Can Multifactor Models of Teaching Improve Teacher Effectiveness Measures?

Valeriy Lazarev and Denis Newman
Empirical Education Inc.
Outeline

1. Motivation: States adopting “multi-measure teacher evaluation systems” (MMTES)
2. Issues: What are teacher evaluations measuring?
3. Approach: factor analysis of MET data
4. Results: three factor model of teacher effectiveness
5. Implications for decision-making
Motivation

• Teacher evaluation is currently a major policy issue, driven in large part by DoE requirements (NCLB waivers, RTTT grants).

• States adopt multi-measure teacher evaluation systems (MMTES).

• We need to understand what we are measuring and how to combine various measurements.
Related Work

• Measures of Effective Teaching (MET) project: Massive data collection and analysis of correlations between various aggregate measures of teacher performance (value-added, observation and survey scores). Correlations are significant positive but small (Kane and Staiger, 2012).

• Optimal composites – weighting summative measures of student achievement, classroom observations, and student surveys - assuming a one-dimensional underlying “teacher effectiveness” (Hansen et al. , 2013; Mihaly et al. , 2013)
Questions

• What are we measuring with observations and surveys?
  o What if “teacher effectiveness” is multidimensional?
  o Can we identify a (small) set of underlying factors that are being measured?

• What impact might additional factors, unrelated to test performance, have on the overall effectiveness of our education system?

• How can the factors underlying the evaluation results be used in personnel decisions?
Combining Measures vs. Singling Out Factors

- Each instrument measures the same concept
- Weights (should) reflect relative reliability of each instrument
Combining Measures vs. Singling Out Factors

- Each instrument measures one or more underlying concepts/factors
- Summative score is a combination of factor scores
Approach

• Factor analysis – identify several latent independent (orthogonal) factors

• Assumptions:
  o Several underlying factors/dimensions, $f$, of effective teaching
  o Each item measures a combination of factors
  o Only one factor is associated with short-term student achievement gains (teacher value added)

• Method:
  o Factor analysis
  o Target rotation: $\lambda_{VAM} \sim (1,0,0\ldots)$
Data

- MET data – a “model” of state MMTES
- Middle-school math and ELA teachers assessed on multiple metrics by MET project
- Value-added scores: based on study-administered BAM (math) and SAT9 (ELA) tests
- Observation rubrics:
  - FFT (eight components of two domains: “Classroom environment” and “Instruction”)
  - CLASS (all 12 components).
- Tripod survey (“7C”) – 36 items
- Total 57 measures/variables
Results

Three factor model:

• Factor 1 ("Effective"): TVA, classroom procedures and behavior management (observation), control (survey)

• Factor 2 ("Constructive"): pedagogical devices (instructional dialog, feedback, and discussion)

• Factor 3 ("Positive"): teacher’s connection to students, students’ positive feelings and perception of the teacher’s empathy (most survey items)
What Impact Might These Factors Have on the Overall Effectiveness of Education System?

- All three factors may predict student outcomes
- “Effective” – spring test scores
- “Constructive” – study skills, future educational choices?
- “Positive” – longer-term impacts on positive behavior, attendance, and staying in school?
- A fourth factor? – peer interactions and leadership (based on performance outside of classroom – data not available)
Implications for Personnel Decisions

- Decision-makers should weight the factors depending on the relative value in a particular context.

- For raises and bonuses, a single score with a high weight of test-related factor can be generally useful.

- For terminations, a matrix can be used (minimum passing score on each factor).

- A specialist position calls for outstanding scores on “constructive”

- For promotion to a leadership position, the hypothetical fourth factor may be given a higher weight, etc.
Future Work

• Closer link between theories of teaching and learning and empirical studies of teacher performance data

• Longitudinal data analysis to establish associations between the additional factors and distal outcomes – results will provide a basis for weighting

• Collection and analysis of broader evaluation data sets, in particular, including results of out-of-classroom observation, peer and parent surveys
Thank you

Val Lazarev
Denis Newman

Empirical Education Inc.

www.empiricaleducation.com
@empiricaled