



**CENTER FOR
TEACHING QUALITY**
TEACHERS TRANSFORMING TEACHING

PRESENTATION FOR Legislative Education Committee
DATE July 14, 2014

Teacher Evaluation and Teaching Quality:

The Evidence

Questions for Teacher Experts

- ◉ What lessons from top performing nations can be applied to the teaching profession in the United States?
- ◉ How was Colorado's teacher evaluation bill (SB 191) shaped - - and is being shaped -- by teachers?
- ◉ What are the essentials of a sound evaluation for ELL teachers?
- ◉ How should student test scores be used -- or not used -- in assessing teachers?

What Do We Know About Teaching Quality and School Improvement?

- ◉ Teachers matter the most of all in-school factors but they only explain about 12% of the differences in student achievement.*
- ◉ Long-term school improvement is built on dynamic collaboration among teachers, administrators, and parents (and students).**
- ◉ Students score higher on achievement tests when their teachers have opportunities to work with each other colleagues over a longer period of time.***

*Hill, H. et. al. (2014).

**Bryk (2002); Goddard (2007); Leena (2011); Adams (2014)

*** Jackson & Bruegmann (2009)

And Then There Are the Top-Performing Nations

2012 Program for International Student Assessment (PISA)

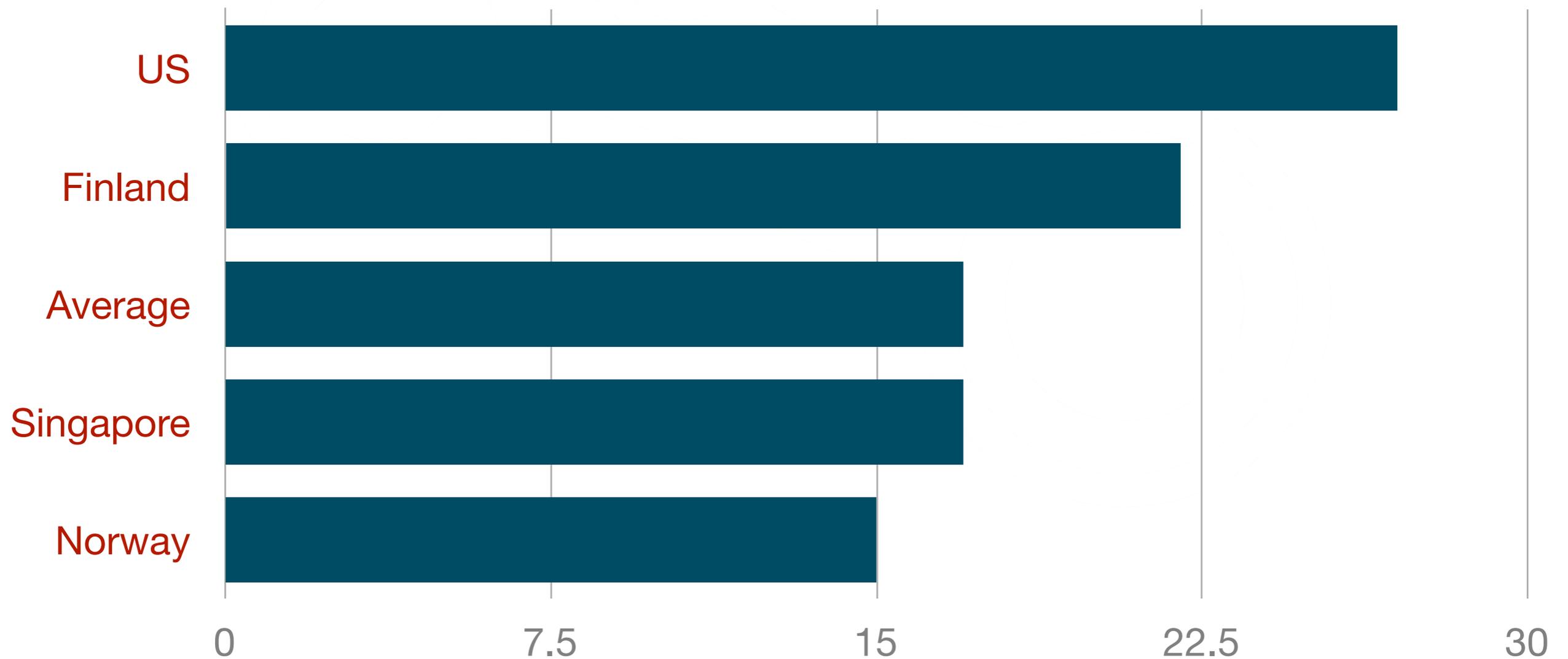
#	Reading-Overall	Rank	Mathematics	Rank	Science	Rank
1	China: Shanghai	570	China: Shanghai	613	China: Shanghai	580
2	Hong Kong	545	Singapore	573	Hong Kong	555
3	Singapore	542	Hong Kong	561	Singapore	551
4	Japan	538	Taiwan	560	Japan	547
5	South Korea	536	South Korea	554	Finland	545
6	Finland	524	China: Macau	538	Estonia	541
36	USA	481	USA	498	USA	497

