Teacher and Principal Evaluation in Tennessee

April 22, 2014

Sara Heyburn
Assistant Commissioner, Teachers and Leaders
Tennessee Department of Education

First to the Top Act, January 2010

- Annual evaluations for all educators, multiple measures
  - First full year of implementation was 2011-12
  - This year marks the third year of statewide implementation

- 50% based on student outcomes:
  - 35% TVAAS or alternative growth measure*
  - 15% other achievement measure

- 50% based on observation*

- Requires annual evaluation be **a factor** in personnel decisions:
  - Promotion
  - Retention
  - Tenure
  - Compensation

*Note: For teachers in non-tested grades and subjects without an individual growth score, student growth now counts 25% and observation counts 60%.

Sea Change

TEAM, the state's new evaluation system, has played a pivotal role in facilitating feedback and continuous improvement for all educators.

Old system:

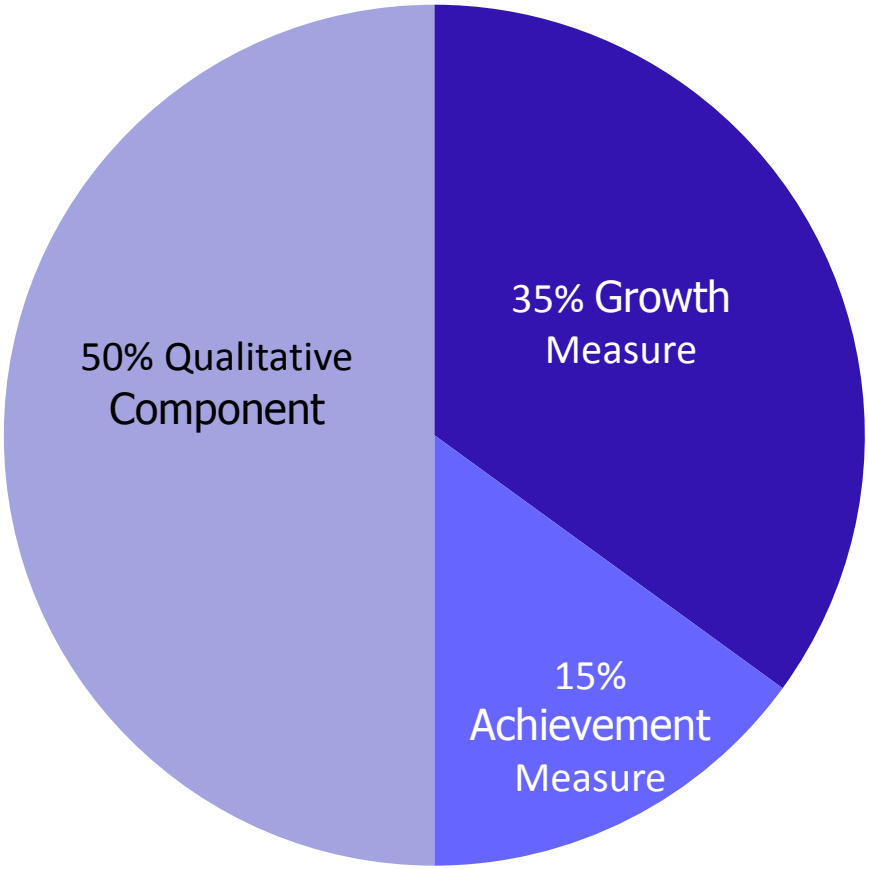
- Tenured teachers evaluated twice every 10 years
- No meaningful feedback or targeted improvement

New system:

- Every teacher, every year receives multiple observations
- Feedback tied to clear vision for effective instruction
- Differentiation among strong and struggling teachers

TN's TEAM evaluation system includes multiple measures and five performance levels

Components



Levels

5: Significantly above expectations

4: Above expectations

3: At expectations

2: Below expectations

1: Significantly below expectations

Reflections on Early Implementation

- Initial challenges
  - Communication
  - Time
  - Accurate scoring and high-quality feedback
- Listening Tour
  - Partnered with SCORE upon Governor's charge
  - Engage educators and gather input
- Year 1 report issued – July 2012
  - Modeling continuous improvement
  - Feedback and data drove changes to the system

Balancing Support and Accountability

- Misalignment index and targeted support coaches
- Annual reports on implementation
- Refining regional support structures

Looking Ahead

- Pushing to the next level
  - 360° cameras
  - Student surveys
  - More attention to emerging best practices
  - Additional growth measures for teachers in non-tested areas
  - Revised principal evaluation

