



**CEDR**

Center for Education  
Data & Research

# Using Student Outcomes to Assess Teacher Preparation: What Do We Know?

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

# Context

2

- 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”

# My Own View

3

- Traditional training programs need to change or they will continue to prepare a declining share of new teachers
  - I'm a bit skeptical that we will see big changes given the politics and incentives faced by universities, but I hope I'm wrong
- There is (potentially significant) room for improvement in teacher preparation
  - We don't know much about the connection between features of training and student outcomes
  - Teachers clearly improve with experience
- Improved teacher training key to teacher workforce improvements (low on political cost, potentially high on learning)

# Road Map for Rest of Talk

4

1. What do we know about using student test results (value added) to assess teacher preparation providers (TPPs)
2. What more needs to be know about assessing teacher training?
3. How might analysis be used for program *improvement*: some case study examples
4. What *can't* be resolved based on empirical evidence
5. Closing thoughts and discussion

# Limitations of Using Student Growth as a Measure of TPP Performance

5

- 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!**

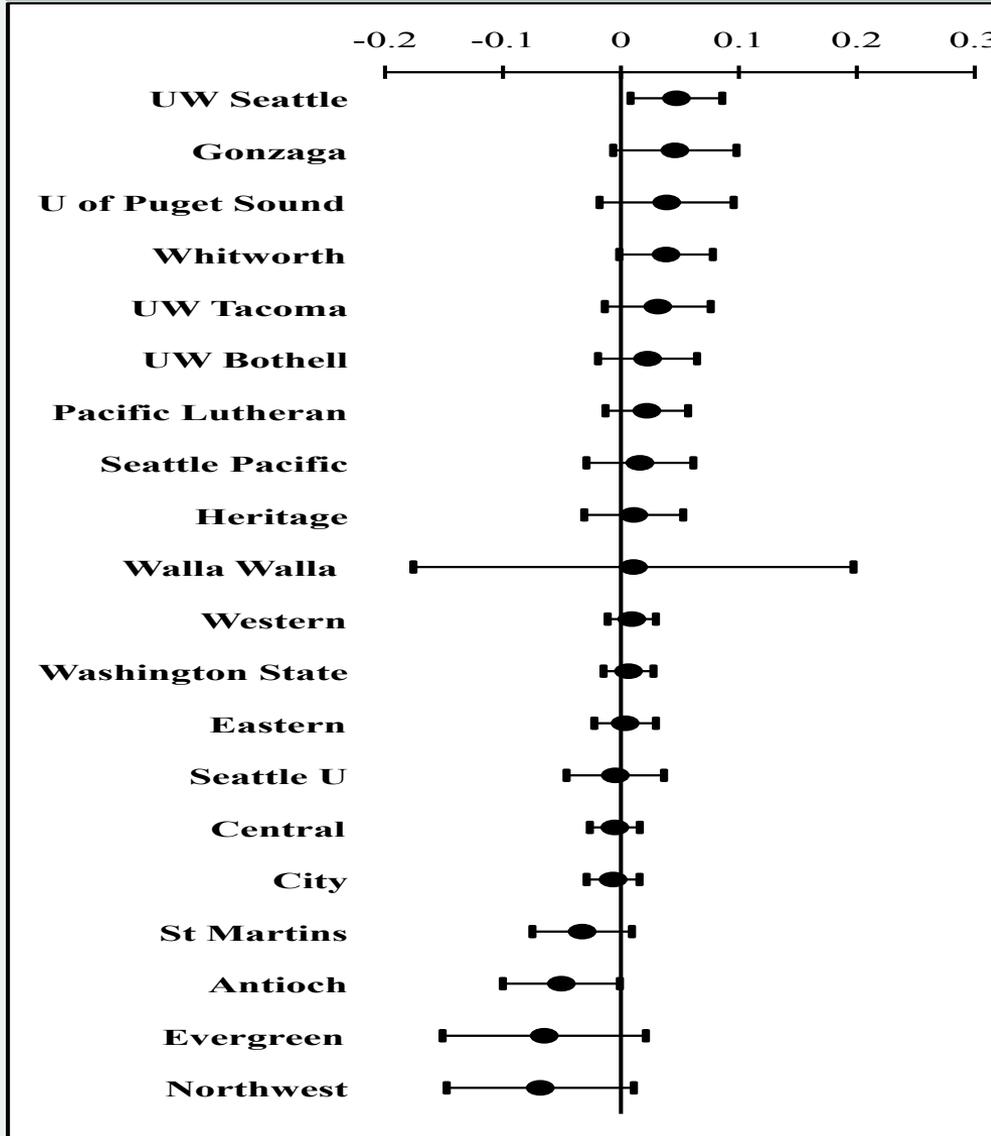
# What's Known About VAM TPP Estimates?

