MLDS Center
Maryland Longitudinal Data System
Better Data • Informed Choices • Improved Results

Clustered Data: Are Multilevel Models Really Necessary?

Dan McNeish & Laura Stapleton
University of Maryland, College Park

Outline

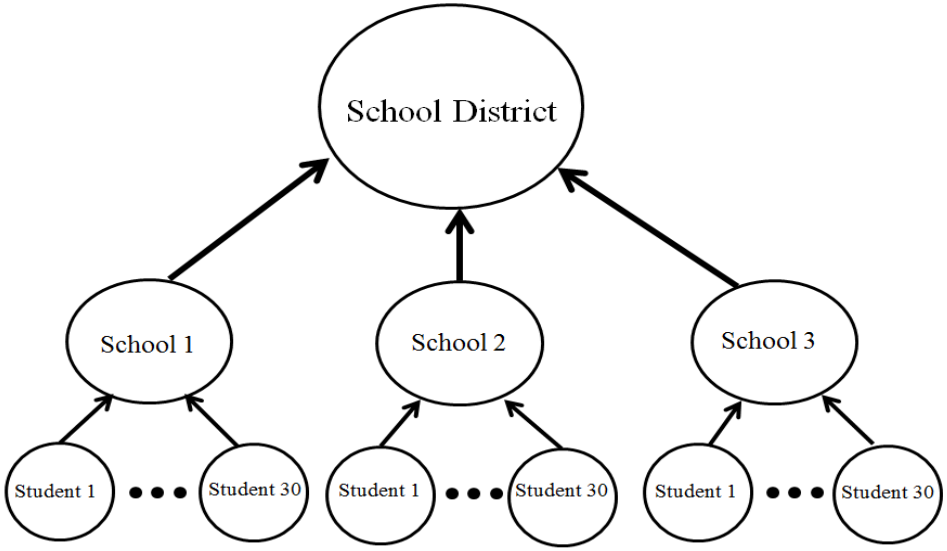
- Overview of Clustered Data and Multilevel Models
- Proliferation of Multilevel Models
- Details and Differences in Methods for Clustered Data
- Application to a Reading Intervention Study
- Implications

MLDS Center
Maryland Longitudinal Data System
Better Data • Informed Choices • Improved Results

Overview of Clustered Data and Multilevel Models

Clustered Data in Education

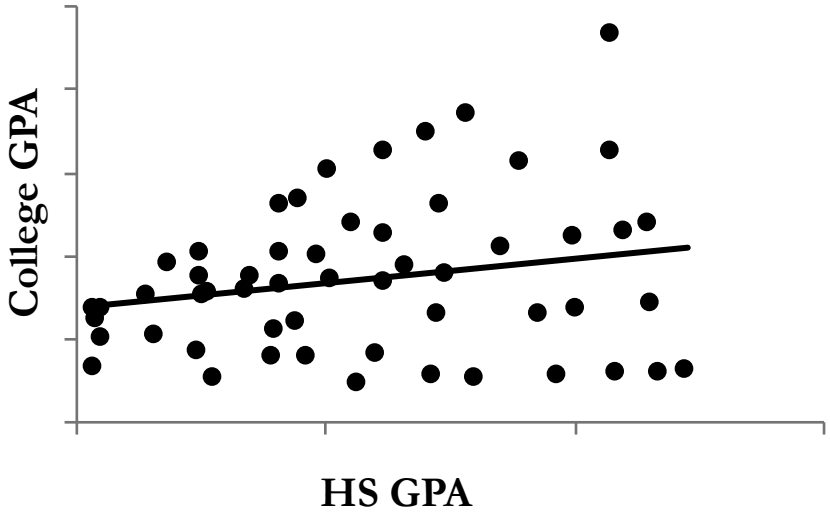
- Clustered data are more often the rule than the exception in education
  - Students clustered within classrooms
  - Classrooms clustered within schools
  - Schools clustered within districts



If the analysis doesn't account for the clustered structure, the model estimates will not be trustworthy

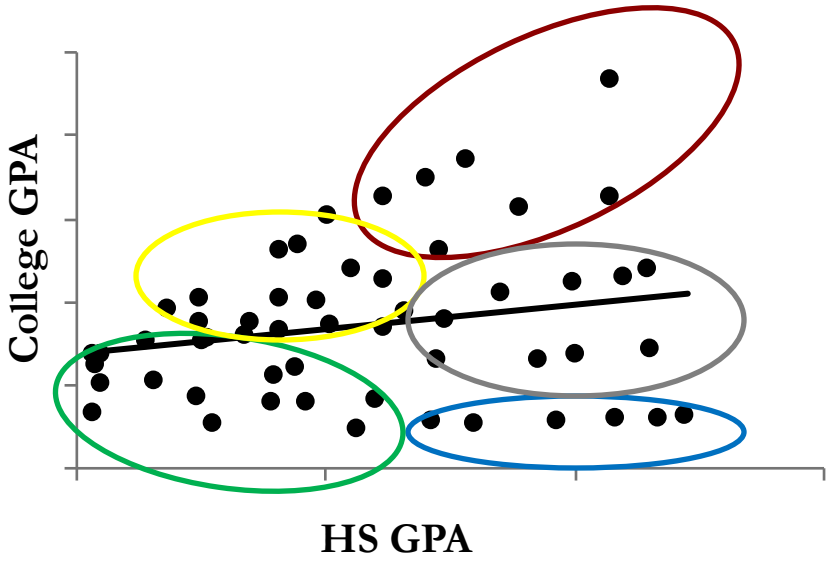
Accounting for Clustering

- Traditional statistical models assume that individuals are not systematically related to one another
  - Not the case if data are clustered
  - Kids in the same school are more like each other than kids in other schools, etc.



Accounting for Clustering

- Traditional statistical models assume that individuals are not systematically related to one another
  - Not the case if data are clustered
  - Kids in the same school are more like each other than kids in other schools, etc.