STRATEGIC DATA PROJECT

Teacher Evaluation Plenary

2014 Annual Convening

www.gse.harvard.edu/sdp/

Boston, MA | April 22, 2014

Center for Education Policy Research
HARVARD UNIVERSITY

STRATEGIC DATA PROJECT

Challenge

- New instruments and measures
- Implementation
- High stakes

Opportunity

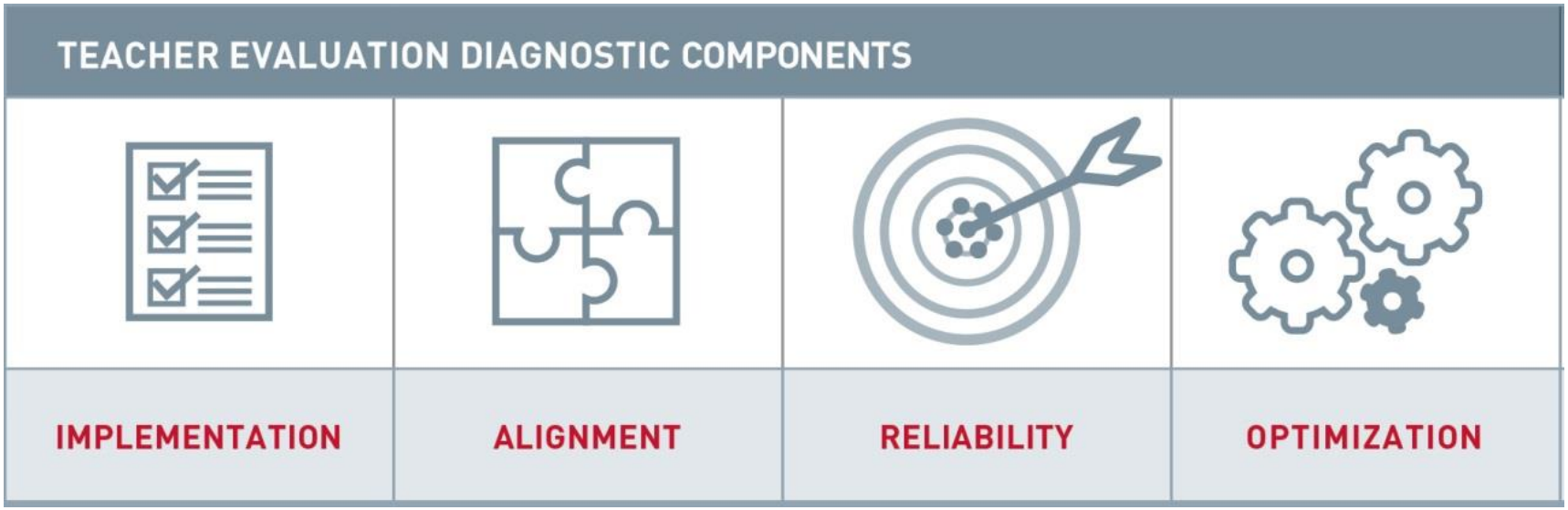
- Policy reforms
- Data infrastructure
- **Missing link?**

  <u>Analysis!</u>

Center for Education Policy Research
HARVARD UNIVERSITY

STRATEGIC DATA PROJECT

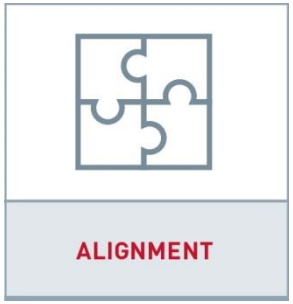
SDP's Teacher Evaluation Diagnostic (TEvD)



*An **analytical tool** designed to **embrace the opportunity** by **addressing the challenges***

The TEvD: Sample Questions

- Are there districts where there is substantially more or less variation in teachers' scores within each component?



- Do teachers' scores on non-growth components (e.g., observation, SLO/SGO, student surveys, etc.) predict individual growth?

The TEvD: Sample Questions

- Are classroom observations providing reliable information about a teacher's effectiveness?

- Are the cut points that define score categories for school growth measures (e.g., median growth percentiles) equitable across schools of different sizes?

TEvD pilot with TN

Analysis

- The relationship/alignment between average **individual growth** scores (i.e., value-added) and average **classroom observation** scores in K-12 public schools across TN

Motivation

- Share emerging results
- Demonstrate TEvD analysis, interactive graphing

Final Words

Thanks!