6

- Empirical research across states (FL, LA, MO, NC, NY, WA) about within state differences in graduates from different programs 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

# Decay Program Estimates (90% CIs): MATH

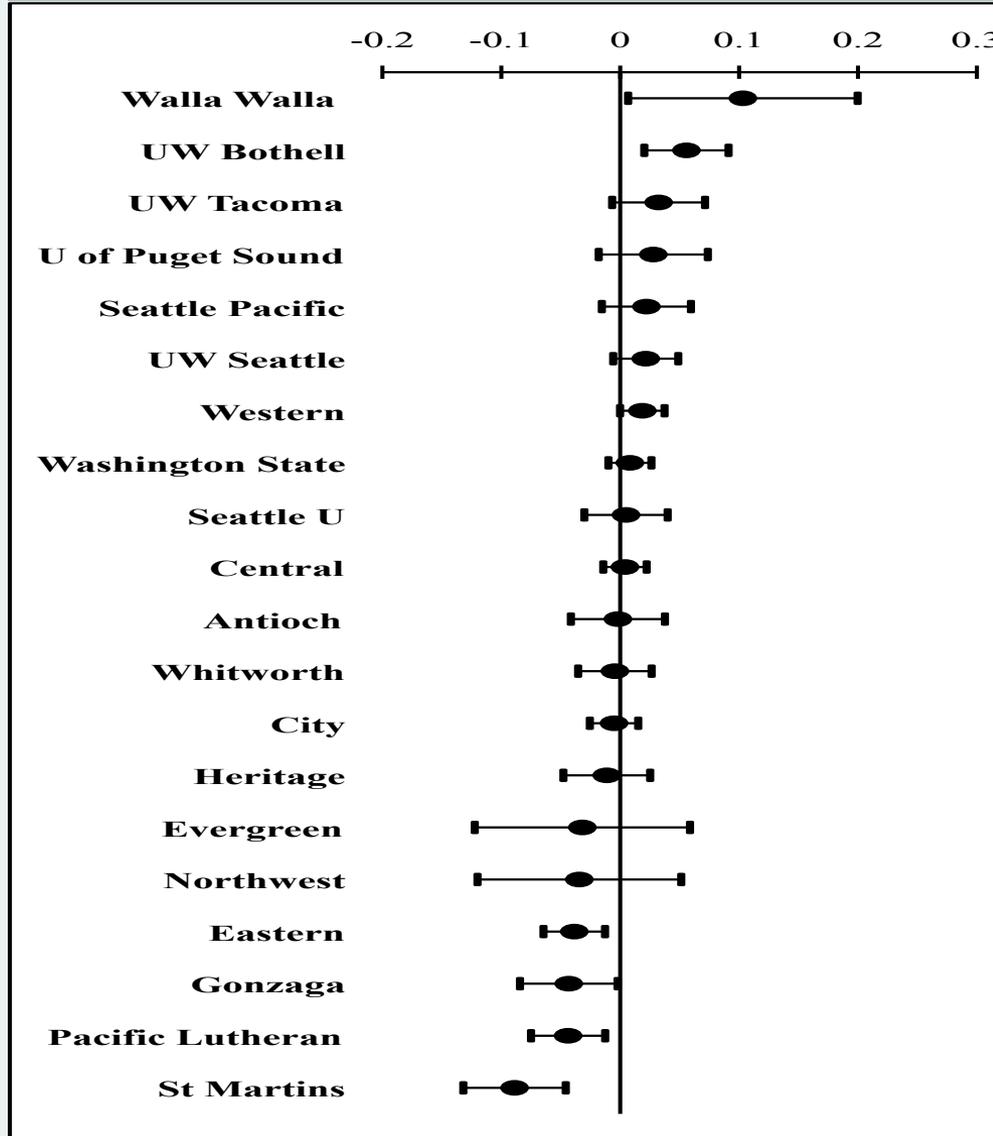
7



- The regression-adjusted difference between the average program and the top program is 4% of a SD of student performance
- The regression-adjusted difference between the top program and bottom program is 12% of a SD of student performance

# Decay Program Estimates (90% CIs): READING

8



- The regression-adjusted difference between the average program and the top program is 10% of a SD of student performance