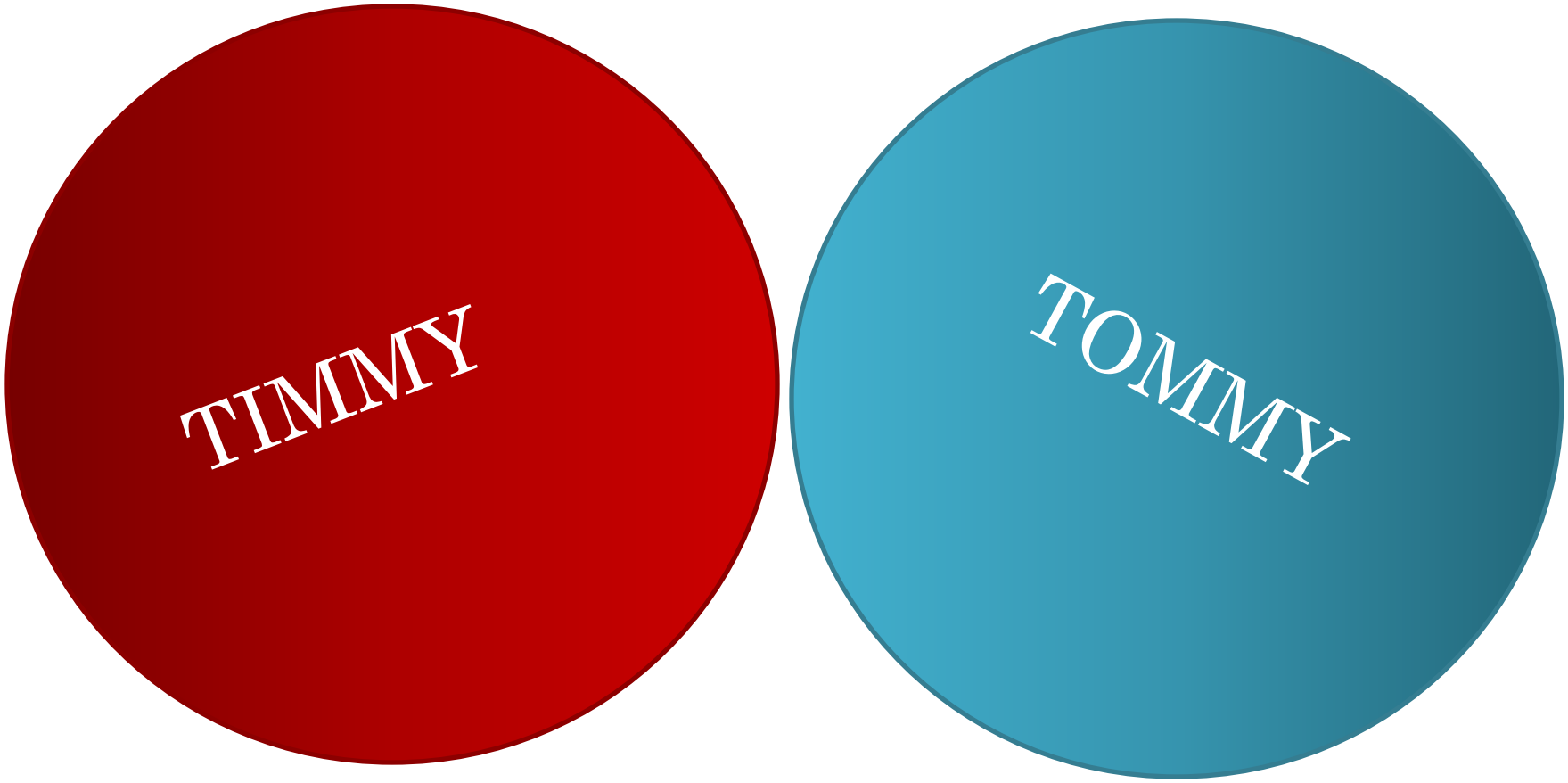


Accounting for Clustering II

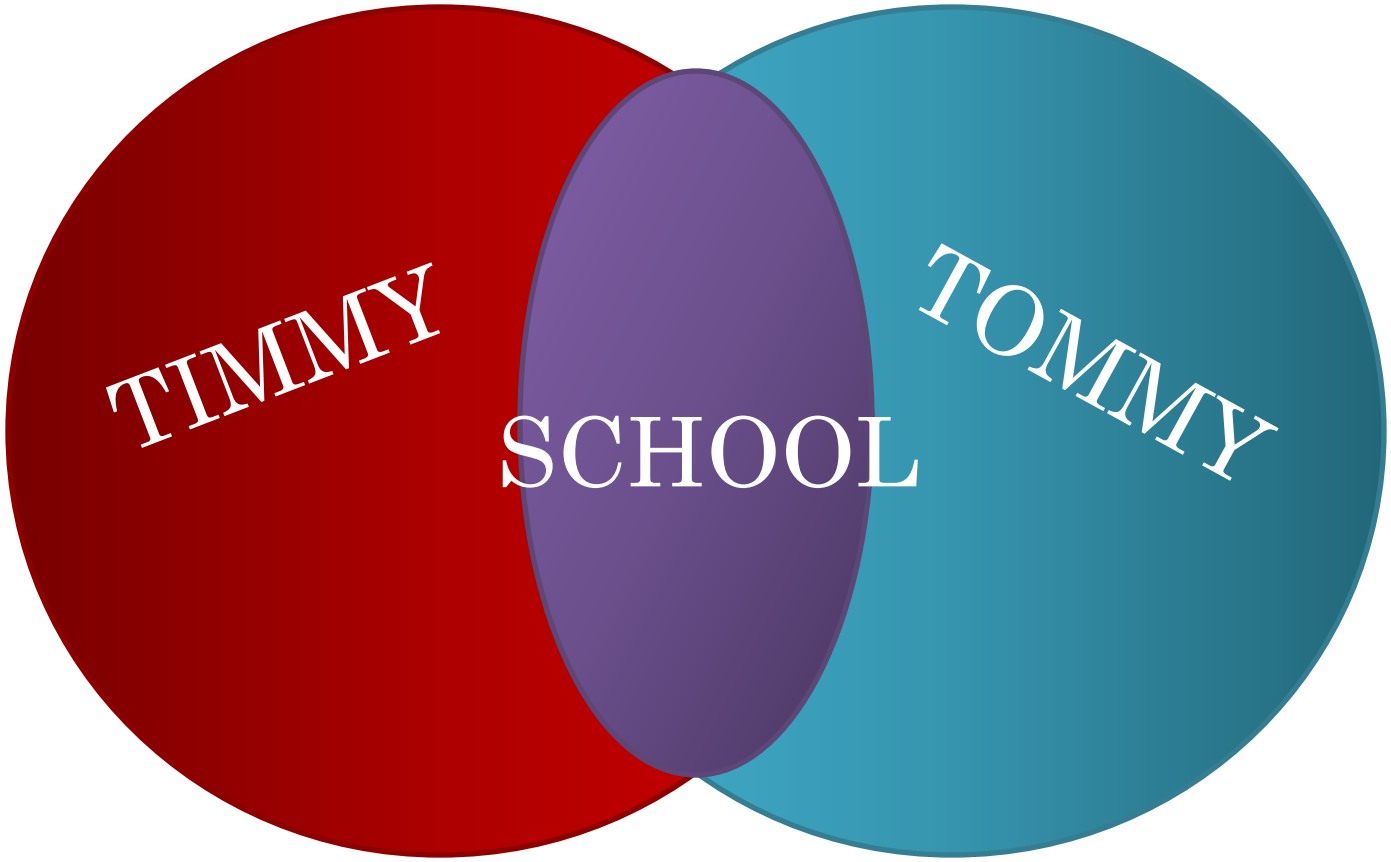
- The Issue:

- Dependencies within clusters reduce the amount of unique information available for a given sample size
  - If I know Timmy's reading score and know that Tommy is in the same school as Timmy, I already know a little bit about Tommy's score too
  - The unique contribution of each student is reduced because some information is shared by knowing which school they are in

What Traditional Model Assumes

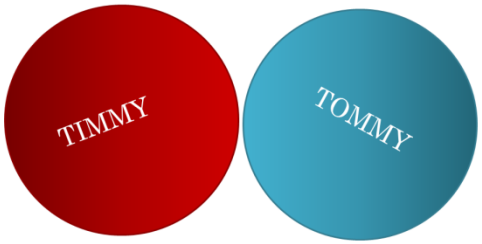


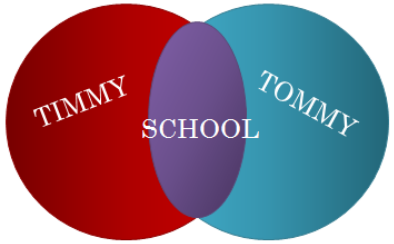
What Happens With Clustering



Accounting for Clustering III

- The total sample size appears in the denominator of formulas for calculating precision in single-level models
  - Using the total sample size assumes that each individual contributes unique information
  - This is not the case with clustering, so the denominator will be too large
- In English – ignoring the clustering makes the model appear more precise that it is



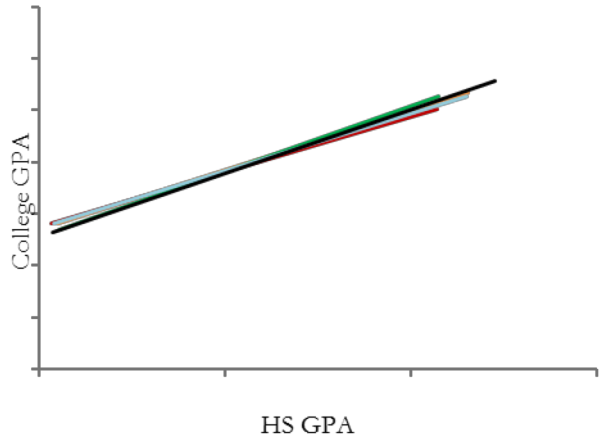


Multilevel Models

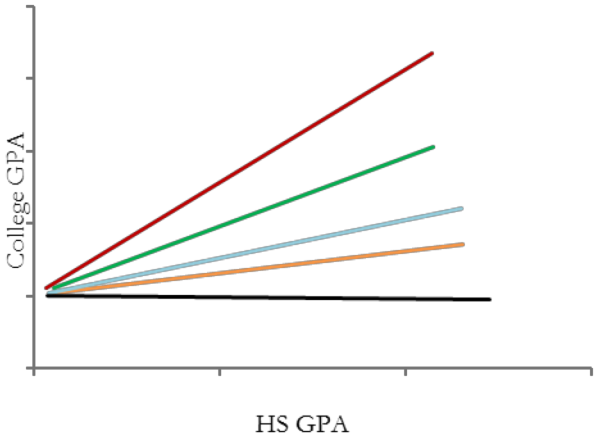
- The predominant method to account for clustering in education research is through multilevel models (MLMs)

- MLMs take the clustering into account primarily by estimating *random effects* for each cluster
  - Random effects allow the prediction model for each cluster to be different from the prediction of the overall sample