- TN collaborators: Nate, Tony, others
- SDP researchers: Andrew, Colin
- Fellows: Andy Baxter, Tom Tomberlin, James Riddlesberger
- Advisors: Tom Kane, Doug Staiger, Jon Fullerton

Technical roundtable

Opportunities to collaborate: talk to Nick Morgan

Teacher and Principal Evaluation in Tennessee

April 22, 2014

Tony Pratt and Nate Schwartz, SDP Fellows

Our Approach to Teacher Evaluation Analysis

- Identifying the **right questions** and completing analyses on the **right timeline**
- All analyses fall in two broad categories
  - Analyses that inform the continued **design and implementation** of our teacher evaluation system
  - Analyses that could leverage teacher evaluation data to provide **additional insights about teacher effectiveness**

Analyses to Inform Design/Implementation of the Evaluation System

- Relationship between value-added and observation scores
- Student survey analysis
- Achievement measure choices
- Implications of changing the weighting of components
- Non-differentiating observers
- Differences in observation scores between teacher groups

Example: Non-Differentiating Observers

- Identified a group of observers who tended to give similar scores to all of the teachers they observed on all indicators of the rubric.
- Established that these observers were not concentrated in particular districts, did not tend to have particular professional roles, and tended to be different than those whose observation scores were most misaligned with value-added scores.
- Result: Deployed TEAM coaches to investigate and support these observers and we are continuing to monitor this metric.

Analyses that provide additional insights about teacher effectiveness

- "SDP-style" human capital analysis
  - Teacher-student assignment
  - District and school retention of effective teachers

- Analysis of teaching practice
  - What can we learn from a database of indicator-level classroom observation data collected across the year for more than 56,000 teachers?

Tennessee Educator Acceleration Model (TEAM)*

Planning

- Instructional Plans
- Student Work
- Assessment

Environment

- Managing Student Behavior
- Expectations
- Environment
- Respectful Culture

Instruction

- Standards & Objectives
- Motivating Students
- Presenting Instructional Content
- Lesson Structure & Pacing
- Activities & Materials
- Questioning
- Academic Feedback
- Grouping Students
- Teacher Content Knowledge
- Teacher Knowledge of Students
- Thinking
- Problem Solving

Professionalism

- Professional Growth and Learning
- Use of Data
- School and Community Involvement
- Leadership

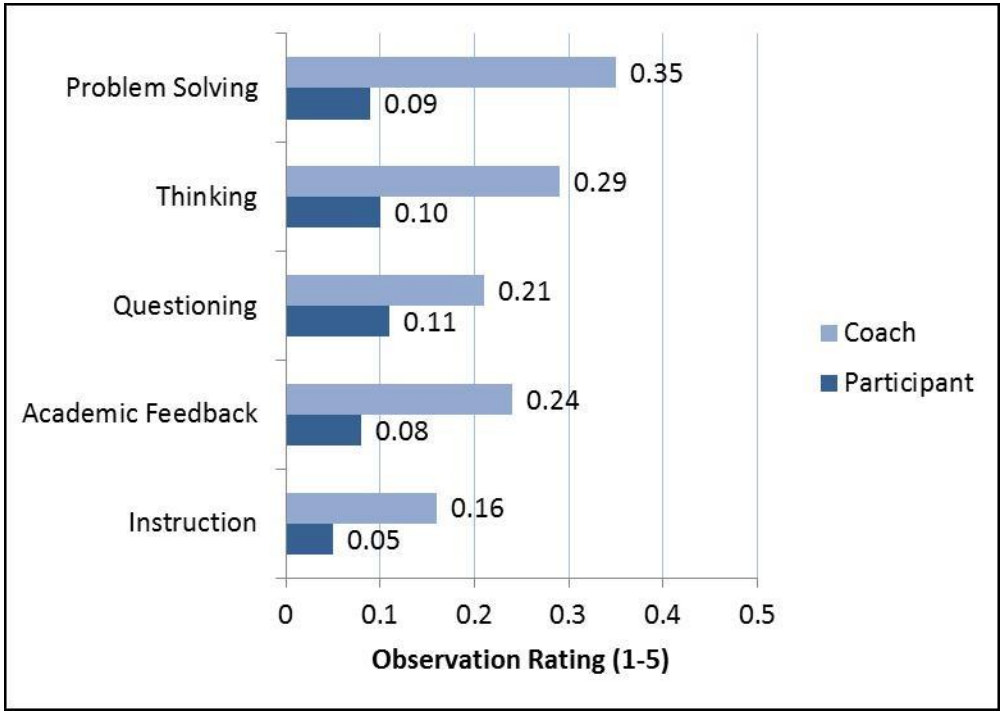
*modified from the TAP Teaching Performance Standards

