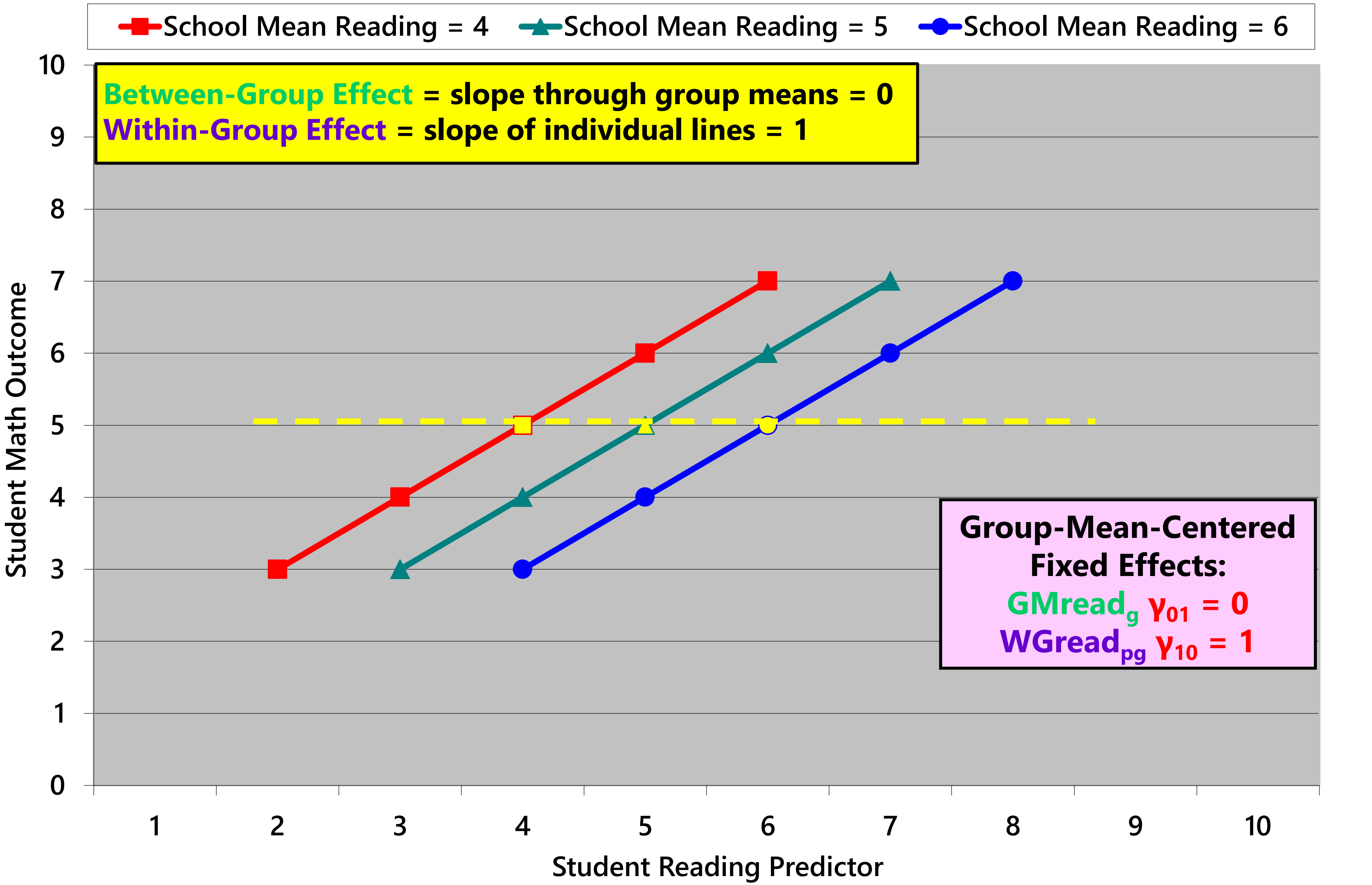
- **Is the Between-Group (BG) effect significant?**
  - Are groups with higher predictor values than other groups also higher on Y than other groups, such that the group mean of the person-level predictor $\mathbf{GMx_g}$ accounts for level-2 random intercept variance ($\tau^2_{U_0}$)?
  - This would be indicated by a significant fixed effect of $\mathbf{GMx_g}$
  - Note: this is NOT controlling for the absolute value of x_{pg} for each person