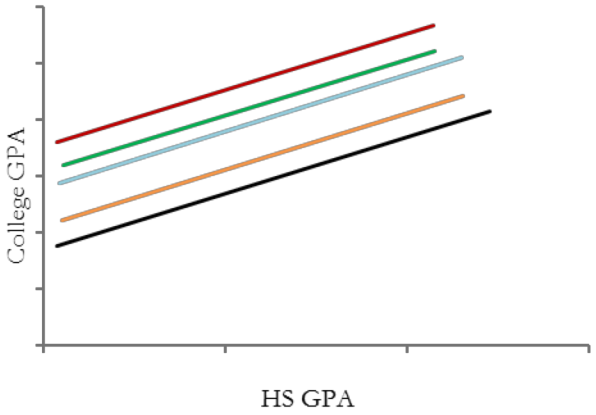
Random Effects



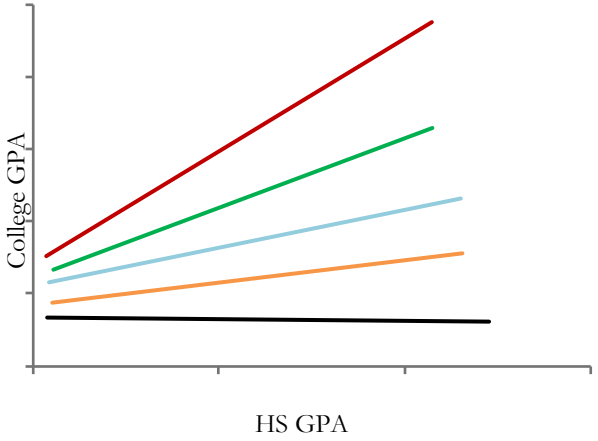
No Random Effects



Random Slopes

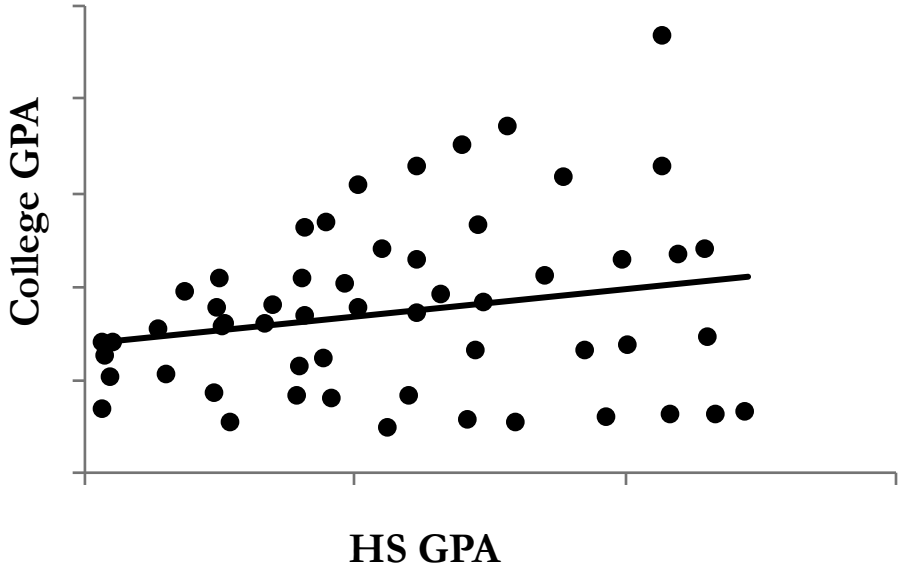


Random Intercepts



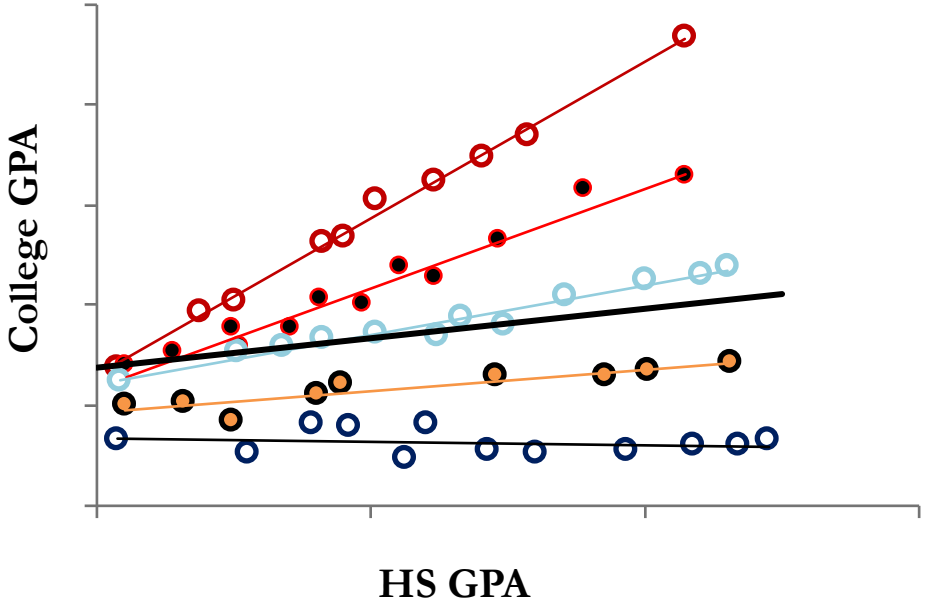
Random Intercepts and Random Slopes

Multilevel Models Graphically

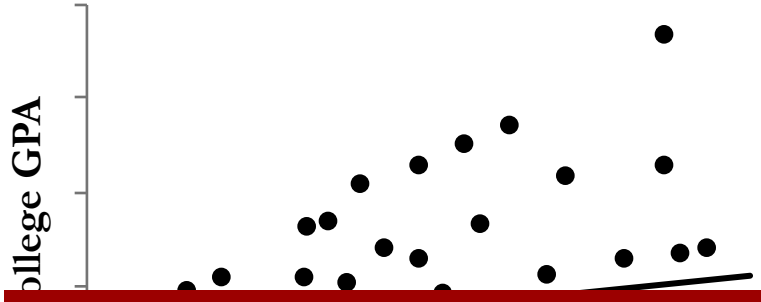


Overall regression line for 5 clusters of data (ignoring clustering)

5 cluster-specific regression lines with overall regression in black.



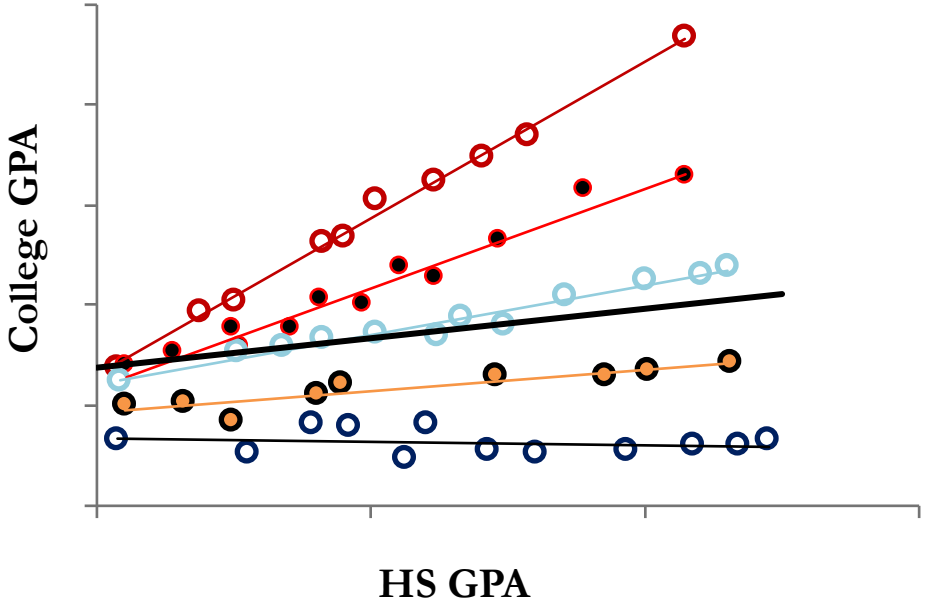
Multilevel Models Graphically



gression line
ers of data
lustering)

MLMs use an overall prediction line but allow the intercept and slope of the line to change for each cluster. The change in the cluster-specific line from the overall line can be quantified and helps to estimate precision more appropriately

5 cluster-specific regression lines with overall regression in black.



MLDS Center
Maryland Longitudinal Data System
Better Data • Informed Choices • Improved Results

Proliferation of Multilevel Models

Dominance of MLMs

- MLMs are overwhelming popular to accommodate clustered data in the behavioral sciences

- Found that from 2006 to 2011, 94% of published studies accounted for clustering using MLMs
  - Found only 14 studies (6%) that used design-based methods (DBMs)

National Assessment of Educational Progress

A Closer Look at Charter Schools Using Hierarchical Linear Modeling

ANALYZING FACULTY WORKLOAD DATA USING MULTILEVEL MODELING

Stephen R. Porter, Ph.D., Office of Institutional Research, Wesleyan University. Paul D. Umbach, Department of Education Policy and Leadership, University of Maryland.