Example: Using Observation Data to Evaluate Large-Scale PD

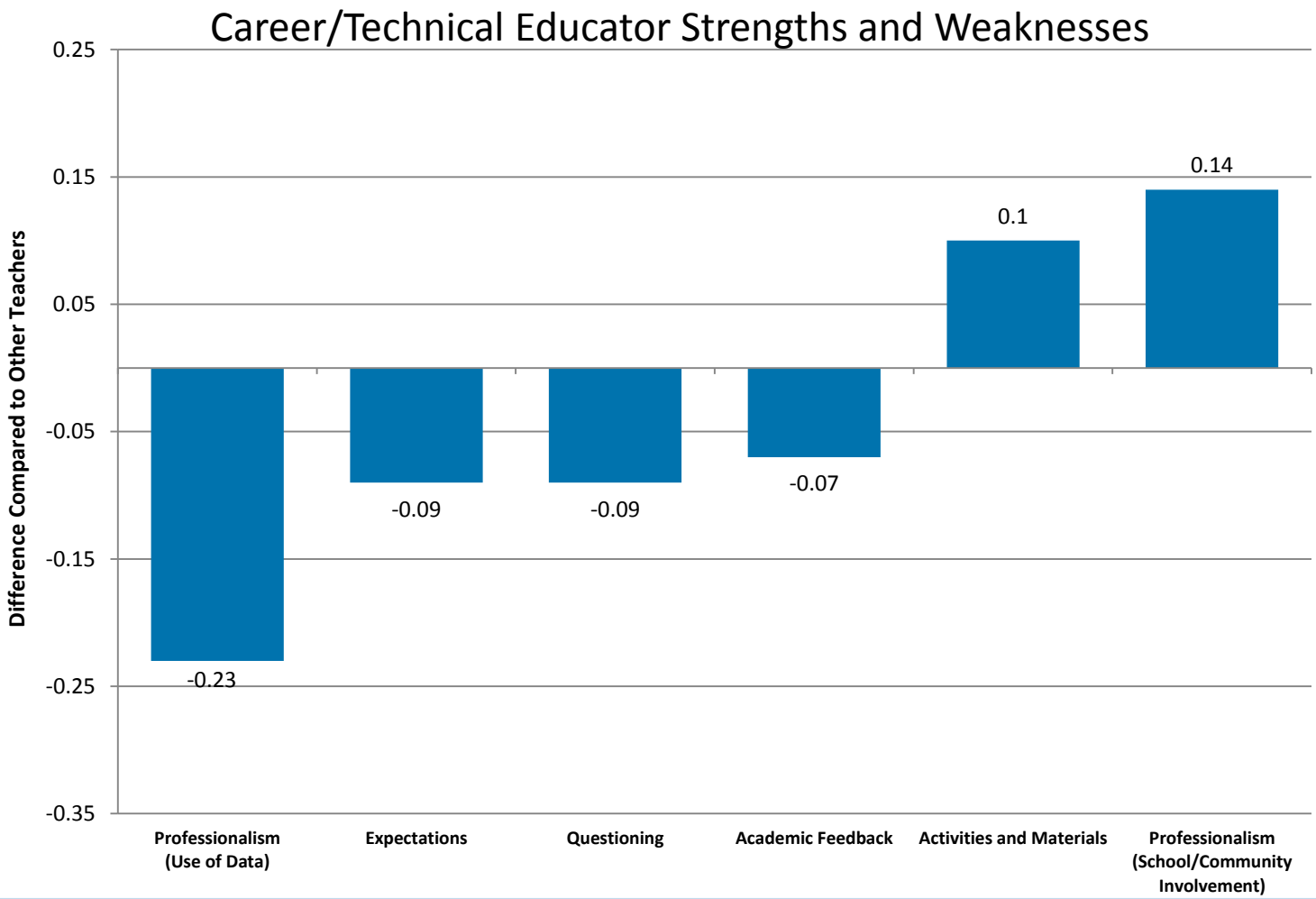
- In 2012, Tennessee trained 200 Common Core Coaches who went on to train around 6,000 math teachers
- Did these trainings affect targeted teaching practices (and student achievement)?
- Analysis controlled for previous year's ratings (and other teacher characteristics) – and included school-level fixed effects to compare teachers within the same school

We can track particular classroom practices – as well as increases in value-added scores

- Gains in targeted teaching practices were equivalent to about **half of the gains** made by teachers between their first and second year of teaching



Extending the picture: Comparing teacher groups



STRATEGIC DATA PROJECT

Teacher Evaluation Plenary

2014 Annual Convening

www.gse.harvard.edu/sdp/

Boston, MA | April 22, 2014

How will you know when reliability is low?