- **Is the Within-Group (WG) effect significant?**
  - If you have higher predictor values than others in your group, do you also have higher outcomes values than others in your group, such that the within-group deviation $\mathbf{WGx_{pg}}$ accounts for level-1 residual variance (σ^2_e)?
  - This would be indicated by a significant fixed effect of $\mathbf{WGx_{pg}}$
  - Note: this is represented by the relative value of x_{pg} NOT the actual value of x_{pg}

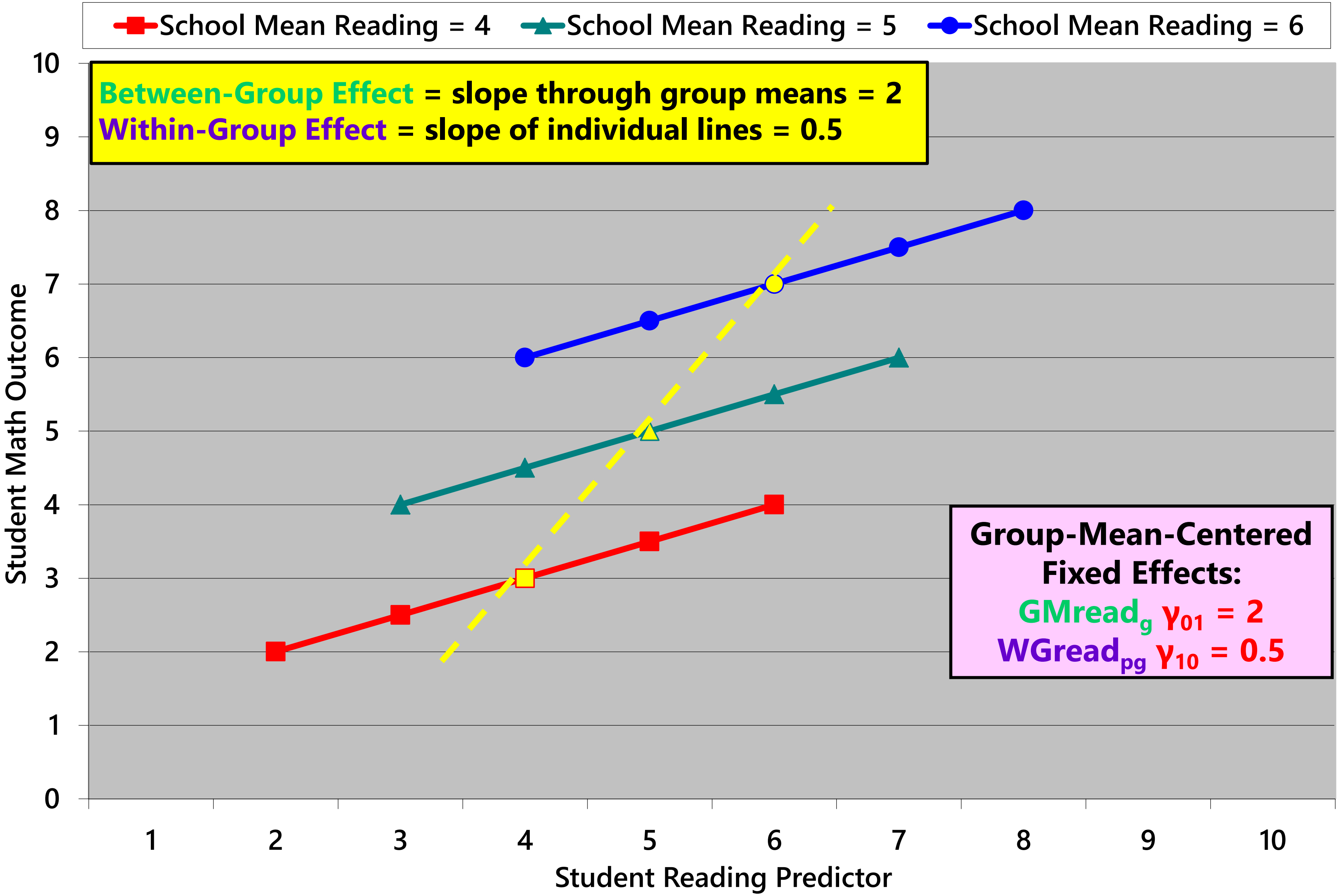
ALL Between-Group Effect, NO Within-Group Effect