Measuring School Effects With Hierarchical Linear Modeling: Data Handling and Modeling Issues

Eugene P. Adcock, Ph.D., Prince George's County Public Schools, Maryland
Gary W. Phillips, Ph.D. National Center for Education Statistics, U.S. Department of Education

Dropping Out of Middle School: A Multilevel Analysis of Students and Schools

American Educational Research Journal

Using Multi-level Analyses to Study the Effectiveness of Science Curriculum Materials

Vasuki Rethinam
Montgomery County Public Schools, USA

Curtis Pyke and Sharon Lynch
Graduate School of Education and Human Development, The George Washington University, USA

RESEARCH REPORT

Is there a "school effect" on pupil outcomes? A review of multilevel studies

A Multilevel Analysis of Students and Schools on High School Graduation Exam: A Case of Maryland

What is Taught in Graduate School?

- Previous study suggested that MLMs are more relevant to behavioral scientists because they are used most often
- There is some evidence that they are used because of tradition and lack of exposure to alternatives
- Reviewed 51 graduate school syllabi for courses specifically focused on methods for clustered data
  - e.g., Longitudinal, complex survey data, cross-sectionally clustered
  - Convenience sample, so generalizability not fully warranted
  - Spanned 8 academic disciplines
    - Economics notably absent because material is embedded within broader econometrics sequences

What is Taught in Graduate School?

Discipline	DBMs Only	MLM Only	MLM and DBMs	Total
Biostatistics	0	1	8	9
Criminology	0	1	0	1
Education	0	8	2	10
Political Sci.	1	2	0	3
Psychology	0	9	0	9
Public Health	1	0	5	6
Sociology	1	1	2	4
Statistics	0	2	7	9

- DBMs= Design-Based Methods (GEE, Cluster Robust Errors, Survey Methods)
- MLMs = Multilevel Models

What is Taught in Graduate School?

Discipline	DBMs Only	MLM Only	MLM and DBMs	Total
Biostatistics	0	1	**8**	**9**
Criminology	0	1	0	1
Education	0	8	2	10
Political Sci.	1	2	0	3
Psychology	0	9	0	9
Public Health	1	0	**5**	**6**
Sociology	1	1	2	4
Statistics	0	2	**7**	**9**

- DBMs= Design-Based Methods (GEE, Cluster Robust Errors, Survey Methods)
- MLMs = Multilevel Models

What is Taught in Graduate School?

Discipline	DBMs Only	MLM Only	MLM and DBMs	Total
Biostatistics	0	1	8	9
Criminology	0	1	0	1
Education	0	**8**	2	**10**
Political Sci.	1	2	0	3
Psychology	0	**9**	0	**9**
Public Health	1	0	5	6
Sociology	1	1	2	4
Statistics	0	2	7	9

- Most disciplines provide coverage of both DBMs and MLMs
- However, Education and Psychology very often exclusively teach MLMs
  - Not surprising to find that 94% of psychology studies use MLMs

Details and Differences in Methods for Clustered Data

Random Effects in MLMs

- The ability to quantify how much each cluster varies from the overall prediction sounds highly advantageous

- Caveat – Researchers must explicitly select which random effects to include and how they related to each other.
  - Do schools start at about the same values or do they differ?
  - Do students grow at similar rates or is there a lot of variation?
  - Do student characteristics have the same effect in all schools?
  - Does where students start affect how they grow over time?

Random Effects in MLMs II

- These decisions are not easy and, because of the nature of random effects, there are not well-developed statistical methods to determine if one selected correctly

- Worse yet, incorrect selections will introduce bias into both the values of the estimates and estimates of their precision.
  - All relevant random effects must be included
  - All irrelevant random effects must be excluded
  - The structure (covariance matrix) of the random effects must be correct.

MLM Assumptions

- MLMs also make several other assumptions, including:

MLM Assumptions
1. All relevant predictors are included
2. All relevant random effects are included
3. The covariance structure of the within-cluster residuals is correct
4. The covariance structure of the random effects is correct
5. The within-cluster residuals follow a (multivariate) normal distribution
6. The random effects follow a (multivariate) normal distribution
7. The random effects are not correlated with any predictors in the model
8. The sample size is sufficiently large for asymptotic inferences at all levels

Overview of Design-Based Methods

- Rather than explicitly model the clustering mechanism, DBMs treat the clustering mechanism as a nuisance
  - Aim to account for clustering rather than explain it
- DBMs treat the model as if it were a single level model, then apply statistical corrections to reflect that the data are clustered
- No random effects are included so models require fewer assumptions, are simpler to specify, and have more straightforward interpretations

DBM Assumptions

- Because DBMs do not require random effects or explicit modeling of the covariance structure(s), far fewer assumptions are required

DBM Assumptions
1. All relevant predictors are included
2. The sample size is sufficiently large for asymptotic inferences at the cluster level
3. Observations between clusters are not related
4. The working correlation matrix is "reasonably close" to the population structure

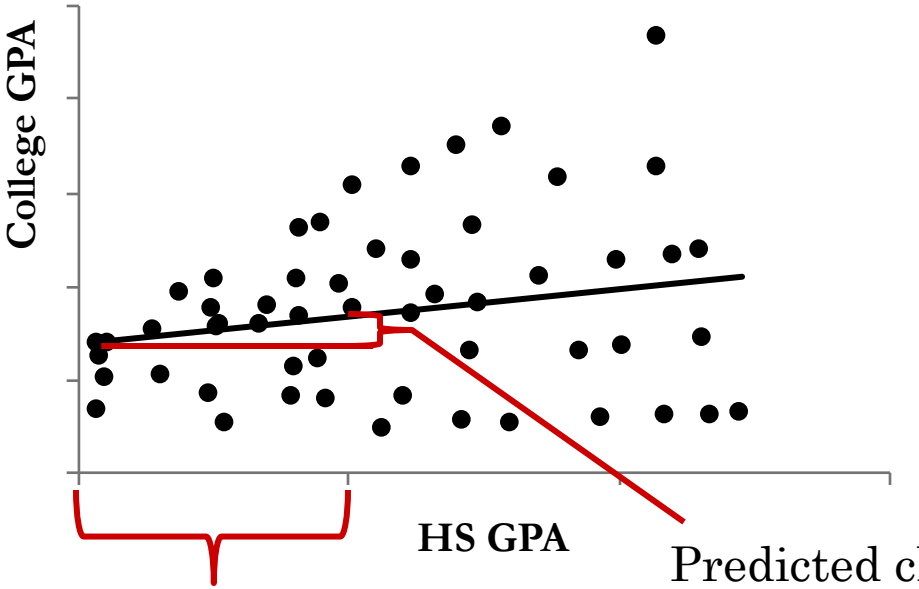
Difference in Coefficient Interpretation

- In a single-level model, the regression coefficients are generally interpreted as

"For a one-unit change in X, Y is predicted to change by β units, holding all other predictors constant"

