THE IMPORTANCE OF EDUCATIONAL MEASUREMENT FOR VALUE-ADDED MODELING

Mark D. Reckase
Michigan State University

Things I Have Heard over the Years

- What the tests measure does not matter as long as the tests are correlated over years.
- Vertical scaling of tests is not necessary because the variables are standardized. Besides, vertical scaling does not work.
- The characteristics of the test score scale does not matter so use what is conveniently available – percentile ranks, number correct scores, IRT estimates, etc.

A Different Perspective on Value-added Estimates

- Value-added estimates are a transformation of the scores on the test that is the dependent variable for the model.
- Many discussions of value-added results include the term “growth”, but the typical value-added analysis does not give information about growth.
- Value-added estimates are measures of teacher performance and they need to be evaluated according to accepted measurement practices.

VAM Estimate as a Measurement

- Consider teacher competence as a latent variable that is to be estimated – this is my item response theory background coming out.
- The VAM estimate for the teacher is an estimate of the location of the teacher on the latent variable.
- Measurement professional standards indicate that when measurements are reported, they should include the following:
  - Measures of reliability and standard error.
  - Evidence for the validity of inferences drawn from the reported "scores".
  - See recent report by Haertel (2013) from a talk at ETS.

Three Considerations for the Future

- If you were designing achievement measures to support VAM estimation, what would be optimal test designs?
  - Current approaches using existing tests that were not designed for this purpose.
  - Some suggest using the same test at the beginning and end of the academic year so there is a meaningful measure of growth.
- Can we set standards on the latent variable of teacher competence to indicate how good is good enough?
  - Now we do not connect student gains to VAM estimates.
  - What level of the VAM estimate corresponds to zero student growth? It is generally not possible to tell because the estimates are standardized.

Three Considerations for the Future

- There is an area in psychometrics that is called “studies of measurement invariance.” Do the results of the tests mean the same thing for students going through different instructional programs.
  - We assume measurement invariance when getting VAM estimates.
  - How much does violation of this assumption affect VAM estimates?

Reactions to the Initial Points

- The characteristics of the achievement measures do matter. Different tests give different results.
  - What is the optimal correlation between the test at year *k* and year *k* -1?
  - Is there an assumption of measurement invariance that we typically ignore?
- Vertically scaled results are useful at the very least for the purpose of interpreting the VAM estimates. When is student growth 0?
- The scale for reporting test results matters. There are numerous studies, including our own, that show differences for different reporting score scales.
  - Is there a scale transformation that optimizes VAM estimates?