

Teacher Prep Programs and Student Growth

Daniel F. McCaffrey

TeachingWorks

University of Michigan February 5, 2015

This talk has not been formally reviewed and should not be cited, quoted, reproduced, or retransmitted without ETS's permission.



Background

- ❑ States are using the achievement test scores of students taught teacher prep program graduates to evaluate the programs
- ❑ Initiated by Louisiana over a decade ago
- ❑ Promoted by Race to the Top and NCLB waivers
- ❑ Could be codified in new US ED regulations
 - Requires growth measures for all graduates who are teaching, not just those teaching tested grades, by using other measures of growth such as student learning objectives

Some Examples of Such Analyses Exist

- ❑ Noell et al. estimate program effects using data using graduates' students' achievement from Louisiana (2003-04 to 2010-11 school years)
 - Find differences among programs
 - Report results by program publicly
<http://regents.louisiana.gov/value-added-teacher-preparation-program-assessment-model/>
- ❑ Boyd et al. (2009) study programs that supply teachers to NYC DOE for 2005-06 school year
- ❑ Goldhaber, Liddle, and Theobald (2012) estimate program effects using data from Washington for the 2006-07 through 2009-10 school years
 - Most thorough analysis
 - Controlled for selectivity of programs

Goals of the Talk

- ❑ Describe the method
 - Contrast to teacher value added
- ❑ Discuss potential issues with using value added to evaluate preparation programs

Value Added

- ☐ Use longitudinal student test score data to estimate the impact of various inputs on student learning
- ☐ The highest profile inputs are teachers and school leaders
- ☐ Inputs can be aggregated to study the effects of interventions, policies, or the characteristics of teachers or principals

The Mechanics of Value Added

1. *Predict*

- ☐ Using prior achievement and other factors predict current scores for each student

2. *Residual*

- ☐ Compute the “residual” which equals the difference between predicted current score and current score

3. *Average*

- ☐ For each teacher average the teacher’s students’ residuals to obtain value added

Many Variations on the General Value Added Approach

- ❑ Variations in the mechanics for making the prediction
 - Variables used in the prediction
 - ❑ Which if any demographics
 - ❑ How many prior test scores
 - ❑ Peer variables, school-level variables
 - Specific model
 - ❑ EVAAS, variations on standard statistical or econometric models
- ❑ Other adjustments – e.g., shrinkage, use of multiple years, corrections for measurement error
- ❑ Different approaches yield estimates of value added that order teacher similarly and are close for many teachers, but some teachers fare quite differently under alternative models

Student Growth Percentiles

- ❑ Many states are choosing to use aggregate *student growth percentiles* (SGPs) for educator evaluations and presumably they will use them for teacher prep programs too
- ❑ An SGP is a measure of the percentile rank of a student's current achievement among students with similar achievement histories
- ❑ The mean or median SGP for a teacher's students serves as a performance metric
- ❑ Like value added:
 - Prior achievement scores are used to identify groups of similar students
 - Data for students linked to a teacher are aggregated (e.g., averaged) to create the measure for the teacher
- ❑ Unlike value added:
 - Instead of predicting current scores, SGPs rank students among similar peers
 - Data for individual students may be reported to students

Use of SGPs Is Hard to Support

- ❑ **SGPs are very inaccurate for individual students**
- ❑ **Some of the errors in SGPs average out in teacher aggregates but errors can remain large**
- ❑ **Aggregate SGPs are susceptible to biases that can be avoided with value added**

Concerns with Value Added: Persistent Error (1)

- ☐ The predicted scores are imperfect
- ☐ Students in different classes differ on their background variables and their potential for growth and achievement
- ☐ People are concerned that the combination of the shortcomings in the predictions and the differences among classroom could result in the errors in predictions being persistent for some teachers
- ☐ Errors in predictions result in errors in value added
- ☐ Consequently, for some teachers value added could have persistent errors leading to persistent incorrect inferences about the teachers' performances
 - That is, for some teachers we consistently over or underestimate value added

Concerns with Value Added: Persistent Error (2)