  - Referred to as "population-averaged" interpretation because every individual in the data shares the same prediction model
  - With DBMs, this interpretation is preserved and the coefficients have a population-averaged interpretation

Back to This Picture

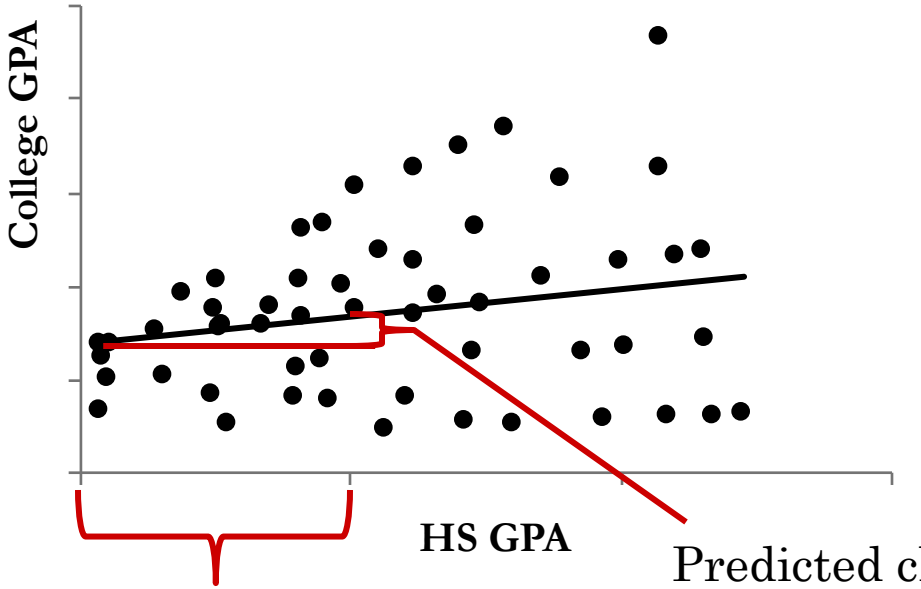


DBM coefficients are interpreted based on the whole sample

Difference in Coefficient Interpretation II

- However, this is **not** how coefficients are interpreted with MLMs
  - MLMs have a cluster-specific interpretation
- Because of the random effects, people with identical data values but who are in different clusters will have different predicted values.
- For a one-unit change in X, Y is predicted to change by β units, holding all other variables **and the random effects** constant

Back to This Picture

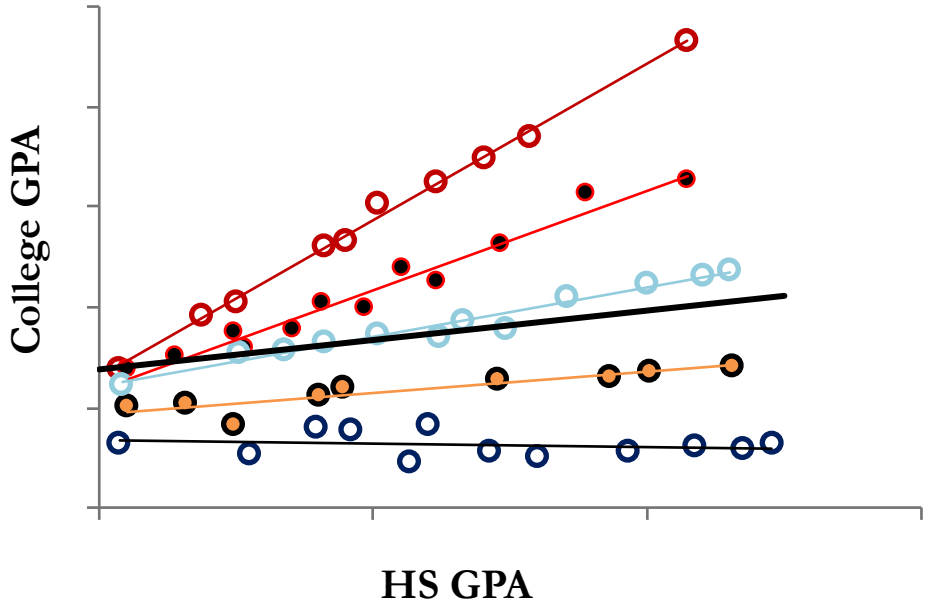


One unit change in predictor

Predicted change in outcome

DBM coefficients are interpreted based on the whole sample

MLM coefficients are interpreted based on the cluster



Difference in Coefficient Interpretation III

- The difference between the cluster-specific and population-averaged interpretations is minimal with continuous outcomes*

- With discrete outcomes(variables with a finite number of categories), the differences between population-averaged and cluster-specific coefficients can be very noticeable **

*Fitzmaurice, Laird, & Ware (2011)
** Carlin, Wolfe, Brown, & Gelman (2001) *Biostatistics*

MLDS Center
Maryland Longitudinal Data System
Better Data • Informed Choices • Improved Results

Application to a Reading Intervention Study

Overview of Research Question

- Data come from a study that was part of an IES grant funded to Dr. Rebecca Silverman
- The overall aim of the study to evaluate the efficacy of a reading intervention to assess whether the growth in reading vocabulary and comprehension was greater for students receiving the treatment compared to students in a control group.
- Will only discuss one of the models for one of the grades

Sample & Method

- 203 kindergarten students
  - 53% of students were ELL
- Students were sampled from 12 different urban, Type-I classified classrooms
  - 6 classrooms assigned to the treatment group
  - 6 classrooms assigned to the control group
- Students' vocabulary was assessed with a pre and post-test
  - Data that will be shown is for the Peabody Picture Vocabulary Test (PPVT)

PPVT Analysis

- Interest is whether PPVT scores are higher for students exposed to the treatment, controlling for pre-test scores and ELL status
  - ICC is 0.21, DEFT is 2.21 so clustering within classrooms is meaningful
- Will demonstrate the difference in running this analysis with a DBM and a MLM

DBM Assumptions

1. All relevant predictors are included
2. The sample size is sufficiently large for asymptotic inferences at the cluster level
3. Observations between clusters are not related
4. The working correlation matrix is "reasonably close" to the population structure

- Assumptions are pretty simple to check
  - (1) the model was theoretically determined
  - (2) sample size was rather small was a correction was used
  - (3) there is no meaningful higher level of the hierarchy
  - (4) the ICC is less than 0.30, so the choice is straightforward

