Measurement Issues Associated with Value-Added Methods

Mark D. Reckase
Michigan State University

A Simplified Example

- Value-added models are typically used for very complex situations
  - Data from complex testing programs
  - Nesting of students within classrooms within schools
  - Complex relationships between measures of student performance and other variables
- Avoid complexities by using a simple example

Girls Gain in Height
Age 9 to 10

- Measure height on 9th and 10th birthday – gain in height is analogous to gain in achievement from one grade to the next.
- Under ideal circumstances, measure height accurately in inches on each birthday.
- Gain is simply the difference between the two heights.

Results of Example

	Age in Years		Gain in Inches
	9	10	
Mean Height	52.07	55.10	3.03
SD	1.78	2.33	1.56

- Large effect size for gain in terms of age 10 standard deviation – 1.3
- H10=H9+Gain+Error – Account for 56% of variance
- Is the distribution of gain normal? A normal distribution implies 2.5% negative gain.

Issue #1 Shapes of Distributions

- Distribution of gain in height (or achievement) is probably not normal.
  - This implies that the distribution of height at age 10 is not normal either.
  - Are there any assumptions of normality in the value-added models?
- What proportion of students would have negative gains in “true” achievement after one year of instruction? This proportion and the mean gain suggests the form of the distribution of gains.

A Little More Complexity

- Suppose height is measured using rulers with different units for the two ages.
  - Have the data, but not the conversion for the units.
  - Measure without error.
- Can the gain in height be determined?
- This mirrors measuring achievement in one grade with one test and achievement in the next grade with another test, if the tests measure the same thing. The results have different units of measurement.

Results – More Complex Example

	Age in Years	
	9	10
Mean Height	52.07	139.96
SD	1.78	5.93

- Regression of age 10 height on age 9 height
  - Est(H10)=2.48(H9)+10.69
  - Account for 56% of variance – same as first example
- Age 10 heights are measured in centimeters. The gain in height in centimeters is 7.7.

Issue #2 Changing Measurement Units

- When the units of measurement are not the same for the two measurements, it is not possible to estimate gain.
  - The intercept term in the regression equation is not the mean gain.
  - What is the effect size for gain?
- But, the ordering of the residuals from the regression is the same as the ordering of gains when the units are the same.

Even More Complexity

- Instead of accurate measurements, height is estimated with a standard error of .8 inches.
- Consider the two situations described above
  - Measurements with the same units of measurement
  - Measurements with different units of measurement with unknown conversion

Results – Same Units with Error

	Age in Years		Gain
	9	10	
Mean Height	52.04	55.14	3.10
SD	1.93	2.47	1.92

- Effect size for gain 1.26 compared to 1.3 for no error case.
- Regression model account for 41% of variance compared to 56% for no error case.
- Gain correlated .8 with true gain.

Issue #3 Measurement Error

- Measurement error reduces the effect size – in this case the reduction is small.
- Measurement error reduces the amount of variance accounted for by the model.
- Measurement error reduces the correlation of gain as a dependent measure with other measures.

Results – Different Units with Error

	Age in Years	
	9	10
Mean Height	52.04	140.05
SD	1.93	6.28

- Regression of age 10 height on age 9 height
  - Est(H10)=2.09(H9)+31.05
  - The coefficients are quite different than those from the values without error, 2.48 and 10.69 respectively.
  - Account for 41% or the variance, compared to 56% without error.

Issue #4 Unit Change and Error

- When units are not equal, the regression line has lower slope and larger intercept.
- Residual is correlated about .81 with true gain.

Much More Complexity

- Value-added with achievement measures is somewhat different than the height example because the thing measured at each grade level is different.
- To be comparable, height would be measured at age 9 and something correlated, like weight, at age 10.
- The shift in thing measured would reduce the relationship between scores at each grade.

Issue #5 Shift in construct

- A shift in construct from grade to grade reduces the relationship between scores.
- The result is underestimation of improvement from grade to grade.

Value-added Models

- Value-added models are elaborate linear models that include achievement measures as both dependent measures and covariates and other measures as covariates.
- These models have a number of important assumptions. The assumptions are related to the previous examples.

Assumptions

- Conditional residual distributions are normally distributed. This assumption may not be met in the when growth data are analyzed.
- Assuming a distribution shape requires that the scale of measurement has interval properties. There are two justifications for interval scales
  - IRT models
  - Produce tests that give desired form of distribution.

Assumptions

- The models also assume the conditional residual variances are equal.
- They also assume a linear relationship between variables.
- Most articles on value-added modeling do not consider whether assumptions are met.

The Ideal Testing Program for Value-added Modeling

- The test either needs to yield scores that are distributed according to a theoretical form or be well fit by an IRT model to approximate an interval scale.
- The test must be highly reliable to minimize the regression effects due to error of measurement.
- Tests at different grade levels must be carefully designed to measure the same constructs.
- The tests must be carefully vertically scaled so units for different tests are on the same scale.
- Students should be motivated and the test properly administered.

Research Topics

- Effects of violations of assumptions
- Fit of the models should be checked
- Need to know percent of variance accounted for and the magnitude of gains.

A Final Thought

- The ordering of gains are much better estimated that the actual magnitude of gains.
- Relationships with gains are pretty well estimated. This supports research on variables that influence educational improvement.
- Absolute amount of gains are less well estimated. This has implications for use of value-added models for teacher and school evaluation.