- The regression-adjusted difference between the top program and bottom program is 19% of a SD of student performance

# Statistical Issues Associated w/ Estimating TPP Effects

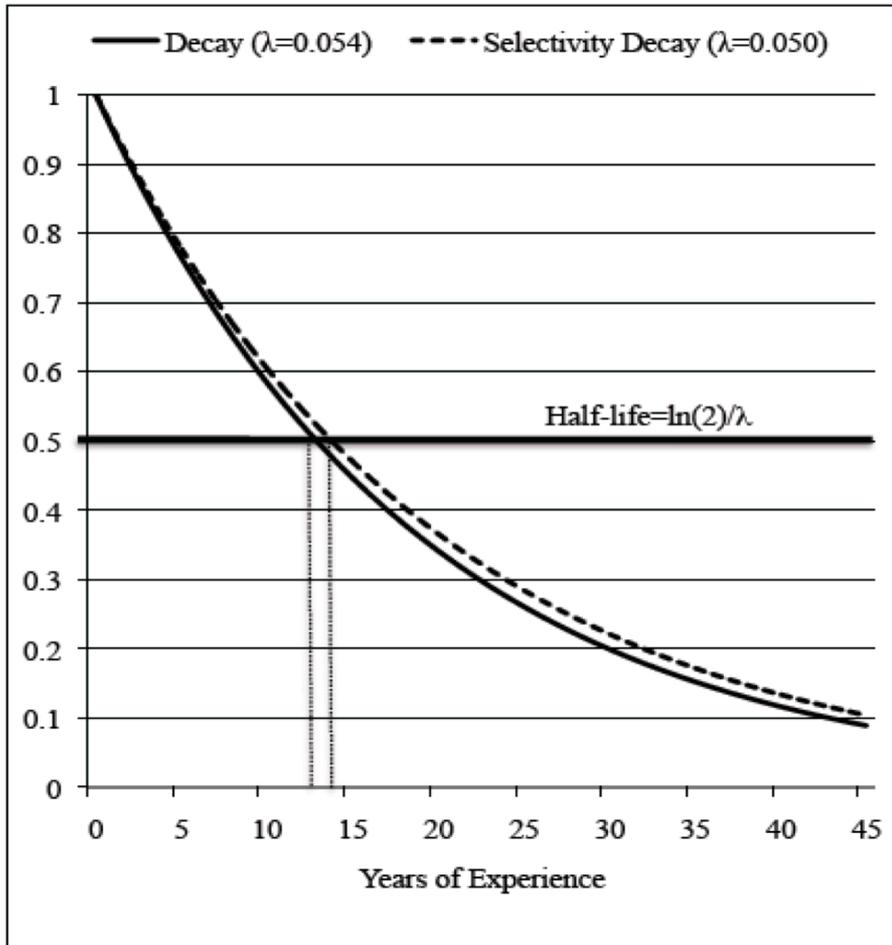
9

- 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

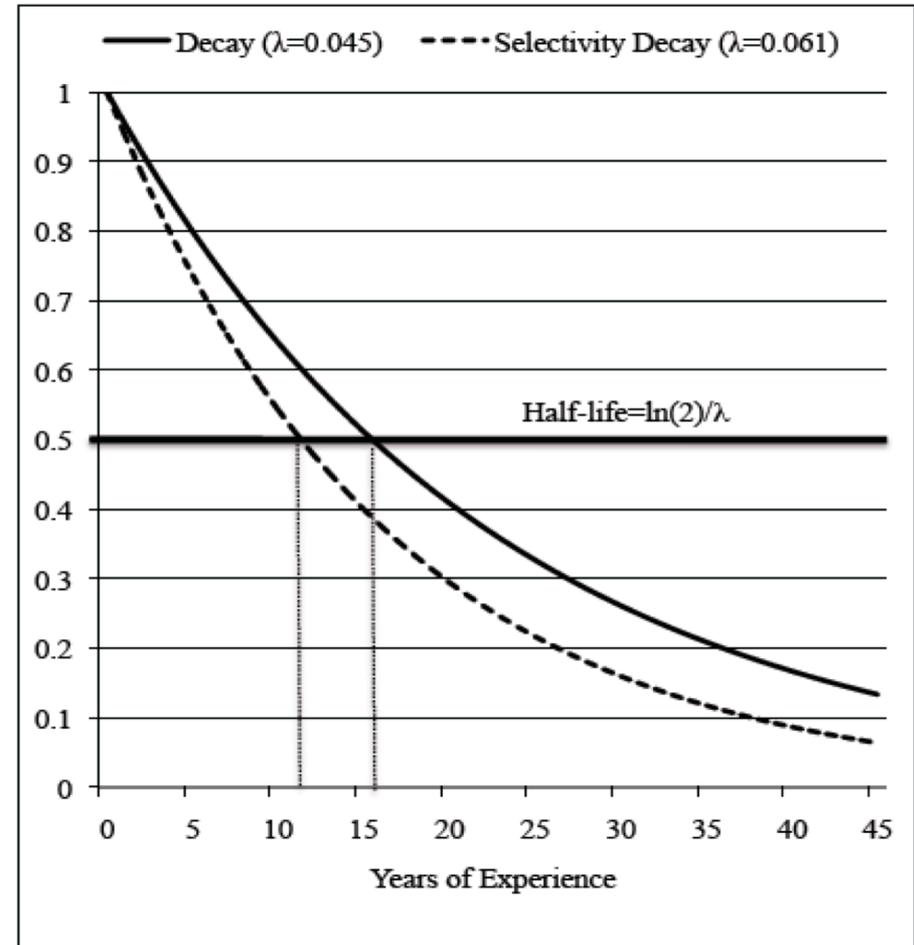