- ❑ Errors could be persistent across potential classes of students who are likely to be assigned to the teacher or across years
- ❑ The directions of persistent errors may be different across identifiable groups of teachers
 - For example, we might underestimate value added for special education teachers

Concerns with Value Added: Random Errors

- ❑ Even if predicted scores were perfect or students were randomly assigned to classrooms, student growth depends on many factors creating random variation in the average residuals
 - Random errors will tend to average out across students
 - But can make a contribution to estimates of value added because some classes are not large
- ❑ Random errors could lead to year-to-year fluctuations in value added and misclassification when teachers classified into groups

Conditions that Are Necessary for Value Added to have Persistent Errors?

- ☐ **Prediction model must be misspecified**
- ☐ **Persistent differences in the background variables of students in different classrooms**
- ☐ **Errors in prediction model must differ across classrooms**

Empirical Results on Value Added

- ❑ Multiple studies do not find evidence of persistent errors on average across teachers**
- ❑ Some studies have found possible persistent errors, but other explanations also exist for their findings**
- ❑ Strongest studies restrict tests for errors to among teachers within the same schools**
- ❑ Concerns about persistent errors remain for some teachers**
- ❑ These random errors are large relative to variation in teachers inputs so that value added has low reliability**
- ❑ Significant year-to-year variation in scores**

Value Added and Teacher Prep Programs

- ☐ Link public school students to their teachers
- ☐ Link teachers to their preparation programs
- ☐ Creates a link from public school students to teacher prep programs
- ☐ Average residuals for students linked to each prep program

Two Statistical Issues with Using Value Added to Evaluate Prep Programs

- ❑ Accuracy of value added for sampled teachers
 - How well does value added measure the effectiveness of graduates to promote student achievement growth?
- ❑ Utility of sampled graduates for evaluating programs
 - How well does the effectiveness of graduates' teaching in public schools measure the effects of programs?

Accuracy of Value Added for Sampled Teachers

Conditions that Are Necessary for Value Added for Prep Programs to Have Persistent Errors?

- ☐ Prediction model must be misspecified
- ☐ Graduates from different programs must teach students who are systematically different from the students taught by graduates of other programs
 - Not only must students differ across classrooms but graduates from the same program must teach similar types of classroom
 - Might occur if graduates cluster in certain types of schools or regions

Steps Can Be Taken to Reduce Risk of Persistent Errors

- Estimating value added for prep programs offers options for statistical modeling that might reduce persistent errors in value added**
 - Control for schools – compare teachers only within the same school**
 - Control for classroom level variables such as average prior scores**
 - Use best practices of modeling teacher value added**

Random Errors and Value Added for Prep Programs

- ❑ Estimated value added will contain random errors
- ❑ Because random errors tend to average out so their impact decreases with the number of students whose residuals contribute to each estimate
- ❑ Estimates of value added for individual prep programs use residuals of student linked to multiple graduates from the program so typically many more students used in prep program value added than individual teacher value added

Errors in Value Added for Prep Programs

- ❑ The risk of persistent error does not appear great
 - Empirical evidence on teacher value added does not find evidence of persistent errors
 - Models can take further steps to reduce risk
- ❑ Typically, at least a modestly large samples of students are used in the estimation of value added for individual programs

Utility of Sampled Graduates for Evaluating Programs

What Do We Want to Learn from Value Added for Prep Programs

- ☐ Quality of the program's training
- ☐ Quality of the program's graduates
- ☐ Quality of the program's graduates who tend to secure teaching jobs in the state

Quality of the Program's Training

- ☐ Sensitive to the training activities of the program
- ☐ “Value added” of the program
- ☐ Answers the question: Is a candidate trained at this program better prepared and better able to teach than if that candidate had attended a different program?
- ☐ Could be used to support improvement efforts for program training

