GETTING TEACHER EVALUATION RIGHT:
USING STUDENT LEARNING EVIDENCE WELL
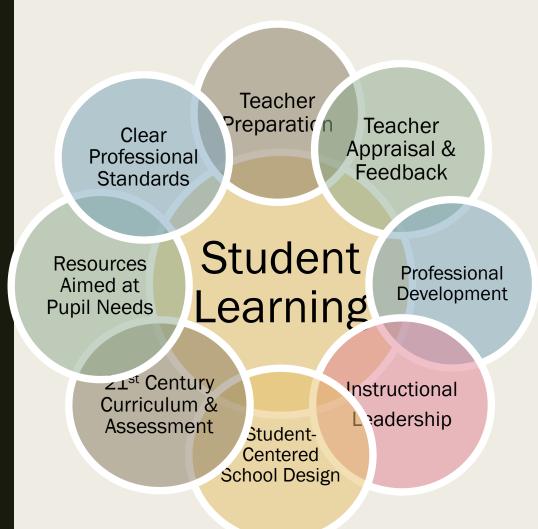
LINDA DARLING-HAMMOND

GETTING TEACHER EVALUATION RIGHT

WHAT REALLY MATTERS FOR EFFECTIVENESS AND IMPROVEMENT

Teacher Evaluation Is Part of a Broader System: Think Systemically

Student Learning

Teacher Preparation

Teacher Appraisal & Feedback

Clear Professional Standards

Professional Development

Resources Aimed at Pupil Needs

Instructional Leadership

21st Century Curriculum & Assessment

Student-Centered School Design

Criteria for an Effective Teacher Evaluation System

1. Based on professional teaching standards - from novice to expert

2. Integrated evidence of teacher practice, student learning, and professional contributions

3. Evaluators who are knowledgeable about instruction & evaluation

4. Useful feedback, connected to professional development opportunities

5. Supportive of collaboration

6. Expert teachers supporting a peer assessment & review process

7. Committee of administrators and teachers who oversee the system

Source: Darling-Hammond, Getting Teacher Evaluation Right, p. 153.

1. Use Standards Well

- Treat teacher evaluation in conjunction with the creation of a standards-based teacher development system

  -- beginners

  -- professional teachers

  -- accomplished practitioners

- Treat teacher evaluation in conjunction with leadership training, development, and evaluation, as well

Three Tiers of Teacher Assessment

Beginning Teacher Performance Assessment (e.g. PACT, edTPA)

Professional License Assessment

Advanced, e.g. National Board Certification

Tier 1
- Assessment of teaching performance for initial teacher licensing
- Coupled with basic skill or content knowledge assessment
- Allows a common high standard of practice for multiple pathways to teaching (teacher education preparation, internships, alternative routes, e.g., TFA)
- Evidence can be used for program approval or accreditation to leverage significant improvement in preparation
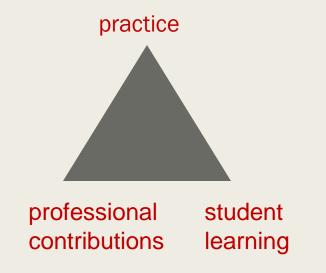
Tier 2
- Assessment of teaching performance for professional license
- Subsequent to induction, prior to tenure
- Systematic collection of evidence about teacher practice and student learning

Tier 3
- Advanced license or designation of accomplishment
- Assessment of high accomplishment as an experienced teacher
- Post-tenure
- State-specific advanced certification or National Board certification
- Career pathways as mentors, teacher leaders, or instructional specialists