- In MET, training was <u>not enough</u>. (Reliability of 1 obs by 1 observer 30 to 37%!)
- Low reliability ⇒ Low correlation with VA, student surveys
  ~~⇐~~
- Only way to know: Audit study
  - Draw a random sample of 90-100 teachers.
  - Ask principals to observe 1 teacher from another school
  - Correlation between own-principal and other principal score=reliability.
  - Does <u>not</u> have to be same lesson.

How to fight the temptation of grade inflation?

- An inflation-proof minimum standard:
  - The Average Novice Teacher
- Self-adjusting for inflation
- Self-adjusting for pool of recruits
- Could vary by district
- Could vary over time

Appendix

3 Key Findings from TEvD Pilot with TN

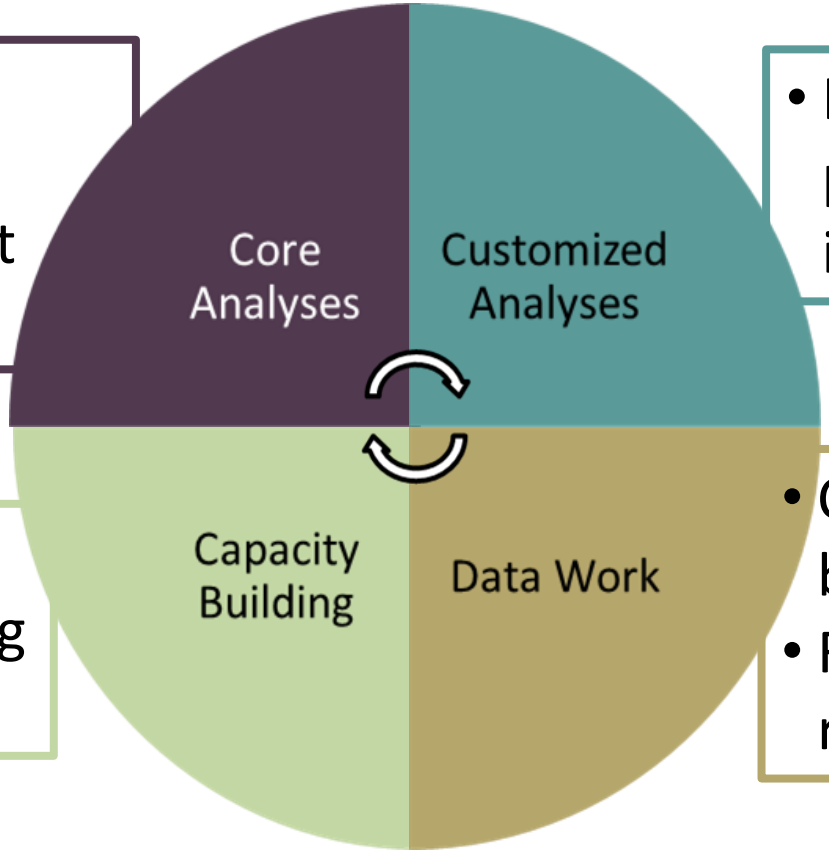
1. Teachers' average observation score differs slightly across observation frameworks
   - After controlling for teachers' individual growth, average observation scores indistinguishable across 2 frameworks
   - Average scores 0.24 points higher on one framework and 0.11 points lower on the other
2. Some districts have average observation scores significantly above or below the state average
   - Many differences persist after controlling for individual growth
   - Suggests that teachers with similar growth scores would receive different observation scores depending on their district
3. Student surveys are a statistically significant predictor of teachers' individual growth
   - Relationship holds after controlling for teachers' observation score

Center for Education Policy Research
HARVARD UNIVERSITY

STRATEGIC DATA PROJECT

The Diagnostic Process

- Highly relevant questions
- Describe current performance



- Key issues partner identifies

- On-site capacity-building trainings
- Dissemination

- Collect, clean, build analysis file
- Return files and related code