Using Research and Local Expertise to Improve Teacher Evaluation Systems

Corinne Herlihy, National Center for Teacher Effectiveness
September 4, 2014
Workshop Agenda

• Introduction
• Research Overview Presentation
  Corinne Herlihy, NCTE
• Practitioner-Researcher Presentation
  Nate Schwartz, Tennessee Department of Education
• Break (15 min.)
• Case study exercise
  • Introduction
  • Examination of the case
  • Extract and apply lessons
• Wrap up
NCTE: Who we are

• Project housed at the Center for Education Policy Research at the Harvard Graduate School of Education, funded by IES
• GOAL: Uniting two strands of research centering around teacher effectiveness
  • Educational economists – focus on value-added and education production function (i.e., teacher experience, degree, preparation)
  • Math educators – focus on instruction, knowledge, teachers’ locus of control

• Can we better understand the different components of teacher effectiveness from these research traditions, and how they relate to one another?
NCTE: Who we are

• Collected data on multiple measures of teacher effectiveness from:
  • 300+ fourth- and fifth-grade math teachers
  • 4 large urban districts on the East Coast
  • Three school years (2010-11 to 2012-13)

• Data sources:
  • Teacher value-added (state math/ELA tests, project math test)
  • Videotapes of instruction (3 per year, ~1700 total)
  • Teacher background information
  • Teacher knowledge
  • Student perceptions
What we’ve learned so far: Understanding Mathematics Instruction

Lesson scores on codes from two different observational instruments – the Mathematical Quality of Instruction (MQI) and the Classroom Assessment Scoring System (CLASS) seem to form 4 distinct factors of instruction...

- Ambitious Mathematics Instruction (i.e., multiple representations, multiple solution pathways, teacher and student mathematical explanations and reasoning)
- Mathematical Errors and Imprecisions
- Classroom Organization (i.e., behavior management, negative climate, productivity)
- Classroom Support (i.e., emotional and instructional support, student engagement)
What we’ve learned so far: Understanding Mathematics Instruction

We have been able to identify some variables associated with teachers’ instruction scores:

• More Ambitious Math Instruction (# of math methods/content courses, math knowledge, knowledge of students)
• Fewer Math Errors (Math knowledge, less preparation for math class)
• Stronger Emotional Support (Elementary math certification, higher efficacy)
• Stronger Classroom Organization (More experience, more test-prep behaviors)
Mathematical Quality of Instruction (MQI)

Ambitious Instruction

- Richness of the Mathematics
- Working with Students and Mathematics
- Common Core Aligned Student Practices
- Math Errors and Imprecisions

Math Errors and Imprecisions

Classroom Support

- Emotional Support
- Instructional Support
- Student Engagement
- Classroom Organization

Classroom Assessment Scoring System (CLASS)
What we’ve learned so far: Understanding Mathematics Instruction

For the most part, we don’t see extremely ambitious mathematics instruction (i.e., lessons with rich mathematics and common core aligned student practices, devoid of errors), nor do we see completely problematic instruction...
What we’ve learned so far: Understanding Mathematics Instruction

... but this changes depending on context.
What we’ve learned so far: Understanding Mathematics Instruction

... but investigations into the functioning of the MQI suggests some problems. Using IRT, we saw for example, that the only teachers who are most likely to get ‘High’ for Explanations are those with underlying abilities at almost 4 SDs above average.
What we’ve learned so far:
Relationship to student achievement

We found it hard to predict teachers’ value-added scores, when blind to the scores, and judging from just their math instruction.
What we’ve learned so far: Relationship to student achievement

Teacher value-added scores also mean different things in terms of instruction from district to district.

• In District 11, the gap in instructional quality between high and low value-added teachers was stark; in District 14, the variation was much smaller.

• The instruction of the highest value-added teachers in District 11 was much stronger than the instruction of the highest value-added teachers in District 13. In fact, the instruction of the lowest value-added teachers in District 11 was comparable to that of the highest in District 13.
What we’ve learned so far:
Relationship to student achievement

We find that the relationship between teachers’ ambitious math instruction and their value-added varies from district to district.
What we’ve learned so far:

Relationship to student achievement

We are able to explain a non-negligible amount of variation in teacher effects on math achievement using our measures. Multiple measures work together in explaining this variation.

<table>
<thead>
<tr>
<th>#</th>
<th>Model</th>
<th>Measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Demographics</td>
<td>Race, Gender</td>
</tr>
<tr>
<td>2</td>
<td>Background</td>
<td>Math Courses, Educational Attainment, Experience, Certification</td>
</tr>
<tr>
<td>3</td>
<td>Instruction</td>
<td>Richness of Math, Mathematical Errors, Student Math Common-Core Aligned Practices, Classroom Organization, Emotional and Instructional Support, Student Engagement</td>
</tr>
<tr>
<td>4</td>
<td>Capacity</td>
<td>Knowledge of Student Abilities and Tendencies, General and Specialized Math Knowledge</td>
</tr>
<tr>
<td>5</td>
<td>Other</td>
<td>Effort, Test Prep Behaviors, Self-Efficacy, Formative Assessment, Content Coverage (Numbers, Algebra)</td>
</tr>
<tr>
<td>6</td>
<td>M3+M4+M5</td>
<td>Non-Demo/Background Measures</td>
</tr>
<tr>
<td>7</td>
<td>M1+M2+M6</td>
<td>All Measures</td>
</tr>
</tbody>
</table>
What we’ve learned so far: Relationship to student achievement