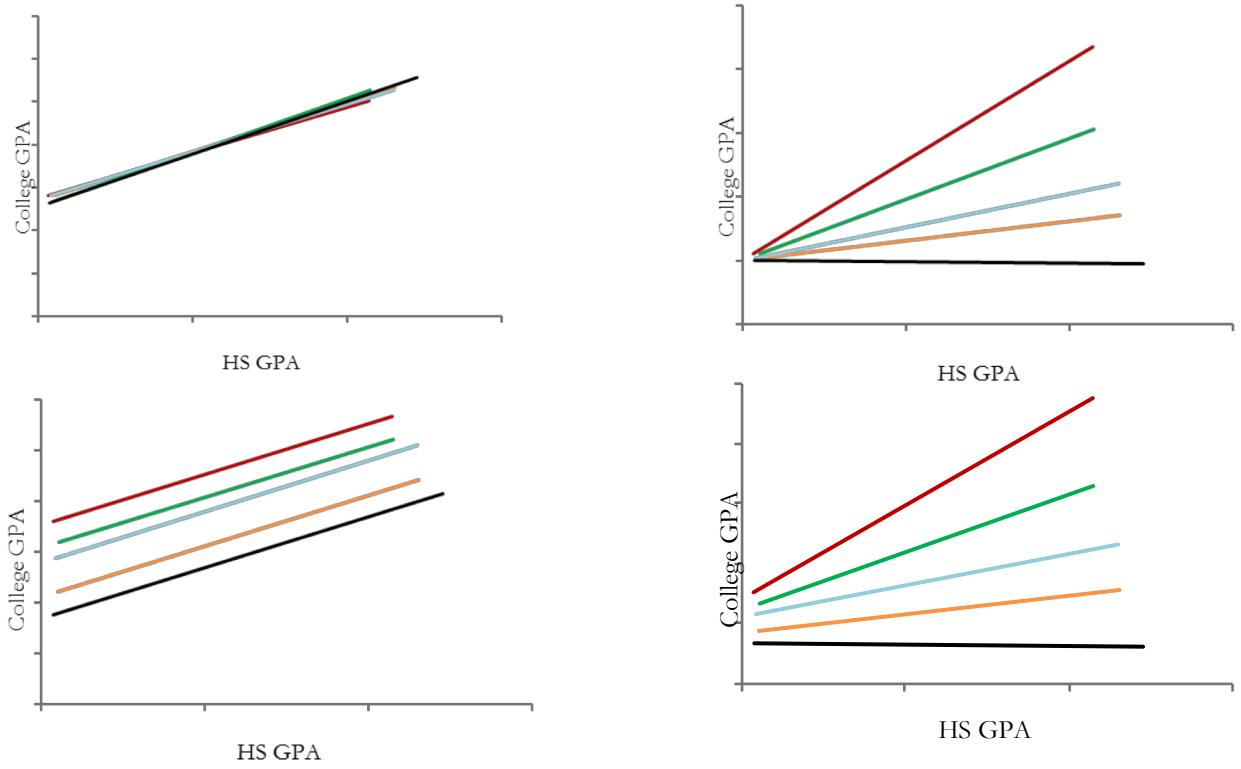
Design-Based Method Results

	PPVT Post-Test	
Effect	Estimate	*p*-value
Intercept	127.43	
Treatment	4.88	.04
ELL	0.04	.99
Treatment × ELL	-2.08	.38
PPVT GSV Pre-Test (mean-centered)	0.77	<.01
Residual Variance	71.89	
Model R^2	0.81	

- Treatment is significant at 0.05 level
- Treatment does not appear to be different for ELL or non-ELL students
- Scores for ELL students are comparable to non-ELL students
- No concerns about assumptions

Multilevel Model

- After predictor variables are selected, the first step is to determine which random effects to include

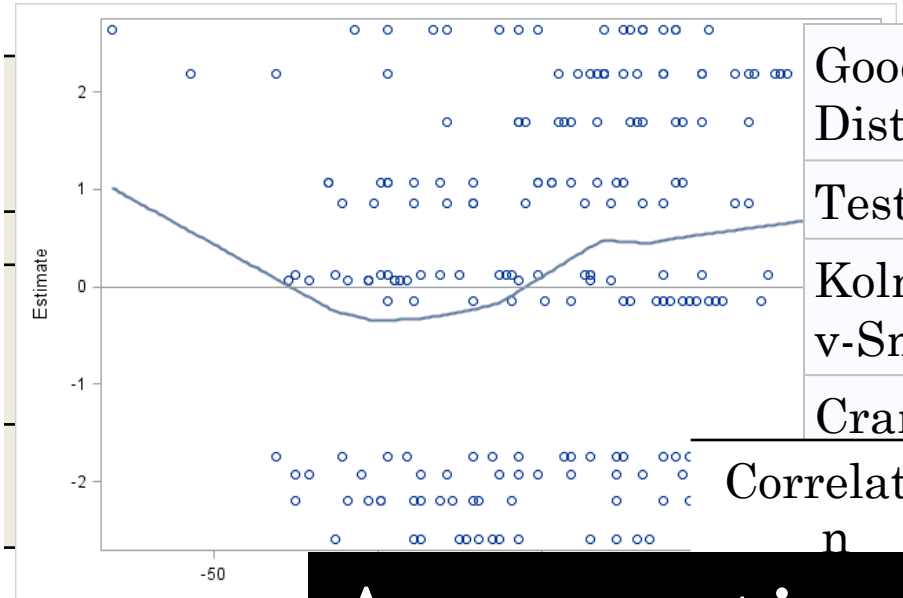


Assumptions

MLM Assumptions

1. All relevant predictors are included
2. All relevant random effects are included
3. The covariance structure of the within-cluster residuals is correct
4. The covariance structure of the random effects is correct
5. The within-cluster residuals follow a (multivariate) normal distribution
6. The random effects follow a (multivariate) normal distribution
7. The random effects are not correlated with any predictors in the model
8. The sample size is sufficiently large for asymptotic inferences at all levels

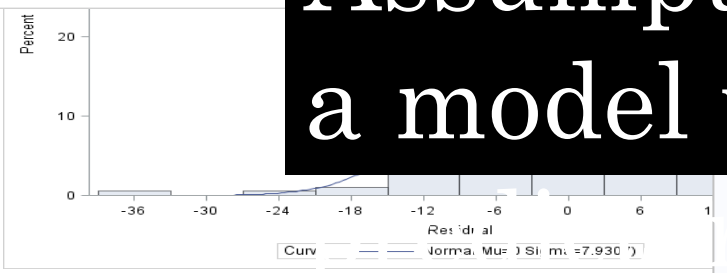
- Won't discuss all of these, but keep in mind that there are a lot of assumptions to check



