Exploring Reading Teacher Effectiveness in Elementary Schools Using Hierarchical Linear Models

Joseph Prather & Marco Muñoz
Jefferson County Public Schools

National Evaluation Institute 2010
College of William and Mary

Research Significance

- Teacher quality is an important area of investigation that has emerged in recent years among educational researchers—NCLB and Race to the Top
- "In this age of accountability, we must define good teaching by results, not by personal characteristics or our preconceived notions. When the goal is student learning, *seeming* to be a good teacher and actually *being* a good teacher can be very different."
  (Jerald, 2003, p. 13)
- Effective Teaching = Student Achievement
  (Stronge, 2010)

Theoretical Framework

- The value-added models (VAM) is based on the growth a student makes from the time of entering a school or a classroom to the time of their exiting that school or classroom
  - Tennessee Value-Added Assessment System (TVAAS) with Bill Sanders and colleagues
  - Dallas Independent School System (DISS) with Webster, Meandro, and colleagues
  - Kentucky Accountability System as part of KERA
  - James Stronge and Jason Millman

Why HLM as a methodology?

- The inherently nested nature of educational data
  - Individual students
  - Within classes
  - Within schools
  - Within district
  - Within state
- This study was individuals within classes and individuals grouped in classes influence each other
  - No more independence of observations assumption
- HLM can provide accurate classroom level effectiveness estimates while minimizing error

General Overview of Design

- Exploratory, correlational in nature
- Progressed through a series of models, testing individual variables and then sets of significant variables (Raudenbush & Bryk, 2002)
- Tested the individual variables for three consecutive years to determine which variables reoccurred
- Compared models across the three years (n=100; range 2335-241 per year)
- Classified teachers as “highly ineffective”, “ineffective”, “neutral”, “effective,” and “highly effective”

Dependent Variable

Kentucky Core Content Test (KCCT)

4th Grade Reading (using scale scores)

Statewide reading testing assessment and part of the accountability system

Independent Variables-Level I

- Student Total Absences
- Student Days Membership
- Education Index (ordinal data based on Census tract median education)
- Median Income
- CTBS Reading Scale Score
- Student Percent Attendance
- Student Days Old
- Student Black
- Student White
- Student Other
- Student Female
- Student Free/Reduced Lunch
- Student English as a Second Language (ESL)
- Student Special Needs
- Student Gifted

Independent Variables-Level 2

- Classroom aggregated data created from selected Level I variables
- Teacher Years of Experience in School District
- Class Size
- Teacher Female
- Teacher Black
- Teacher White
- Teacher Other
- Teacher Rank
- Teacher Degree

One-Way ANOVA w/ Random Effects (Null Model)

- Level-1 equation (student level):

$$Y_{ij} = \beta 0_j + r_{ij}$$

- Level-2 equation (classroom level):

$$\beta_{oj} = \gamma_{oo} + u_{oj}$$

- Expanded Model:

$$Y_{ij} = \gamma_{oo} + u_{oj} + r_{ij}$$

- Intraclass Correlation:

2001 21%

2002 21%

2003 17%

Note: About 10% or above is appropriate for HLM analyses

Regression with Means-as-Outcomes (Null on Level I)

- Level-1 equation

  $Y_{ij} = \beta_{oj} + r_{ij}$

- Level-2 equation

  $\beta_{oj} = \gamma_{oo} + \gamma_{o1}(W_j) + u_{oj}$

- Expanded Model:

  $Y_{ij} = \gamma_{oo} + \gamma_{o1}(W_j) + u_{oj} + r_{ij}$

Variables Significant Three Consecutive Years:

- Teacher Years Experience
- Education Index
- CTBS Reading Test

Conditional Intraclass Correlation:

2001	14%
2002	16%
2003	11%

ANCOVA with Random Effects (Null on Level 2-Random Intercept)

- Level 1:
- $Y_{ij} = \beta_{oj} + \beta_{1j}(X_{ij} - \overline{X}..) + r_{ij}$
- Level 2: $\beta_{oj} = \gamma_{oo} + \mu_{oj}$

  $\beta_{1j} = \gamma_{1o}$

- Expanded:

  $Y_{ij} = \gamma_{oo} + \gamma_{1o}\,(X_{ij} - \overline{X}..) + \mu_{oj} + r_{ij}$