NO Between-Group Effect, ALL Within-Group Effect



Between-Group Effect > Within-Group Effect



3 Kinds of Effects for Level-1 Predictors

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

- **Are the BG and WG effects different sizes: Is there a contextual effect?**
  - After controlling for the absolute value of the level-1 predictor for each person, is there still an incremental contribution from the predictor's group mean (i.e., does a group's general tendency predict $\tau^2_{U_0}$ above and beyond just the person-specific value of the predictor)?
  - In clustered data, the contextual effect is phrased as "after controlling for the individual, what is the additional contribution of the group"?

- **To answer this question about the contextual effect for the incremental contribution of the group mean, we have two options:**
  - Ask for the contextual effect via an ESTIMATE statement in SAS (or TEST in SPSS, or NEW in Mplus, or LINCOM in STATA): `WGx −1 GMx 1`
  - Use "**grand-mean-centering**" for level-1 x_{pg} instead: $\mathbf{L1x_{pg} = x_{pg} - \mathit{C}_1}$

    → **centered at a CONSTANT, NOT A LEVEL-2 VARIABLE**
    - Which constant only matters for what the reference point is; it could be the grand mean or other

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

- In this model: $Y_i = \beta_0 + \beta_1(X_{1i}) + \beta_2(X_{2i}) + e_i$
- If X_{1i} and X_{2i} **ARE NOT** correlated:
  - β_1 is **ALL the relationship** between X_{1i} and Y_i
  - β_2 is **ALL the relationship** between X_{2i} and Y_i

- If X_{1i} and X_{2i} **ARE** correlated:
  - β_1 is **different than** the full relationship between X_{1i} and Y_i
    - "Unique" effect of X_{1i} *controlling for* X_{2i} or *holding* X_{2i} *constant*
  - β_2 is **different than** the full relationship between X_{2i} and Y_i
    - "Unique" effect of X_{2i} *controlling for* X_{1i} or *holding* X_{1i} *constant*

- Hang onto that idea…

Group-MC vs. Grand-MC for Level-1 Predictors

Level 2		Original	Group-MC Level 1	Grand-MC Level 1
$\bar{X}_g$	$GMx_g = \bar{X}_g - 5$	x_{pg}	$WGx_{pg} = x_{pg} - \bar{X}_g$	$L1x_{pg} = x_{pg} - 5$
3	−2	2	−1	−3
3	−2	4	1	−1
7	2	6	−1	1
7	2	8	1	3

Same GMx_g goes into the model using either way of centering the level-1 variable x_{pg}

Using **Group-MC**, WGx_{pg} has NO level-2 BG variation, so it is not correlated with GMx_g

Using **Grand-MC**, $L1x_{pg}$ STILL has level-2 BG variation, so it is STILL CORRELATED with GMx_g

So the effects of GMx_g and $L1x_{pg}$ when included together under Grand-MC will be different than their effects would be if they were by themselves…