# Decay of TPP Effect Estimates

10

## MATH



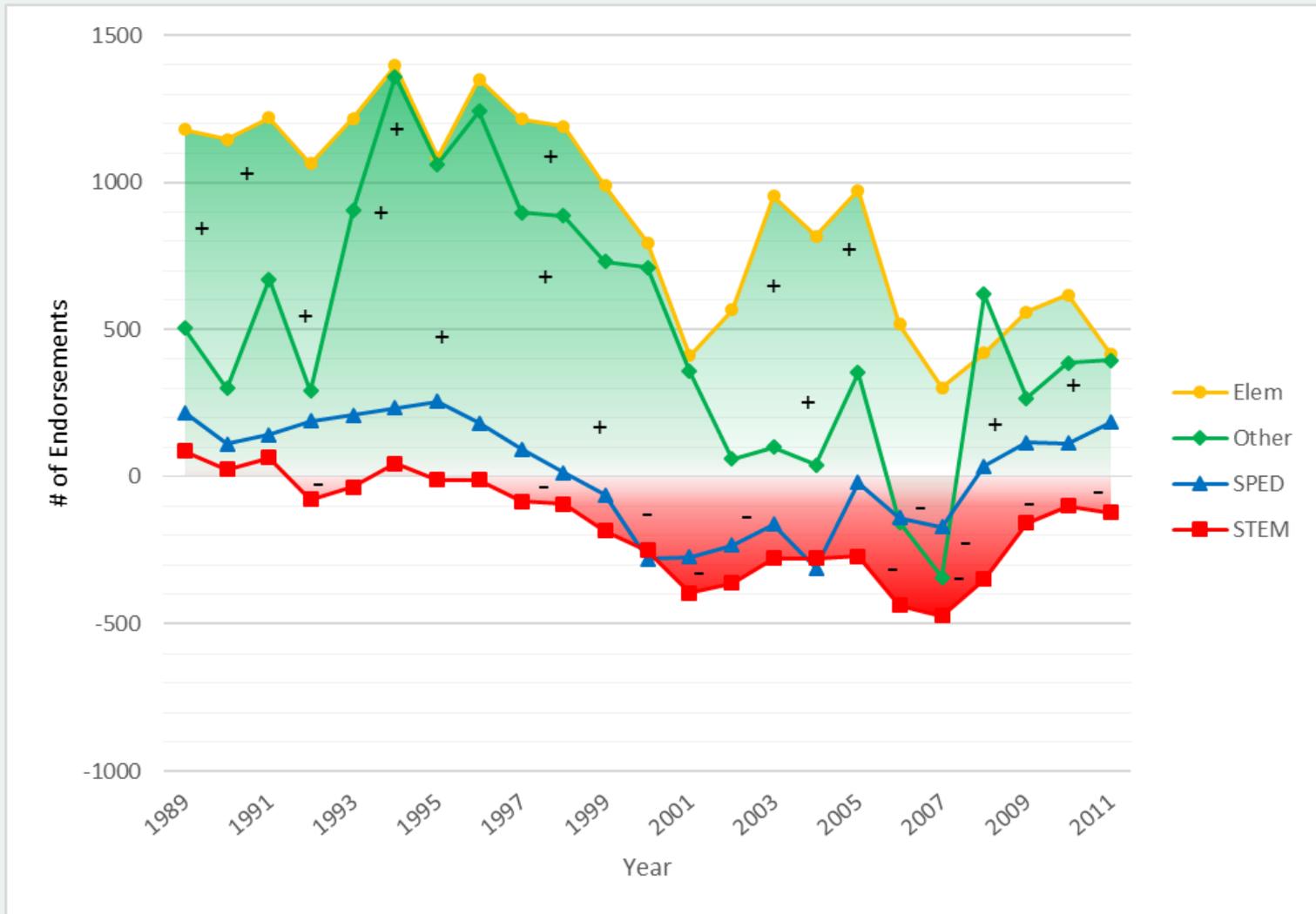
## READING



# Important non-VAM Issues

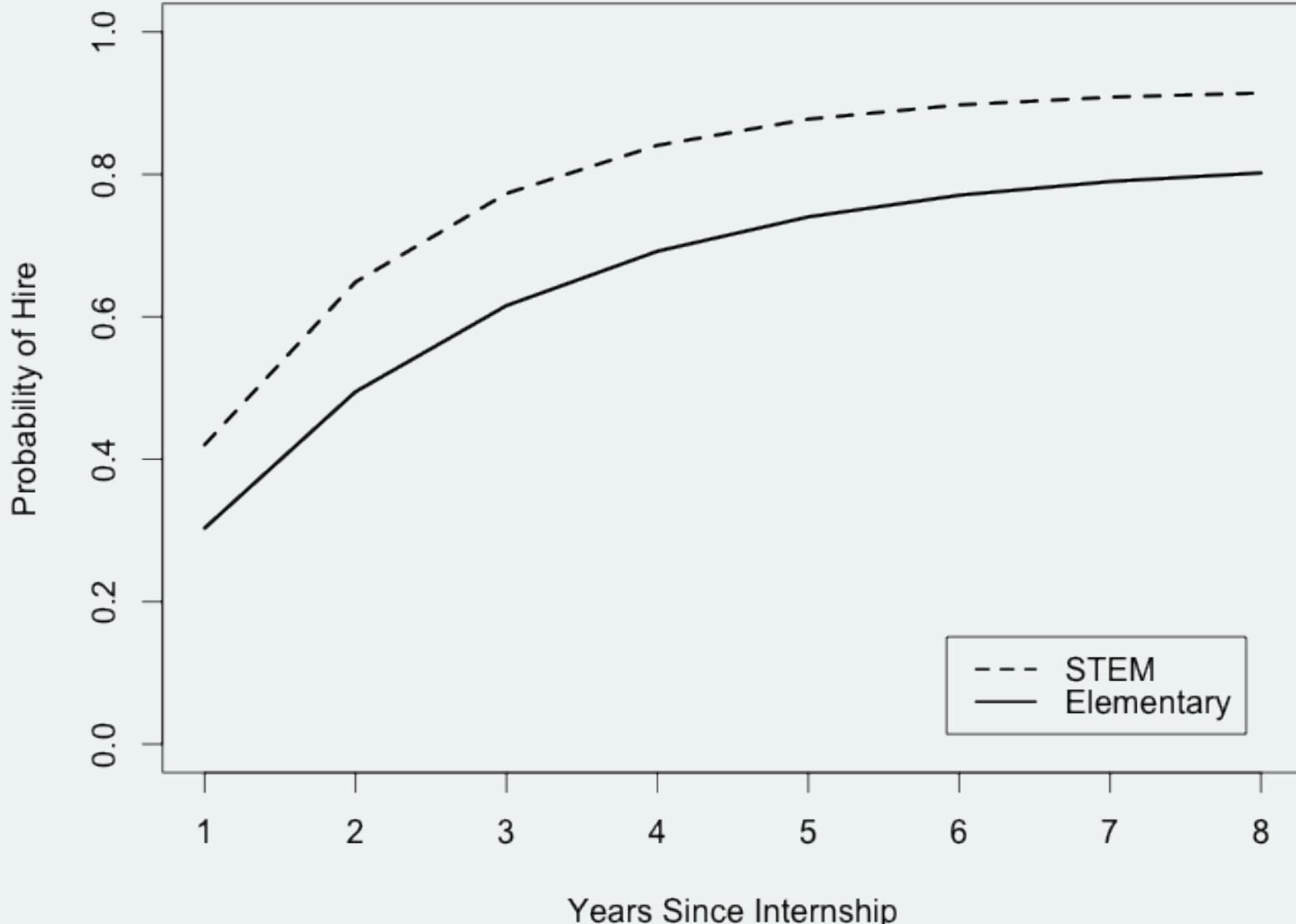
- Turnover is costly, in dollars and student achievement, how do programs and program features affect teacher retention?
  - Some evidence of important differences in teacher mobility by program (Goldhaber and Cowen, 2013), satisfaction with program (DeAngelis et al., 2013), and training features (Ronfeldt et al., 2013)
- There are chronic teacher labor market imbalances, finding STEM and special education teachers is especially difficult
  - Production of teachers doesn't comport with labor market need