What do top-performing nations do to invest in teachers?

- ◉ Fully paid university-based teacher education, year-long “student teaching,” with focus on pedagogy AND research skills
- ◉ Fewer standardized tests, and more emphasis on teachers developing and scoring own assessments
- ◉ Teachers teach about 9-17 hours of lessons per week, remaining time for lesson study and leadership
- ◉ Teacher evaluation and pay places premium on the spread of teaching expertise

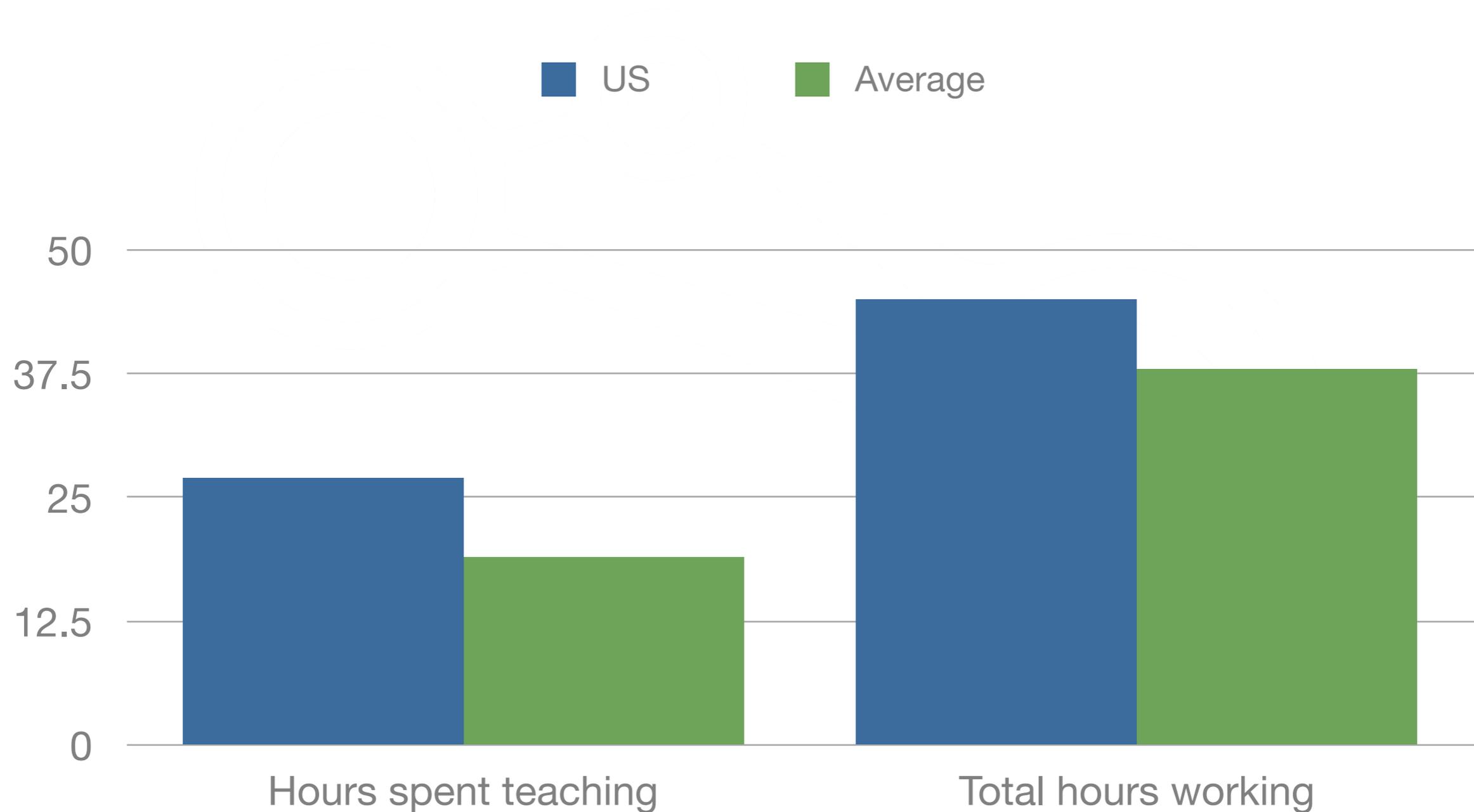
US Teachers Teach the Most

Hours spent teaching



Darling-Hammond, L. (2014). *What can TALIS tell us?* Presentation at the National Press Club. June 27, 2014.

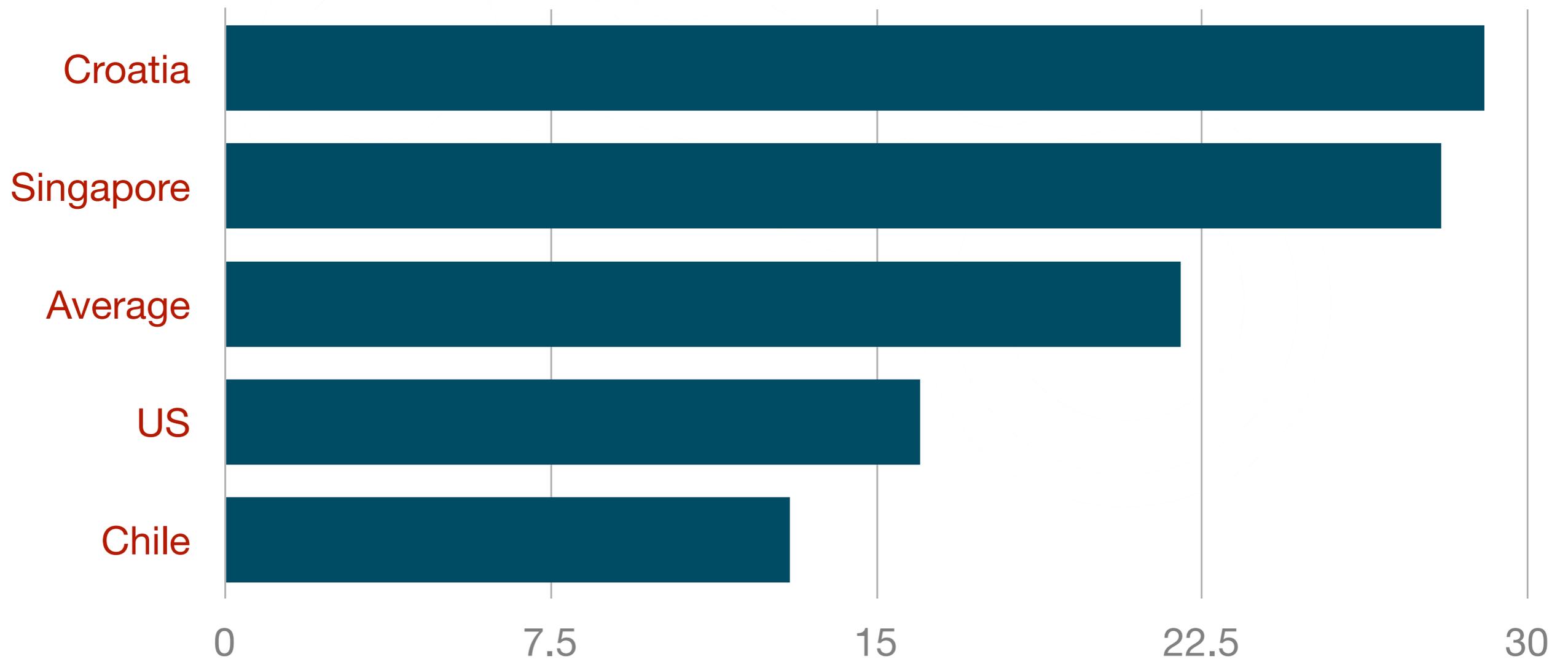
US Teachers Work the Most Hours



Darling-Hammond, L. (2014). *What can TALIS tell us?* Presentation at the National Press Club. June 27, 2014.