Clustered Data Model with **Grand-Mean-Centered Level-1 x_{pg}**

→ Model tests difference of WG vs. BG effects (It's been fixed!)

x_{pg} is grand-mean-centered into $L1x_{pg}$, <u>WITH</u> GMx_g at L2:

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

$L1x_{pg} = x_{pg} - C_1$ → it still has both Level-2 BG and Level-1 WG variation

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

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

$GMx_g = \bar{X}_g - C_2$ → it has only Level-2 BG variation

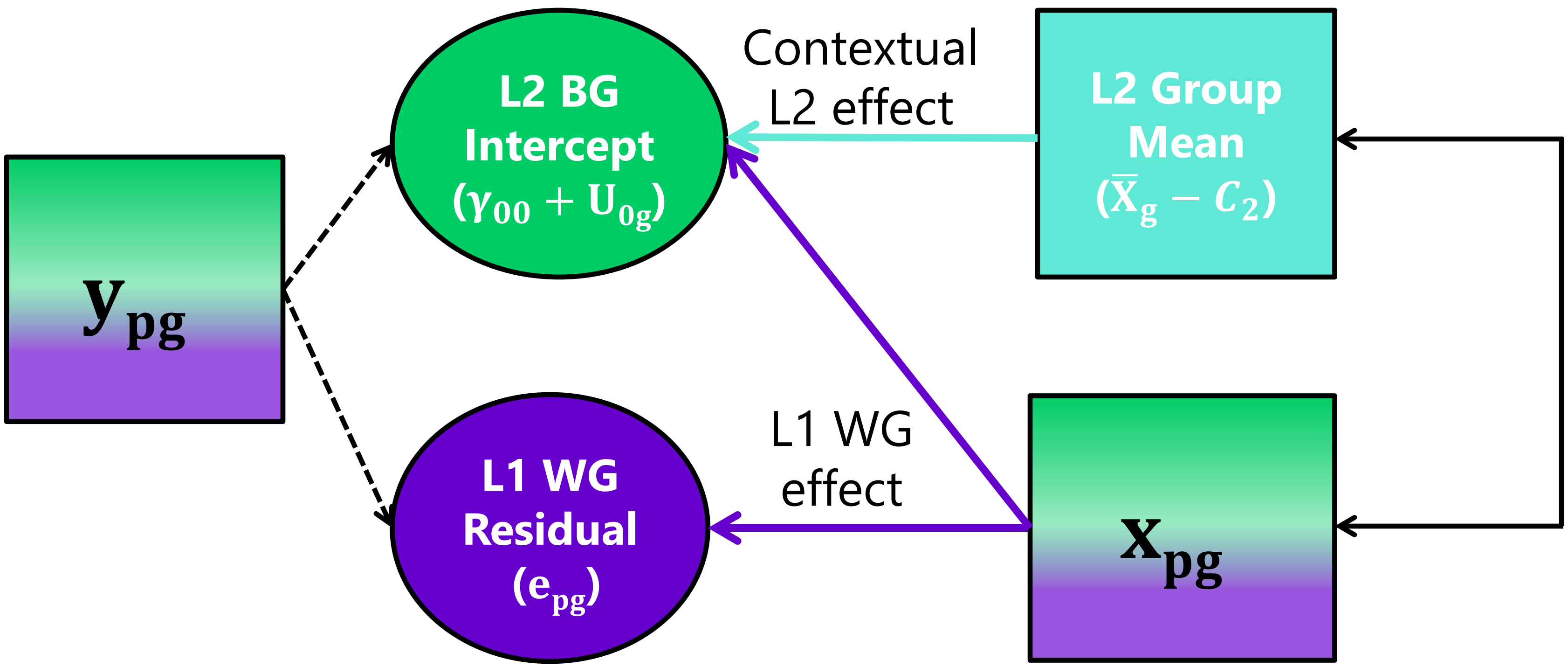
γ_{10} becomes the WG effect → *unique* level-1 effect after controlling for GMx_g

γ_{01} *becomes* the contextual effect that indicates <u>how the BG effect differs from the WG effect</u>
→ *unique* level-2 effect after controlling for $L1x_{pg}$
→ does group matter beyond individuals?