2. Build Integrated Evidence of Teaching

-- Standards-based measures of teaching practice

-- Professional contributions

-- Student progress

Linked by goals and considered in context

Integrated evidence of:

practice

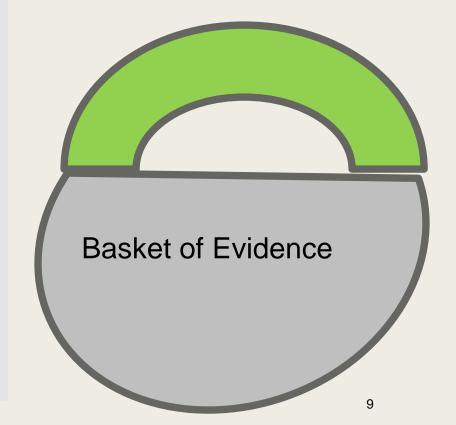professional contributions

student learning

Use Multiple Measures to Reflect *Instructional Practice*

- <u>Standards-based observation</u> by experts trained in evaluation and, ideally, the content area

- <u>Artifacts of teaching:</u> Curriculum plans, assignments, and student work samples

- <u>Evidence of practices that support student learning</u> in and outside of the classroom

  -- work with parents

  -- work with colleagues to support students

  -- engagement in professional learning

Use Multiple Measures to Assess Student *Learning*

- Authentic assessments of learning: essays, research projects, investigations; portfolios

- Local, state, national tests (pre - post measures; AP; IB; DRA)

- English proficiency gains

- Accomplishment of learner goals (IEP, other)

- Documentation of learning

- Goal-Setting Matched to Curriculum and Students Taught

Basket of Evidence

What About Value-Added Test Scores:

Value related to

  – *Measuring progress rather than status*

  – *Researching practices and programs*

Limitations related to

  – *Design of state tests*

    -- Measure only grade-level standards

    -- Measure a narrow part of the curriculum

  – *Other influences on student learning*

Leading Scholars on the Use of VAM for Teacher Evaulation

"VAM estimates of teacher effectiveness ... should not be used to make operational decisions because such estimates are far too unstable to be considered fair or reliable."

"Ranking teachers by their VAM scores can have unintended consequences that reduce quality."
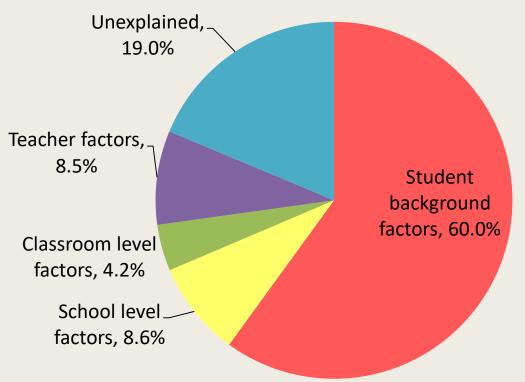
-- American Statistical Association (2014).

-- National Research Council, Board on Testing and Assessment. (2009).

What Have We Learned?

- VAM measures of "teacher effects" include many school and outside-of school conditions other than teachers

- VAM measures are not highly stable, and vary based on:
  - *Student backgrounds*
  - *School contexts*
  - *Different courses (as these intersect with teacher skills and / or student backgrounds)*
  - *Different years (as these represent experience, context, course, or student differences)*

- VAM measures exhibit biases against teachers of particular kinds of students. These are worse when tests measure narrow skills along a narrow scale.

Many Things Influence Achievement Gains

Variance in Achievement*
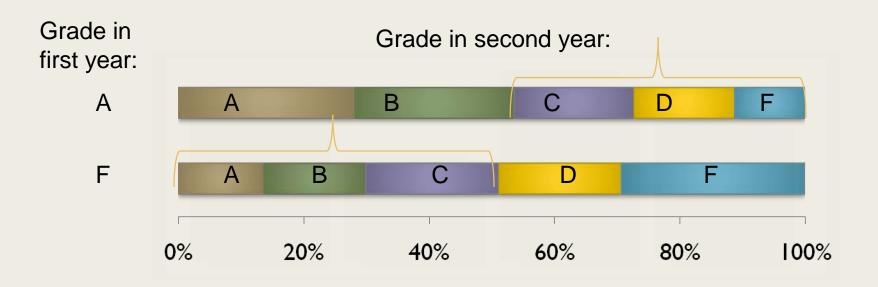**(10th grade Mathematics)**

- Model Controls for Prior Achievement
- 95% of school, classroom, and teacher effects were unobservable (i.e. could not be controlled with specific variables)