Variables Significant Three <u>Consecutive Years:</u>

- Intercept
- Days Absent
- CTBS Reading Score
- Black
- Female
- Free/Reduced Lunch
- Advance Program

Note: All t-values p<.001 except 2002 Days Absent (p<.01)

ANCOVA with Random Effects

Model A

$Y = \beta_0 + \beta_1*(CTBS) + r_{ij}$

Level 2 for both models:

$\beta_{oj} = \gamma_{oo} + \mu_{oj}$

All coefficients are fixed

Model B

$Y = \beta_0 + \beta_1*(\text{Student Total Absents}) + \beta_2*(CTBS) + \beta_3*(\text{Student Black}) + \beta_4*(\text{Student Female}) + \beta_5*(\text{Student Free/Reduced Lunch}) + \beta_6*(\text{Student Advance Program}) + r_{ij}$

Intercepts- and Slopes-as-Outcomes Models (Model C)

Level One:

$$Y = \beta_0 + \beta_1*(\text{Student Total Absents}) + \beta_2*(\text{CTBS}) + \beta_3*(\text{Student Black}) + \beta_4*(\text{Student Female}) + \beta_5*(\text{Student Free/Reduced Lunch}) + \beta_6*(\text{Student Advance Program}) + r_{ij}$$

Level Two:

$$\beta_0 = \gamma_{00} + \gamma_{01}*(\text{Teacher Years Experience}) + \mu_0$$

$$\beta_1 = \gamma_{10} \qquad \beta_2 = \gamma_{20}$$

$$\beta_3 = \gamma_{30} \qquad \beta_4 = \gamma_{40}$$

$$\beta_5 = \gamma_{50} \qquad \beta_6 = \gamma_{60}$$

3 Model Comparison

	Model A	Model B	Model C
Level 1	CTBS	CTBS	CTBS
		Student Absences	Student Absences
		Student Black	Student Black
		Student Female	Student Female
		Student Free/Red	Student Free/Red
		Student Adv. Prog.	Student Adv. Prog.
Level 2	Random Intercept	Random Intercept	Teacher Yrs. Exp.
	All others fixed	All others fixed	All others fixed

Examining Teacher Effects

- Using the residual as a teacher estimate of effectiveness (which ideally would be random)

Correlations of Residuals by Year

	Model A	Model B	Model C
2001/2002	.63	.56	.56
2002/2003	.54	.44	.42
2001/2003	.59	.52	.50

Teacher Ratings

Classified Teachers as:

2	2 or more SD above the mean
1	between 1 and 2 SD above the mean
0	between 1 SD above and below the mean
-1	between 1 and 2 SD below the mean
-2	2 or more SD below the mean