Grand-Mean-Centering + L2 GMx

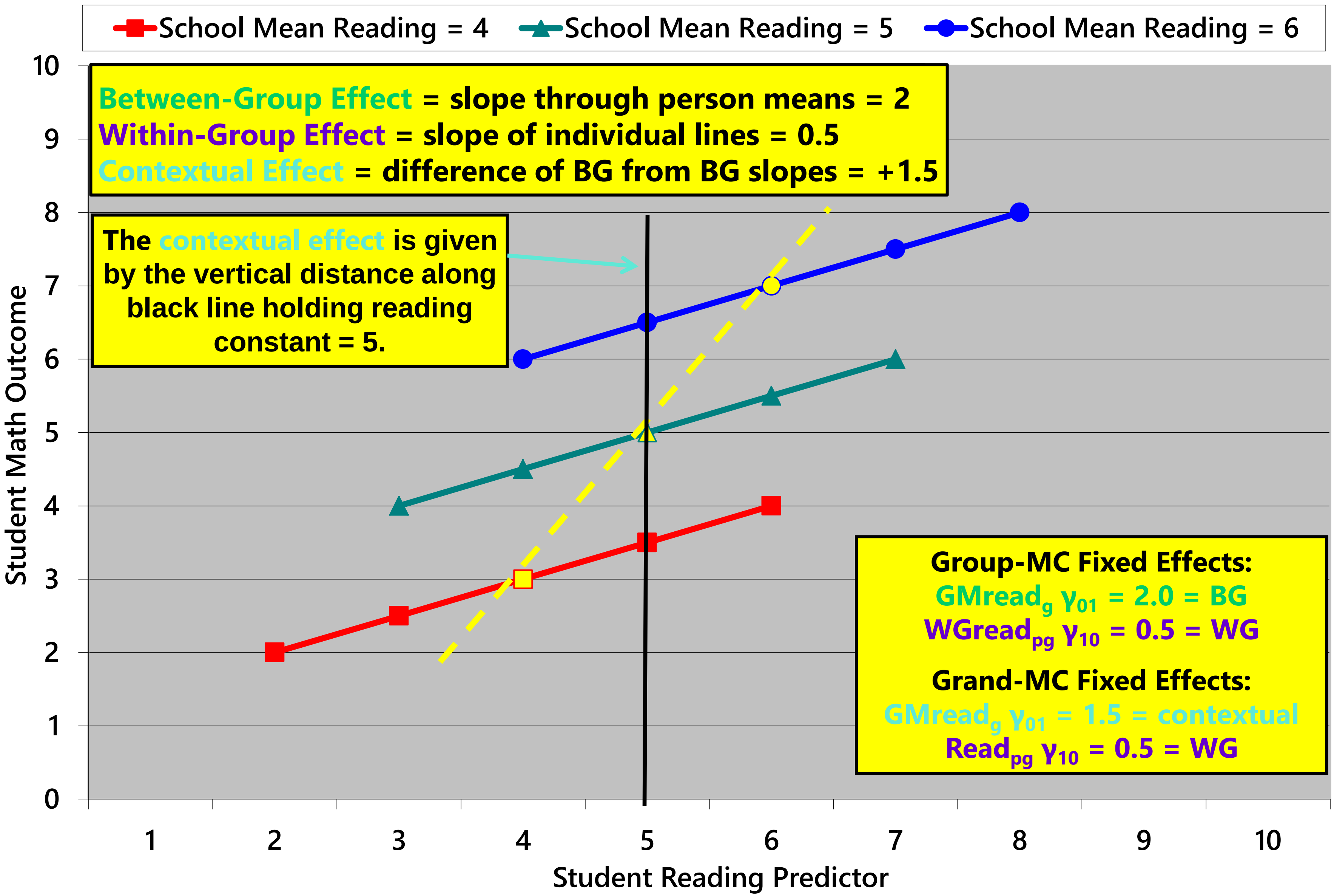
Model-based partitioning of y_{pg} outcome variance into **variance components**:

Original x_{pg} is not partitioned, but group mean $\mathbf{\bar{X}_g - \mathit{C}_1}$ is added to allow an extra (different) effect at L2



Because original x_{pg} still has BG variance, it still carries part of the BG effect…

Between, Within, and Contextual Effects



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

Group-MC: $WGx_{pg} = x_{pg} - GMx_g$

Level-1: $y_{pg} = \beta_{0g} + \beta_{1g}(x_{pg} - GMx_g) + e_{pg}$

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

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

$\rightarrow y_{pg} = \gamma_{00} + \gamma_{01}(GMx_g) + \gamma_{10}(x_{pg} - GMx_g) + U_{0g} + e_{pg}$

$\rightarrow y_{pg} = \gamma_{00} + (\gamma_{01} - \gamma_{10})(GMx_g) + \gamma_{10}(x_g) + U_{0g} + e_{pg}$

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

Effect	Group-MC	Grand-MC
Intercept	γ_{00}	γ_{00}
WG Effect	γ_{10}	γ_{10}
Contextual	$\gamma_{01} - \gamma_{10}$	γ_{01}
BG Effect	γ_{01}	$\gamma_{01} + \gamma_{10}$

Grand-MC: $L1x_{pg} = x_{pg}$

Level-1: $y_{pg} = \beta_{0g} + \beta_{1g}(x_{pg}) + e_{pg}$

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

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

$\rightarrow y_{pg} = \gamma_{00} + \gamma_{01}(GMx_g) + \gamma_{10}(x_{pg}) + U_{0g} + e_{pg}$