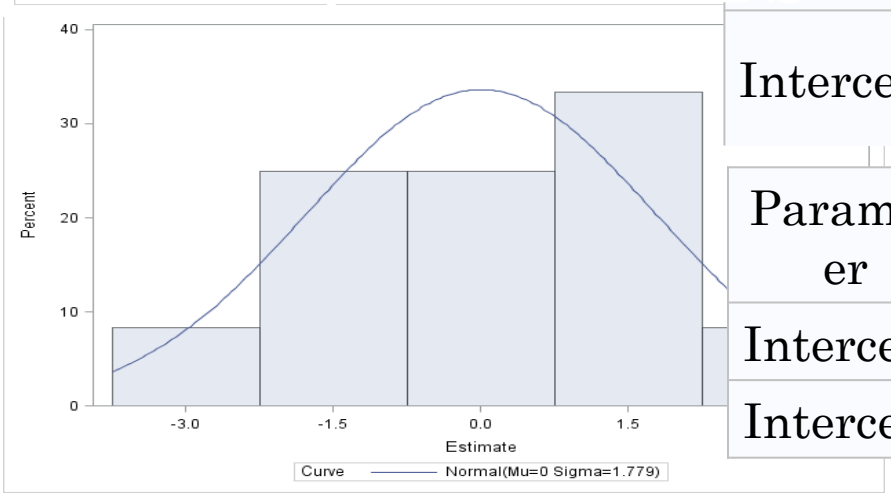
Goodness-of-Fit Tests for Normal Distribution				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.06		>0.15
Cramer-	W-	0.11		0.08

Correlation	P-value		
		0.68	0.08

Assumptions 2,5,6,7 for a model with only 4



			Standard Error
Intercept	Treatment = 0	2.70	4.68
Intercept	Treatment = 1	9.34	8.74

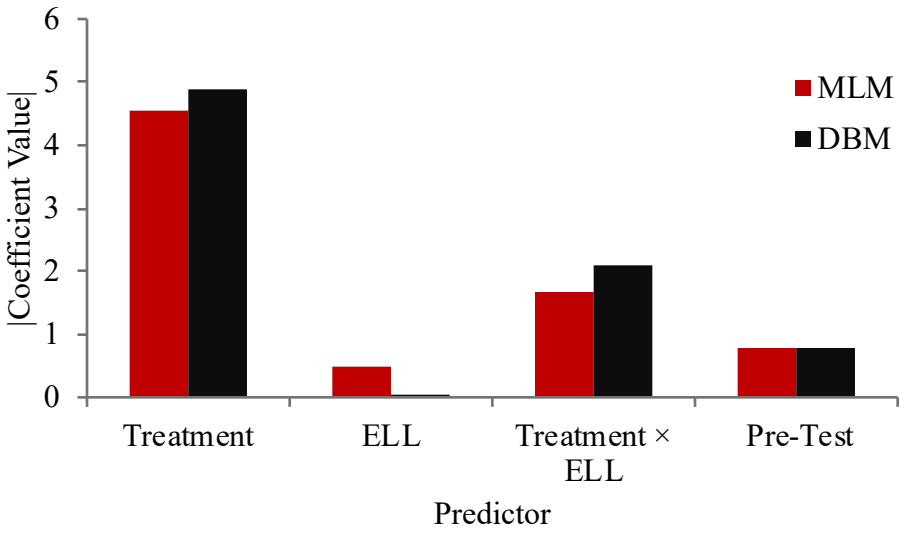


Parameter	Group	Estimate	Standard Error
Intercept	ELL= 0	5.78	7.24
Intercept	ELL= 1	4.34	5.84

MLM		
Effect	Estimate	*p*-value
Intercept	127.16	
Treatment	4.53	.07
ELL	0.49	.83
Treatment × ELL	-1.65	.54
PPVT GSV Pre-Test (mean-centered)	0.77	<.01
Intercept Variance	6.11	
Within-Cluster Variance	66.32	
Total Residual Variance	72.43	

DBM		
Effect	Estimate	*p*-value
Intercept	127.43	
Treatment	4.88	.04
ELL	0.04	.99
Treatment × ELL	-2.08	.38
PPVT GSV Pre-Test (mean-centered)	0.77	<.01
Residual Variance	71.89	

Effect	MLM	DBM
Intercept	127.16	127.43
Treatment	4.53	4.88
ELL	0.49	0.04
Treatment × ELL	-1.65	-2.08
PPVT GSV Pre-Test (mean-centered)	0.77	0.77
Intercept Variance	6.11	---
Within-Cluster Variance	66.32	---
Total Residual Variance	72.43	71.89



MLDS Center
Maryland Longitudinal Data System
Better Data • Informed Choices • Improved Results

Implications

MLM/DBM Comparison

- Main point is that the DBM analysis was much simpler, involved far fewer assumptions, and, in this case, allowed calculating an OLS-type R^2
  - With these data, the cluster-specific and population-averaged interpretation are interchangable

- The *p*-value changing was a coincidence in these particular data
  - Not a typical result
  - Strong assumptions of MLMs could have contributed to this though*

*LeBeau (2013) Doctoral Dissertation, UMN

MLM/DBM Comparison II

- DBMs are also much less sensitive to researcher specifications
  - Fewer decisions need to be made explicitly by the researchers
  - These decisions are often based on sub-optimal statistical theory
  - Results for the same data with the same predictors can vary based on who is conducting the analysis
- Quantifying the random effects sounds great in theory
  - In realistic models with many predictors, it can become a very complex task that can have discernible effects on the results

MLDS Center's Goals

- Internal discussion has been devoted to making MLDS' findings broadly interpretable to the public

- DBMs make fewer assumptions, are interpreted the same as far simpler statistical models, and give simpler interpretations
  - The model building and assumption made by MLMs alone can be difficult to understand
  - MLM coefficient interpretation is also far less intuitive

Why This Matters

- As a general guideline, researchers strive for the simplest statistical method that will accommodate their data and handle their research question
  - Much criticism is directed toward using models that are too simple and don't meet assumptions/ data requirements
- With clustered data in education, the exact opposite occurs in practice
  - MLMs are almost universally used to account for clustering
  - Cluster-specific information is not always of interest
  - Researchers are making strong assumptions and tackling difficult problems to get random effects they don't need or even care about

This Talk in One Sentence

Use methods that address the needs of your research questions.

Concluding Recommendations

- MLMs are useful if:
  - Directly modeling the covariance structures/ explaining the clustering mechanism is relevant to the research question
  - Inferences about specific clusters are desired
  - The partitioning of variance is substantively important

- DBMs are useful if:
  - The clustering is just an aspect of the data to accommodate, especially with discrete outcomes
  - The correlation within-clusters is not a modeling interest
  - Interest does not pertain to specific clusters

Thank You!

dmcneish@umd.edu