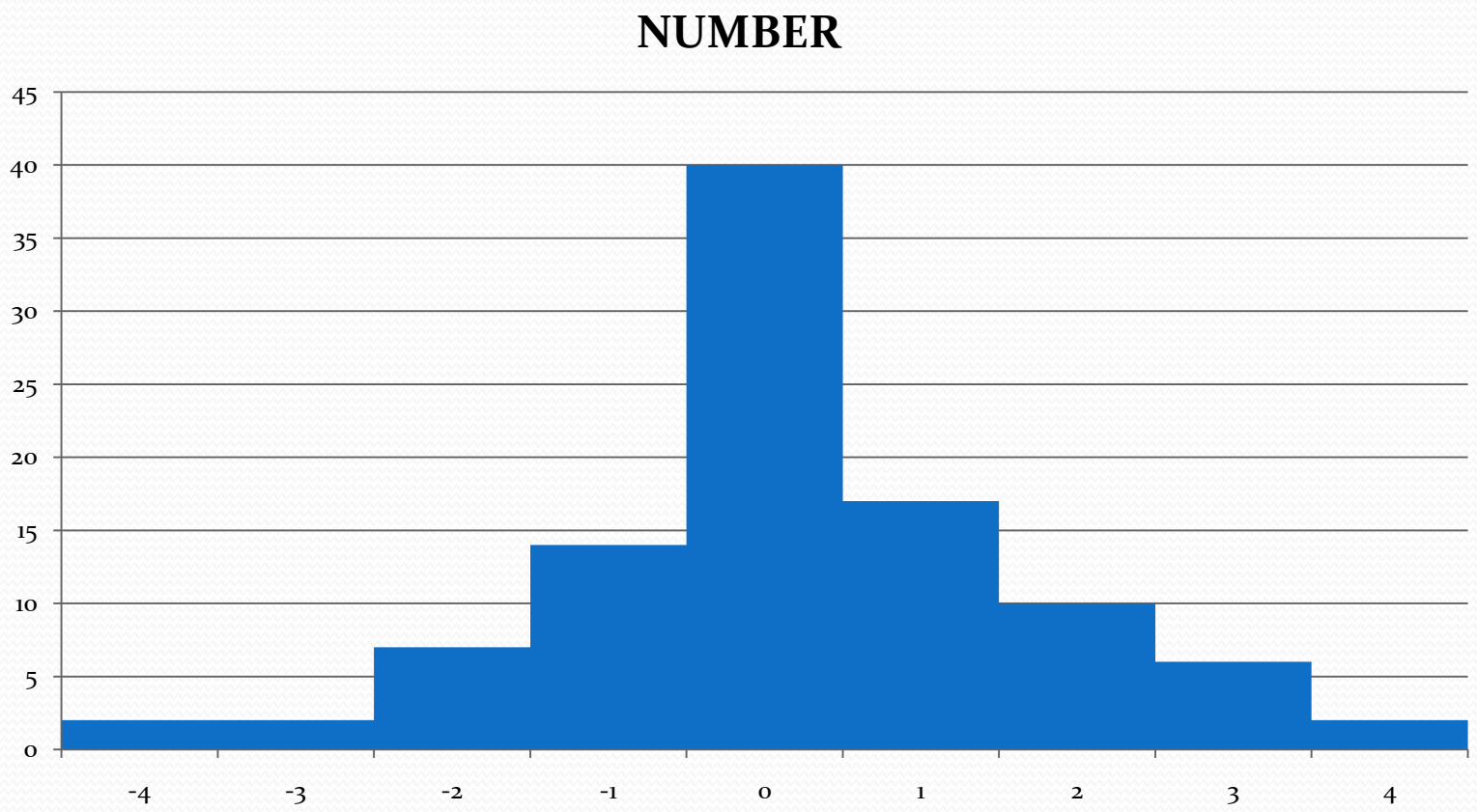
Teacher Ratings *(Continued)*

- 100 Teachers in all three years of study
- Summed the 3 years ratings

Sum and Frequency

Σ	Freq.	Σ	Freq.	Σ	Freq.
4	2	1	17	-2	7
3	6	0	40	-3	2
2	10	-1	14	-4	2

Teacher Ratings *(Continued)*



Teacher Ratings *(Continued)*

- Only 1 teacher crossed over the zero range (1,-1,0)
- 8 teachers were classified in ranges that had a difference of 2
- 39 teachers were rated in the middle range for all three years (0,0,0)

Discussion & Implications

- Definitely not use to rank order all teachers
- Even though it there is a moderate correlation from year-to-year (approximately .5), there is still too much variability to use exclusively
- Would need to be validated with other measures
- Numerous teachers had to be omitted (class size, did not teach 4th grade three years consecutively)
- Very limited in outcome – effectiveness means increasing reading ability as measured by state test

Discussion & Implications (Continued)

- Classifies teachers, but not instructional practices—black box model
- High mobility in our district (followed 100 day accountability rule)
- Does this model "level the playing field"—questions emerging for low-income schools?
- Used for identification of support more than reward/punitive actions—formative rather than summative uses

Conclusion

- Stronge and Tucker (2000) provide several guidelines for using testing data models as part of teacher evaluations, including: "use student learning as only one component of a teacher evaluation system that is based on multiple data sources" AND "when judging teacher effectiveness, consider the context in which the teaching and learning occur" (pp. 53-54).
- Millman (1997) also provide valid guidelines: fairness with teachers, comprehensiveness (of assessment), competitiveness (pros and cons), and consequential validity (Campbell's Law).

Conclusion

- Great progress from NCLB to the Race to the Top—from input to outcomes. Also, how about the inclusion of processes to avoid black-box outcome evaluations?
- "The existing research base on VAM [value-added models] suggests that more work is needed before the techniques can be used to support important decisions about teachers or schools" (McCaffrey et al., 2003, p. 111)
- If these models are used when evaluating teachers, they should be used along with other measures of teacher performance.

THANK YOU!

Joseph Prather
Joseph.prather@jefferson.kyschools.us

Marco Muñoz
Marco.munoz@jefferson.kyschools.us