Quality of the Program's Graduates

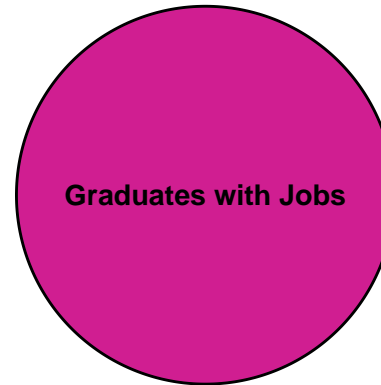
- ☐ Sensitive to the training and “filtering” activities of the program
 - ☒ Reflects program selectivity for entrants and stringency for graduation
- ☐ Answers the question: Will a graduate from this program better able to teach in a given position than graduates from other programs would be?
- ☐ Source of information about the program's contribution to the potential teacher labor force
- ☐ May be useful to school administrators who hire teachers
- ☐ May be important for accreditation to ensure quality of the product

Quality of the Program's Graduates Who Tend to Secure Teaching Jobs in the State

- ☐ Sensitive to training, filtering, and who gets jobs
- ☐ Answers the question: Will a graduate from this program who gets a job be better able to teach in that position than graduates from other programs would be?
 - There is ambiguity about which “other programs” should be considered: all other programs or other programs whose graduates might fill the position
- ☐ Source of information about the program's contribution to the potential teacher labor force
- ☐ May be useful to school administrators who hire teachers
- ☐ Which candidates secure jobs is outside of the control of the program

Target Population for Each Question

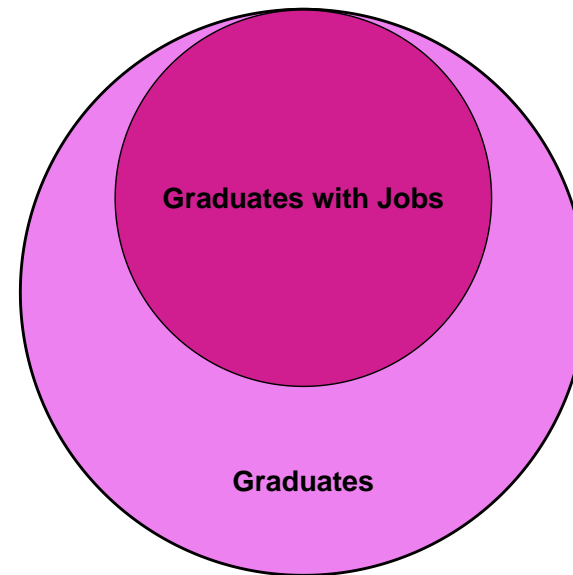
Q3 All graduates hired in the state



Target Population for Each Question

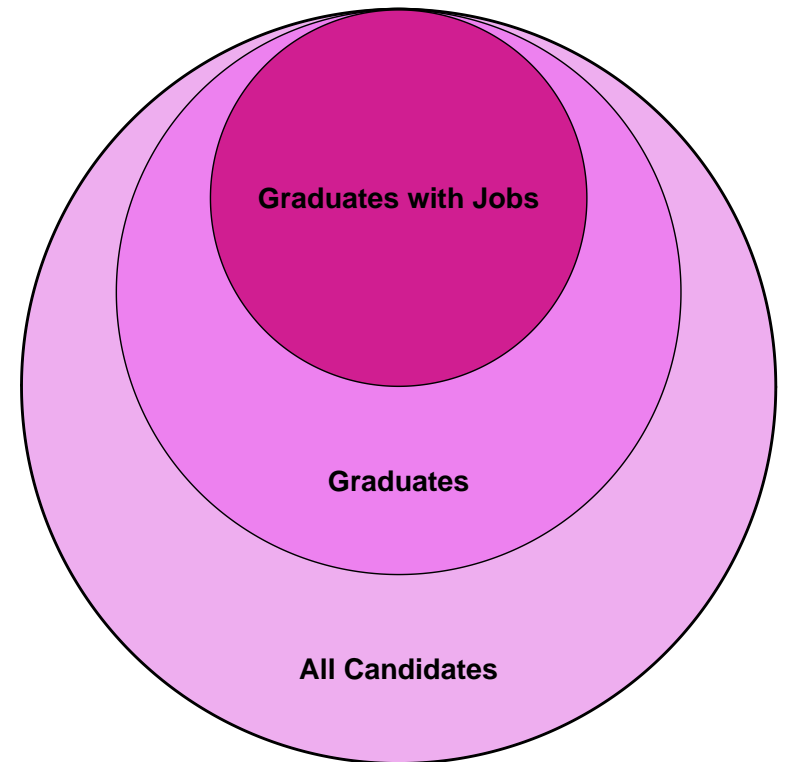
Q3 All graduates hired in the state

Q2 All graduates of the program



Target Population for Each Question

- Q3 All graduates hired in the state**
- Q2 All graduates of the program**
- Q1 All candidates who start the program**



