

Discussion of  
**Vertical Scaling in Value-Added Models for  
Student Learning**  
(Briggs, Weeks and Wiley)

Henry Braun  
Boston College

National Conference on Value-Added Modeling

Madison WI  
April 22-24 2008

# Growth Models under NCLB

- AYP (2002) based principally on status models
- Accountability relying (partially) on change in status (e.g. growth to a standard) seen as fairer
- Conventional growth models require a single cross-grade scale
- Vertical scale construction involves applying a linking procedure to grade-specific results
- There is no optimal linking procedure – and many different procedures are in use

How do scale characteristics and value-added results depend on the choice of linking methodology?

# The Experiment

Explored a 2x2x2 design

- IRT (Rasch vs. Lord)
- Calibration (Separate vs. Hybrid)
- Estimation (EAP vs. MLE)

Investigated two VAMs

- Three-level HLM
- Layered model (EVAAS)

# Results

- HLM
  - Statistical characteristics depend on linking procedure
  - Variance components for slopes (students & schools) are relatively small
  - Medium to high correlations for school value-added estimates
  - **Questions**
    - Why not use a non-linear model for fixed effects?
    - How consistent are the results at the extremes?
- Layered Model
  - Medium to high correlations for school value-added estimates
  - Three-level categorizations do not vary much across procedures
  - **Questions**
    - How does the “thinness” of the data affect results?
    - How consistent are the results at the extremes?

# Vertical Linking

Vertical linking is problematic on both substantive and psychometric grounds

- What information is lost when a unidimensional vertical linking is carried out (Schmidt et al., 2005)?
- How does assumption of unidimensionality distort results (Martineau, 2006)?
- How do errors in linking affect value-added estimates (Doran & Cohen, 2005)?
- Degrees of freedom in linking methodology (Briggs et al., 2008)?

# Growth Trajectories

- Characteristics of paths depend on linking procedure
- Non-linear trajectories complicate interpretations of growth since expected growth depends on starting point
- School evaluations based on meeting fixed growth objectives can be unfair – even for same-grade comparisons
- If proficiency standards were set in a principled and coherent manner, then (improved) growth-to-a-standard indicators might (actually!) be preferred

# Value-added Models

- Important to distinguish general growth models from VAMs
- VAMs yield normative results that are not well-suited for accountability
- VAM results are better used for preliminary identification
- VAM results can/should contribute to school evaluations – but only in conjunction with other types of evidence
- Other sources of evidence should also be critically examined

**Can VAM-related sensitivity studies help us choose among linking procedures?**

# Conclusions

- Vertical linking and growth modeling are here to stay (at least for a while) – But we need to convey our concerns about how they are used and misused
- Sensitivity analyses make an important contribution to our understanding of the operating characteristics of VAMs
- We should continue the search for VAM alternatives that don't require a vertical scale
  - Variable persistence models
  - Conditional growth models (Betebenner)
  - Discrete models (Fielding et al.; Betebenner; Braun & Qu)

# Parting Thoughts

NCLB accountability system is an uneasy marriage of federal mandates and state-specific responses.

Flexibility given to states with respect to test construction, standard setting, AYP targets and (now) growth models means that there is no comparability in results across states.

Although comparability is not a priority, its absence means that it is more difficult to meaningfully interpret states' results in terms of the (relative) educational progress of their students.

So even if all children achieve proficiency (!!), we can not say wether they truly will not be “left behind”.