Another Example of Contextual Effects

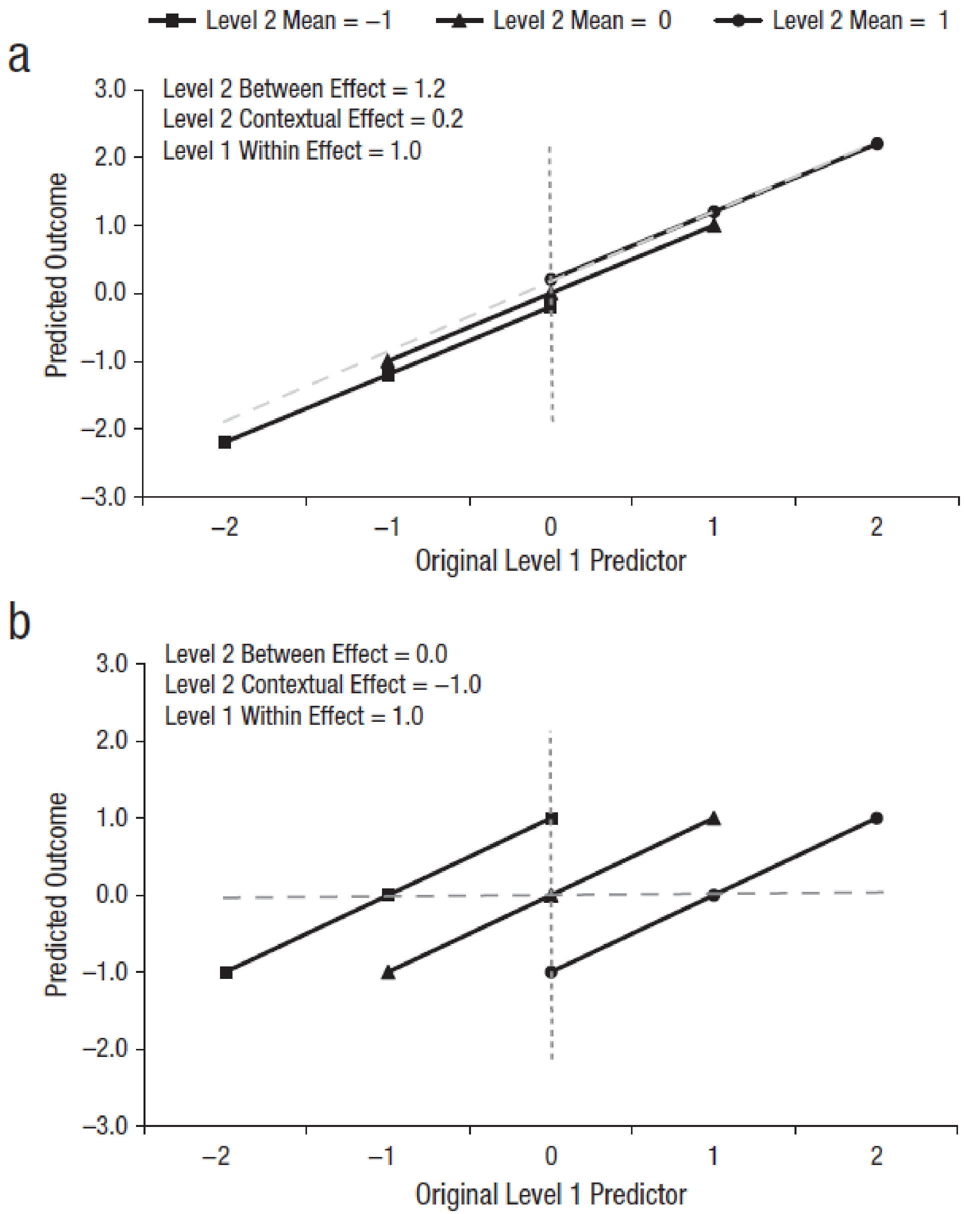
- Group-MC is equivalent to Grand-MC if the group mean of the level-1 predictor is included (and the level-1 effect is not random, stay tuned)
- Grand-MC may be more convenient in clustered data due to its ability to directly provide level-2 contextual effects (of group controlling for person)

- Example: Effect of SES for students (nested in schools) on achievement:

- **Group-MC** of level-1 student SES_{pg} , school mean $\overline{SES}_g$ included at level 2
  - Level-1 **WG** effect: Effect of being rich kid relative to your school (is already purely WG because of centering around $\overline{SES}_g$)
  - Level-2 **BG** effect: Effect of going to a rich school NOT controlling for kid SES_{pg}

- **Grand-MC** of level-1 student SES_{pg} , school mean $\overline{SES}_g$ included at level 2
  - Level-1 **WG** effect: Effect of being rich kid relative to your school (is purely WG after *statistically* controlling for $\overline{SES}_g$)
  - Level-2 **Contextual** effect: Incremental effect of going to a rich school (after *statistically* controlling for student SES)

Between vs. Contextual Effects



- Image from Hoffman (2019)
- *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*
- *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

Other Reasons to Choose Grand-MC

- Grand-MC for level-1 predictors creates level-2 contextual effects for the cluster mean instead of level-2 between-cluster effects. This is preferable in 3 instances:
- **When you really do want level-2 contextual effects**
  - The incremental effect of the group-level predictor (cluster mean) after controlling for the person's absolute (not relative) predictor
- When the group-MC for the level-1 predictor doesn't make sense conceptually, such as for **categorical level-1 predictors**
  - e.g., 0-1 predictors when group-MC become impossible values
- **When the cluster mean is not a reliable group-level predictor**
  - The sample of persons within groups is not complete enough to form a cluster mean that would be useful
  - Using externally-provided info does a better job of representing the group
  - When you want to control for the same *construct* at the group level, but cannot use the same variable to do so, group-MC does not make sense, and grand-MC would be preferred (just please don't smush)