<table>
<thead>
<tr>
<th>#</th>
<th>Model</th>
<th>State Test</th>
<th>NCTE Low-Stakes Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Demographics</td>
<td>1.4%</td>
<td>2.3%</td>
</tr>
<tr>
<td>2</td>
<td>Background</td>
<td>8.0%</td>
<td>12.9%</td>
</tr>
<tr>
<td>3</td>
<td>Instruction</td>
<td>7.0%</td>
<td>6.8%</td>
</tr>
<tr>
<td>4</td>
<td>Capacity</td>
<td>6.7%</td>
<td>21.3%</td>
</tr>
<tr>
<td>5</td>
<td>Other</td>
<td>8.2%</td>
<td>6.2%</td>
</tr>
<tr>
<td>6</td>
<td>Non-Demo/Background</td>
<td>19.6%</td>
<td>29.5%</td>
</tr>
<tr>
<td>7</td>
<td>All Measures</td>
<td>28.1%</td>
<td>40.3%</td>
</tr>
</tbody>
</table>
What we have yet to learn

• Most of analyses thus far have been correlational
• In the final year of NCTE (SY 2012-2013), we randomly assigned 68 rosters of 1,612 students to teachers in 30 blocks to investigate causal relationships between measures
  • Many students and a few teachers ‘left’ the study (i.e., changed grades, moved out of district)
  • Approximately 80% of students who were not originally hand-placed and did not leave the study because of ‘random’ reasons ended up with their ‘random-assignment’ teacher at the end of the year
  • Students who switched were less likely to be eligible for subsidized lunches, less likely to be ELL students, and less likely to be Hispanic
Three papers to consider

  - [http://edr.sagepub.com/content/current](http://edr.sagepub.com/content/current)


  - [http://www.nber.org/papers/w19423](http://www.nber.org/papers/w19423)
The test matters:
The relationship between classroom observation scores and teacher value added on multiple types of assessment

• Other studies generally find modest to low correlations between observational data and VAMs
• Low correlations may reflect a lack of alignment between the goals of particular teaching practices captured by observational protocols and the kinds of student outcomes measured by many standardized tests
• Different student assessments may be differentially sensitive to more ambitious forms of instruction measured by classroom observation instruments
• Lack of strong relationship between different measures of effectiveness creates a dilemma for teachers and districts and a puzzle for researchers
Protocol for Language Arts Teaching Observation (PLATO)

Factors

1. Instructional Scaffolding
   - Modeling
   - Strategy Instruction

2. Disciplinary Demand
   - Intellectual Challenge
   - Classroom Discourse

3. Classroom Environment
   - Time Management
   - Behavior Management
Research questions

1. How does the overall relationship between scores of PLATO teaching practices and teacher VAMs differ depending upon the assessment used to construct VAM?

2. What are the relationships between specific instructional factors as measured by PLATO and teacher value-added scores on different student assessments?
Is a good teacher a good teacher for all?
Comparing value-added of teachers with their English learners and non-English learners

Researchers assess the extent to which a teacher’s effectiveness at improving student performance in math and reading is similar for English learners (EL) and their non-EL counterparts.

1. How much does teacher effectiveness vary across classrooms for EL and non-EL students?
2. Are teachers equally effective with ELs and non-ELs?
3. Can measured teacher characteristics help explain differences in teacher effectiveness?
Measuring the impacts of teachers I:
Evaluating bias in teacher value-added estimates

- Do the differences in test-score gains across teachers measured by value-added capture causal impacts of teachers or are they biased by student sorting?
- Using school district and tax records for more than 1 million students, find that VA models which control for a student’s prior test scores exhibit little bias in forecasting teachers’ impacts on student achievement.
- Researchers test for bias in two ways
  - Estimating bias using observable characteristics
    - Does not rule out the possibility that students are sorted to teachers based on unobservable characteristics orthogonal to parent characteristics and lagged score gains.
  - Estimating bias using teacher switching quasi-experiments
    - Exploits naturally occurring teacher turnover to estimate forecast bias
Case Learning Objectives

• Better understand the complexities and challenges of implementing initiatives to improve instructional practice of teachers and learning outcomes for students.
  • As researchers, build strategies to support his work in districts.
  • As practitioners, identify opportunities and strategies to better align this work in your own organization.

• Examine how the components of these initiatives by curriculum leader and teacher effectiveness specialists influence successful outcomes for teachers and students.

• Understand the importance of external factors in supporting implementation and success.

• Consider the potential for alignment between the two initiatives.

• Develop recommendations for effective methods to track outcomes and measure effectiveness, including identifying which data to collect and which analyses to conduct.
Our motivation for the case...