# Net production by endorsement (production – attrition)



# Modeling Time and Probability of Hire: STEM vs. Elementary Ed Interns

13



# What More Needs to be Known?

14

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

# Research to Inform Program Improvement

15

- Relatively little quantitative research on the features of TPPs that are associated with student achievement, but what does exist offers *suggestive* evidence that some features may matter
  - Some licensure tests are predictive of student achievement raising questions about selectivity of ed schools
  - Student teaching environment (measured by “stay ratio”) predictive of teacher effectiveness
  - Strong connection between student teaching and methods coursework also predictive of effectiveness
- We need to be more thoughtful about how we learn what matters for teacher training

# Teacher Training Learning Collaborative

16

- Stay ratio negatively correlated with prob of employment
  - 15% of teachers hired into schools in which student-teaching occurred
  - Important implications for equity of teacher distribution
- Few characteristics of cooperating teacher are significant
  - *Negative* correlation between prior experience of cooperating teachers supervising interns and probability of hire; are school systems compensating for teacher weaknesses?
- Licensure scores are not predictive of hiring in general, but *are* predictive of student achievement and hiring into one's internship school
  - Student teaching as a screening device

# Learning About Novice Teachers Before They are Hired

|                          | Dep. Var.: Math Value-Added |          |         |        | Dep. Var.: Reading Value-Added |       |        |          |
|--------------------------|-----------------------------|----------|---------|--------|--------------------------------|-------|--------|----------|
|                          | 1                           | 2        | 3       | 4      | 5                              | 6     | 7      | 8        |
| 21-Pt Screen             | 0.044***                    |          |         | 0.024  | 0.047***                       |       |        | 0.048*** |
| 54-Pt Screen             |                             | 0.064*** |         |        |                                | 0.021 |        |          |
| Certificate & Education  |                             |          | -0.004  | -0.012 |                                |       | -0.019 | -0.015   |
| Training                 |                             |          | 0.032   | 0.041* |                                |       | -0.009 | -0.009   |
| Experience               |                             |          | -0.034  | -0.027 |                                |       | 0.005  | 0.021    |
| XXXXXXXXXXXXXXXXXX       |                             |          | 0.076** | 0.075* |                                |       | 0.026  | 0.017    |
| Flexibility              |                             |          | -0.024  | -0.028 |                                |       | 0.025  | -0.007   |
| Instructional Skills     |                             |          | 0.044   | 0.014  |                                |       | 0.017  | 0.005    |
| Interpersonal Skills     |                             |          | -0.01   | 0.021  |                                |       | -0.023 | 0.005    |
| Cultural Competency      |                             |          | -0.011  | -0.024 |                                |       | 0.001  | -0.001   |
| Preferred Qualifications |                             |          |         |        |                                |       | -0.006 |          |
| N                        | 282                         | 208      | 203     | 172    | 296                            | 211   | 187    | 174      |
| R-sq                     | 0.043                       | 0.083    | 0.156   | 0.171  | 0.095                          | 0.013 | 0.053  | 0.113    |
| adj. R-sq                | 0.04                        | 0.078    | 0.122   | 0.125  | 0.092                          | 0.008 | 0.005  | 0.064    |

# What Can't Be Resolved Based on Empirical Evidence?

18

- To what extent value-added or output measures should be used 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

# Closing Thoughts

- 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

# References/Resources

- <http://cedr.us/publications.html>
  - “Knocking on the Door to the Teaching Profession? Modeling the entry of Prospective Teachers into the Workforce.” CEDR Working Paper 2013-2.3.
  - “The Gateway to the Profession: Teacher Preparation Programs Based on Student Achievement. CEDR Working Paper no. 2012-4.”
  - “The STEM and Special Education Teacher Pipelines: Why Don't We See Better Alignment Between Supply and Demand?” CEDR Working Paper 2014-3.
  - “Excavating the Teacher Pipeline: Teacher Training Programs and Teacher Attrition.” CEDR Working Paper 2013-5.