What Do Value Added Measures of Teacher Preparation Programs Tell Us?

Dan Goldhaber
Center for Education Data & Research
University of Washington Bothell
Context

- Teacher training programs are increasingly being held under the microscope
  - “by almost any standard, many if not most of the nation's 1,450 schools, colleges and departments of education are doing a mediocre job of preparing teachers for the realities of the 21st century classroom” (Arne Duncan, 2009)
  - New NCTQ evaluation suggests few programs are up to snuff

- New emphasis on focus on student outcomes
  - Some RTTT plans require states to report on the the student growth/value added of program graduates
  - New CAEP standards: “surmounting all others, insist that preparation be judged by outcomes and impact on P-12 student learning and development”
Questions

1. What do we currently know about value-added teacher effect estimates generated by different models?

2. What more needs to be known about this issue (question 1)?

3. What can’t be directly resolved by empirical evidence on this issue?

4. What are the practical implications of the research on this issue for decision making?
Limitations of Using Student Growth as a Measure of TPP Performance

- All methods of estimating teacher performance based on student test achievement entail *predicting* achievement.
  - Some limitations/challenges are similar to individual teacher case.
    - Measures can be sensitive to the student test that is used.
    - Difficult to know whether models have fully distinguish a teacher’s contributions to student learning from other school, home, etc. factors.
  - Value-added methods *may* be able to tell us something about the effectiveness of a program’s graduates, but this information is a function both of graduates’ experiences in a program (training) and of who they were when they entered (selection).

- Limitations of value added are likely true of all other means of judging TPPs based on outputs.
What’s Known About VAM TPP Estimates?

• Empirical research reaches somewhat divergent conclusions about the extent to which training programs explain meaningful variation in teacher effectiveness
  - TPPs producing effective math teachers also tend to produce effective reading teachers
  - Little evidence of program specialization

• Relatively little quantitative research on the features of TPPs that are associated with student achievement
  - What does exist offers suggestive evidence that some features may matter
    - e.g. licensure tests, student teaching environment and connection between student teaching and methods coursework
Statistical Issues Associated w/ Estimating TPP Effects

• TPP school system feeder patterns and distinguishing TPP effects from district or school effects
  - Evidence that structure of the teacher labor market differs from one state to the next
  - Models can include district or school effects, but theory cannot tell us whether these ought to be included

• How do we weight graduates from the past when judging programs?
  - Seems wrong to count graduates from long ago toward TPP effect estimates, but only using recent graduates exacerbates small program problem; very imprecise estimates for small programs & unintended incentives
  - TPP effects appear to fade out w/ teacher experience
Decay of TPP Effect Estimates

**MATH**
- Decay ($\lambda=0.054$)
- Selectivity Decay ($\lambda=0.050$)

**READING**
- Decay ($\lambda=0.045$)
- Selectivity Decay ($\lambda=0.061$)

Half-life=$\frac{\ln(2)}{\lambda}$
What More Needs to be Known?

1. More TPP analysis based on secondary teachers
   - Majority of VAM studies are based on elementary teachers

2. What goes on inside TPPs? Necessary to try to disentangle selection and training
   - Candidate selection processes, timing and nature of student teaching, required coursework

3. Reaction of TPPs to new accountability pressures
   - Do they create feedback loops that lead to positive institutional changes?

4. More analysis of non-VAM TPP outcomes
   - Measured student learning is important but not everything
What Can’t Be Resolved Based on Empirical Evidence?

• To what extent value-added should be used at all to evaluate TPPs
• How to try to separate the impact of TPP graduates from school and district environments
• What level of statistical confidence is the right level to take action or report results publicly (95% level probably isn’t right)
• How to weight student achievement against other, potentially competing objectives, such as the diversity of the teacher workforce
What Are the Practical Implications of the Research on Decision-making?

• How, potentially different (selection vs. training), TPP effects are used will depend on who is using them
  - Some (e.g. principals) likely care about total TPP effects, while others (potential teacher candidates) might want to know more about value of training

• Those believing in formal (college or university based) teacher training ought to advocate for more research on the value of different training features
  - Unlocking the secrets to effective training (if they exist) is a key to dramatic teacher workforce improvements
  - My guess is that in the absence of this kind of work we continue to see more reliance on alternative pathways