Risks for Confounding for Each Question

Q3 **Graduates may secure atypical set of jobs**

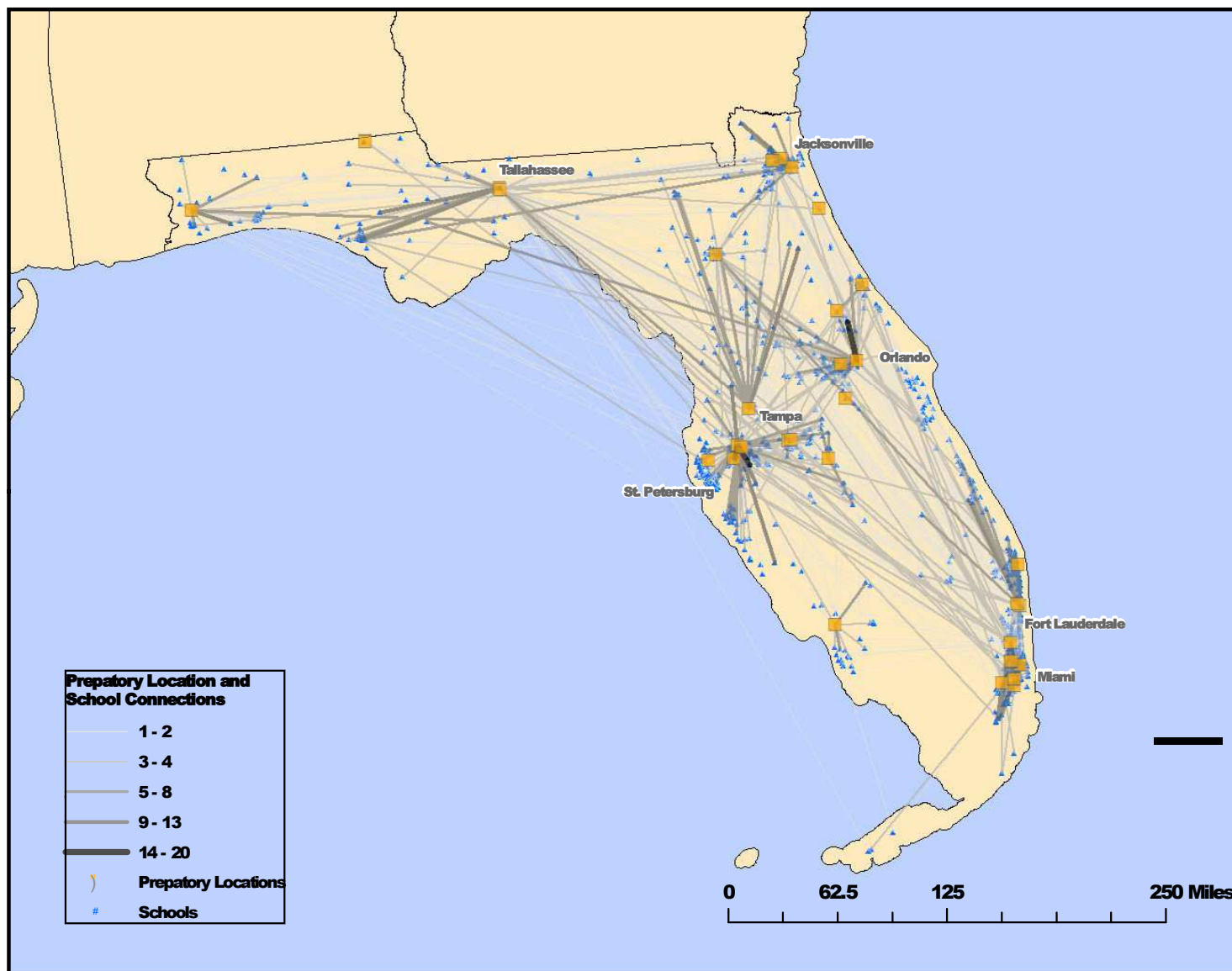
Q2 **Nonrepresentative set of graduates may obtain jobs**
Graduates may secure atypical set of jobs