Pseudo-R^2 for Level-1 Predictors

- Level-1 main effects and interactions (i.e., among level-1 predictors only) will reduce level-1 residual variance

$$\text{Pseudo } R_e^2 = \frac{\text{residual variance}_{\text{fewer}} - \text{residual variance}_{\text{more}}}{\text{residual variance}_{\text{fewer}}}$$

- **However—what happens to the level-2 variance depends on what kind of variance the level-1 predictor has:**
  - If the level-1 predictor ALSO has level-2 variance (e.g., Grand-MC predictors), then its level-2 variance will also likely reduce level-2 random intercept variance (in combination with the contextual effect)
  - If the level-1 predictor DOES NOT have level-2 variance (e.g., Group-MC predictors), then its reduction in the level-1 residual variance will cause an **INCREASE in level-2 random intercept variance**
    - **Say what????** Btw, this is why you are never "safe" from ignoring clustering!
    - This is why I enter the L1 and L2 orthogonal parts of a predictor into the model at the same time: to avoid this artificial increase in L2 variance (and subsequent pseudo-R^2)

Increases in Random Intercept Variance

- Level-2 random intercept variance $\tau^2_{U_0}$ will increase as a consequence of reducing level-1 residual variance σ^2_e
- Observed level-2 $\tau^2_{U_0}$ is NOT just between-group variance
  - Also has a small part of within-group variance (level-1 σ^2_e), or: **Observed $\tau^2_{U_0}$ = True $\tau^2_{U_0}$ + (σ^2_e/*n*)** where *n* = level-1 sample size
  - Likelihood-based estimates of "true" $\tau^2_{U_0}$ use (σ^2_e/*n*) as correction factor: **True $\tau^2_{U_0}$ = Observed $\tau^2_{U_0}$ − (σ^2_e/*n*)**
- For example: observed level-2 $\tau^2_{U_0}$=4.65, level-1 σ^2_e=7.06, *n*=4
  - True $\tau^2_{U_0}$= 4.65 −(**7.06**/4) = **2.88** in empty means model
  - Add fixed level-1 slope → reduce σ^2_e from 7.06 to 2.17 (Pseudo-R^2_e = .69)
  - But now True $\tau^2_{U_0}$= 4.65 −(**2.17**/4) = **4.10** in new model
  - So the **ICC** (and thus the **design effect**) is higher than it was originally…

Summary: Fixed Effects of Level-1 Predictors

- **Is the Between-Group (BG) effect significant?**
  - Are groups with higher predictor values than other groups also higher on Y than other groups, such that the group mean of the person-level predictor $\mathbf{GMx_g}$ accounts for level-2 random intercept variance ($\tau^2_{U_0}$)?
  - Given directly by level-2 effect of GMx_g if using Group-MC for the level-1 predictor (or can be requested via ESTIMATE if using Grand-MC for the level-1 predictor)

- **Is the Within-Group (WG) effect significant?**
  - If you have higher predictor values than others in your group, do you also have higher outcomes values than others in your group, such that the within-group deviation $\mathbf{WGx_{pg}}$ accounts for level-1 residual variance (σ^2_e)?
  - Given directly by the level-1 effect of WGx_{pg} if using Group-MC —*OR* — given directly by the level-1 effect of $L1x_{pg}$ if using Grand-MC and including GMx_g at level 2 (without GMx_g, the level-1 effect of $L1x_{pg}$ if using Grand-MC is the smushed effect)

- **Are the BG and WG effects different sizes: Is there a contextual effect?**
  - After controlling for the absolute value of the level-1 predictor for each person, is there still an incremental contribution from the group mean of the predictor (i.e., does a group's general tendency predict $\tau^2_{U_0}$ above and beyond the person-specific predictor value)?
  - Given directly by level-2 effect of GMx_g if using Grand-MC for the level-1 predictor (or can be requested via ESTIMATE if using Group-MC for the level-1 predictor)