Unexplained, 19.0%

Teacher factors, 8.5%

Classroom level factors, 4.2%

School level factors, 8.6%

Student background factors, 60.0%

Source: Goldhaber, Brewer, & Anderson, *Education Economics, 7 (3),* 1999.
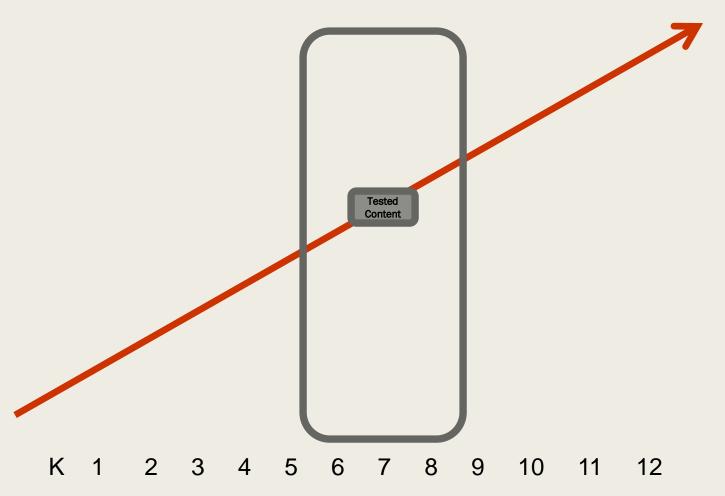
Value-Added Estimates Are Highly Unstable

- Many teachers indicated as effective or ineffective in one year are not for others

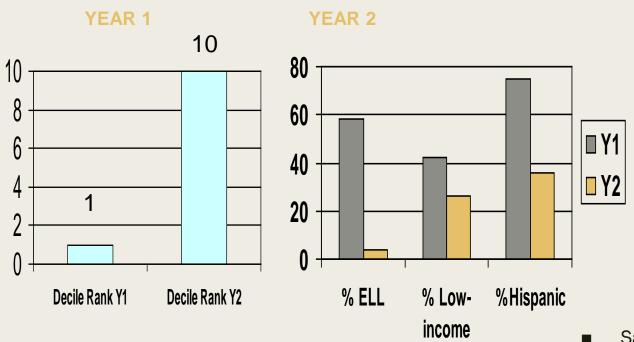- Consider classification of teachers into 5 categories (A-F) in two consecutive years.

Grade in first year:

Grade in second year:

| | A | B | C | D | F |

A

| | A | B | C | D | F |

F

0%   20%   40%   60%   80%   100%

Grades A-F correspond to quintiles 1-5. Source: Tim Sass (2008).

Grade-Level Tests Cannot Measure Growth for Many Students

Tested Content

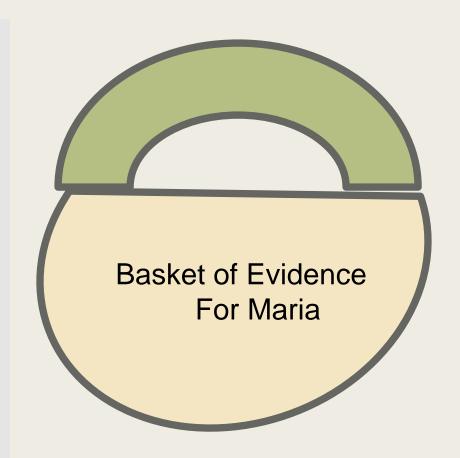K 1 2 3 4 5 6 7 8 9 10 11 12

A Teacher's Measured "Effectiveness" Can Vary Widely: 'Maria's' Case

YEAR 1

YEAR 2

- Same high school
- Same course (English I)
- Not a beginning teacher
- Model controls for:
  - *Prior achievement*
  - *Demographics*
  - *School fixed effects*

Potential Solution: Use Multiple Measures to Assess Student Learning

- English language proficiency gains

- Qualitative Reading Inventory gains

- 9th grade Autobiography paper / exhibition (video) illustrating progress across multiple drafts (linked to curriculum artifacts and teacher feedback)

- School fall and spring scored writing samples

Basket of Evidence For Maria

Use of Invalid Metrics Mismeasures Teachers: Carolyn's Case

"Ms. Abbott is a great teacher."