Q1 **Nonrepresentative set of candidates may select the program**
Nonrepresentative subset of candidates who start
the program may finish it
Nonrepresentative set of graduates may obtain jobs
Graduates may secure atypical set of jobs

Where Graduates Work

- ❑ Graduates tend to work in schools geographically clustered around the prep program
- ❑ In Florida, the schools where graduates from different programs worked differed in terms of student demographics and student achievement

Where Graduates Work



School Environment

- Professional development after graduation
 - Teacher value added improves substantially during initial years of teaching**
 - Teachers in different schools or districts may have different professional development, initiation, and mentoring experience that affect this development****
- Different schools use different curricula and have different resources**

Graduates with Teaching Positions

- ❑ Not all program graduates obtain a teaching position in public schools
 - In some states large percentages of their teacher prep graduates do not teach in the states' public schools
 - For example, 50% of graduates in Ohio and 75% graduates in Delaware do not teach in state
- ❑ Many teachers drop out of teaching during their first few years; those who remain may not be representative of all the programs' graduates
- ❑ Many graduates substitute and work part time before securing a position

Candidates Who Complete Program

- ☐ Not all candidates who enter a program complete it
- ☐ The candidates who drop out of the program are not a random sample of the candidates
- ☐ Can create differences among programs in the potential effectiveness of their graduates that is not a result of the program training

Candidates Who Enter the Program

- ❑ The background characteristics of candidates differ across programs
- ❑ Universities are not equally selective
 - For Michigan public universities 25th percentile ACT composite scores range from 17 to 28 and 75th range from 23 to 28
- ❑ Teacher prep programs have differing entrance requirements

Analytic Fixes for Potential Confounding

- ❑ Compare graduates working in the same schools
 - This is feasible through statistical modeling
 - Does require assumptions such as graduates working in any school are a random sample of program graduates
- ❑ For Q1 and Q2, collect data on candidate during program and control for this when comparing graduates who have jobs
- ❑ For Q1, use state longitudinal data system to control for candidates background when comparing across programs
- ❑ For Q1, correcting for students who drop out can be tricky, collecting data during training would be helpful

Sample Size

- ☐ Many programs produce few candidates per year
- ☐ Restricting to teachers with jobs in selected grades and subjects can further reduce samples
- ☐ Variability among students will contribute to the variability in the program measures
- ☐ With few candidates this contribution can be large
- ☐ Combining data across years can help but it has complications

Other Issues

- ❑ Value added is not the only outcome of interest
 - Teacher value added is sensitive to the test
 - Value added predicts other outcomes but does not account for a large share of the variance
- ❑ Many different programs may co-exist at an institution or within a department or college
- ❑ Content training may exist outside of the “program”
- ❑ The teachers who are eligible to contribute the measure must be identified
 - Should it be graduates within a fixed number of years of graduation or within a fixed number of years starting teaching?
 - Should teachers teaching out-of-subject areas be included?

Is There Any Value in Such Program Measures

☐ Hard to support for formal evaluations

- Even if we want to answer Q2 about quality of graduates potential confounders seem very plausible
- Any use should have supporting data to demonstrate comparability of program graduates or candidates and validity of methods used to adjust for differences

☐ May have utility for formative purposes

- Provides external measures on what may be insular organizations
- Most likely it provides a reasonably accurate measures on teaching for a subset of program graduates
- Interpretation requires judgement and knowledge of program that program administrators can obtain
- Identify places to start conversations