"Ms. Abbott is my son's teacher, and she has been fabulous. He has never learned more math than he has this year.

"Ms. Abbott is one of my daughter's favorite teachers. She helped my daughter develop a love of math, and the kids really enjoy her class."

The "Worst" 8th Grade Math Teacher in New York Based on VAM

"I always loved Ms. Abbott. She is and was an amazing teacher, and no teacher evaluation will ever change that."

"I am a sixth grader at Anderson, and I have been looking forward to having math class with Ms. Abbott since the fourth grade.... She is one of the best math teachers ever, in my opinion, and I would do anything to keep her here at Anderson."

Example of Multiple Measures: Carolyn's Basket

- Regents Exam outcomes

  with fall pre-test on Regents test material

- Pre- and post-test for unit on quadratic equations (Carolyn's focus this year) linked to curriculum and teaching evidence from that unit

- Scores & work samples for the 8th grade math/science inquiry project, with focus on students' applications of mathematics, as compared to the previous year

- Student survey of attitudes about math (pre- and post)

Basket of Evidence
For Carolyn

3. Ensure Knowledgeable Evaluators

Evaluation Expertise

Content Expertise

How Might Evidence Be Combined?
Massachusetts Example

4. Link Feedback and Learning

Link Professional Development Plans, Licensure Tiers and In-School Practice

Teacher Effectiveness is enhanced by:

- *Professional development tied to appraisal and feedback*

- *Goal-setting with support*

- *Reflection on teaching in relation to student learning outcomes*

- *Collaboration with colleagues*

What Kind of PD ?

A Paradigm Change

■ FROM:

■ Sit and get

■ Drive-by

■ One size fits all

■ Disconnected from teachers' classroom and students

■ TO:

■ Content-focused

■ Active

■ Collaborative

■ Using models & modeling

■ Coaching

■ Feedback and reflection

■ Sustained over time

OH WOW! PARADIGM SHIFT!
THAVES

5. Evaluation Can Support Collaboration

[Student learning evidence] can be used to promote collaboration and reflection on practice among educators. By analyzing work in teams and by setting schoolwide, subject-level, grade-level, or team-based goals, educators can work together to better understand and meet the needs of all students. The goal-setting and inquiry cycle encourages educators to share strategies and seek guidance and assistance from specialists who support student learning....

-- American Institutes of Research

New Mexico Approach: South Valley Academy

- 3-tiered state portfolio-based licensing system

- Within the school, PDP (Professional Development Plan) used for goal-setting; focused collaborative work; and assessment of results

- A practitioner action research evaluation model

Stanford Center for Opportunity Policy in Education

Case in Point

■ Andres' focus on improving vocabulary learning in chemistry

-- measurable goals [58% to 80% mastery]

-- a variety of strategies, developed with peers

-- tracking of outcomes on multiple measures

[14 artifacts collected in his portfolio with analysis of student learning]

Andres' Reflection

"These collaborations are perhaps the most powerful and useful part of the PDP process. Staff became aware of the many learning gaps that we are noticing in our students, and we began working collaboratively to share and find ways to address those learning gaps. Many times other staff members noticed things that were happening that I hadn't noticed before, but as soon as they said it, I was like, "Yeah, that's happening with my students, too." In many of the meetings, teachers together can start noticing trends, and then the project is even more meaningful....

Having a forum to share effective strategies for student learning opens the door for *every teacher at the school to help every other teacher get better*."

6. Peer Assistance and Review

- Mentoring for beginners and veterans with specific needs

- Tap National Board Certified Teachers and other accomplished teachers

- Content and evaluation expertise

- Time and support for coaching

- Network of mentors for developing skills

7. Oversight Panel

- Ensure good training

- Support mentoring

- Support decision making

- Create more uniformity and support

- Engage in continuous Improvement

Thoughts
and Questions?