US Teachers have less planning time

■ Planning minutes per teaching hour



Darling-Hammond, L. (2014). *What can TALIS tell us?* Presentation at the National Press Club. June 27, 2014.

HOW DOES SINGAPORE ENSURE AND SUSTAIN A HIGH-QUALITY TEACHING FORCE?

Singapore: Teaching Quality *Begins* with Investments in Teachers

- ◉ Paid pre-service education
 - ◉ Includes monthly salary
- ◉ Competitive salaries
 - ◉ Equivalent to that of a beginning engineer or accountant (\$30,000 - \$50,000 per year) for a graduate
- ◉ Mid-career entrants
 - ◉ Salaries adjusted to previous working experiences

Singapore: Teaching Quality *Continues* with Investments in Teachers

- ◉ 100 hours of paid PD
- ◉ After 12 years all teachers entitled to one school term of PDL
- ◉ Pursue post-graduate studies
- ◉ Principals who have served 6 years can take a 2 month sabbatical

Singapore: Teacher Development

- ◉ Teaching Track
 - ◉ Achieve teaching excellence in the classroom
 - ◉ Extends through levels of senior, lead, and master teacher
- ◉ Leadership Track
 - ◉ Grooms teachers to take on leadership positions in schools
 - ◉ For example, principals, heads of departments, and at MOE departments
- ◉ Senior Specialist Track
 - ◉ Developed to be experts in curriculum and instructional design, educational psychology and guidance, educational testing and measurement, and research and statistics

Singapore: Purposes of Teacher Evaluation

- ◉ Teacher **evaluation is a vital step** in the drive to improve the effectiveness of teaching and learning
- ◉ Raising student performance leads to **substantial gains in student learning**
- ◉ Effective monitoring and evaluation of teaching is **central to the continuous improvement** of the effectiveness of teaching.

Singapore: Focus on Teacher Experience and Retention

- ◉ Attrition is less than 3% annually
- ◉ Renumeration packages include:
 - ◉ A retention bonus payable every 3-5 years
 - ◉ A lump sum withdrawal option at the end of a teacher's career
 - ◉ Provide reasonably competitive career earnings by the time educators reach the age of 40 or 50

So What Do We Know About Singapore's Teacher Evaluation and Development System?

- ◉ All teachers well-prepared before they begin to teach, including specific training to teach second language learners
- ◉ Most teachers are coached and assessed primarily by peers
- ◉ Teacher evaluation focused on teaching the whole child, partnering with parents, and spreading teaching expertise

So What Do We Know About Singapore's Teacher Evaluation and Development System?

- ◉ Standardized test scores are *not* used to judge teachers
- ◉ Master teachers, with principals, lead the evaluation process
- ◉ Teachers assemble evidence of accomplishments, using electronic portfolios
- ◉ Master teachers expected to “influence” policy and programs



[WHY JOIN?](#) [GET EMAIL UPDATES](#)

[LOG IN](#)

[Join the Community](#)



[Resources](#) [Collaboratory](#) [Blogs](#)

[ABOUT](#) [NEWS](#) [EVENTS](#) [DONATE](#)



WE TRANSFORM EDUCATION BY ELEVATING
TEACHERS' BOLD IDEAS AND EXPERT PRACTICES.



Questions for Teacher Experts

- ◉ What lessons from top performing nations can be applied to the teaching profession in the United States?
- ◉ How was Colorado's teacher evaluation bill (SB 191) shaped - - and is being shaped -- by teachers?
- ◉ What are the essentials of a sound evaluation for ELL teachers?
- ◉ How should student test scores be used -- or not used -- in assessing teachers?

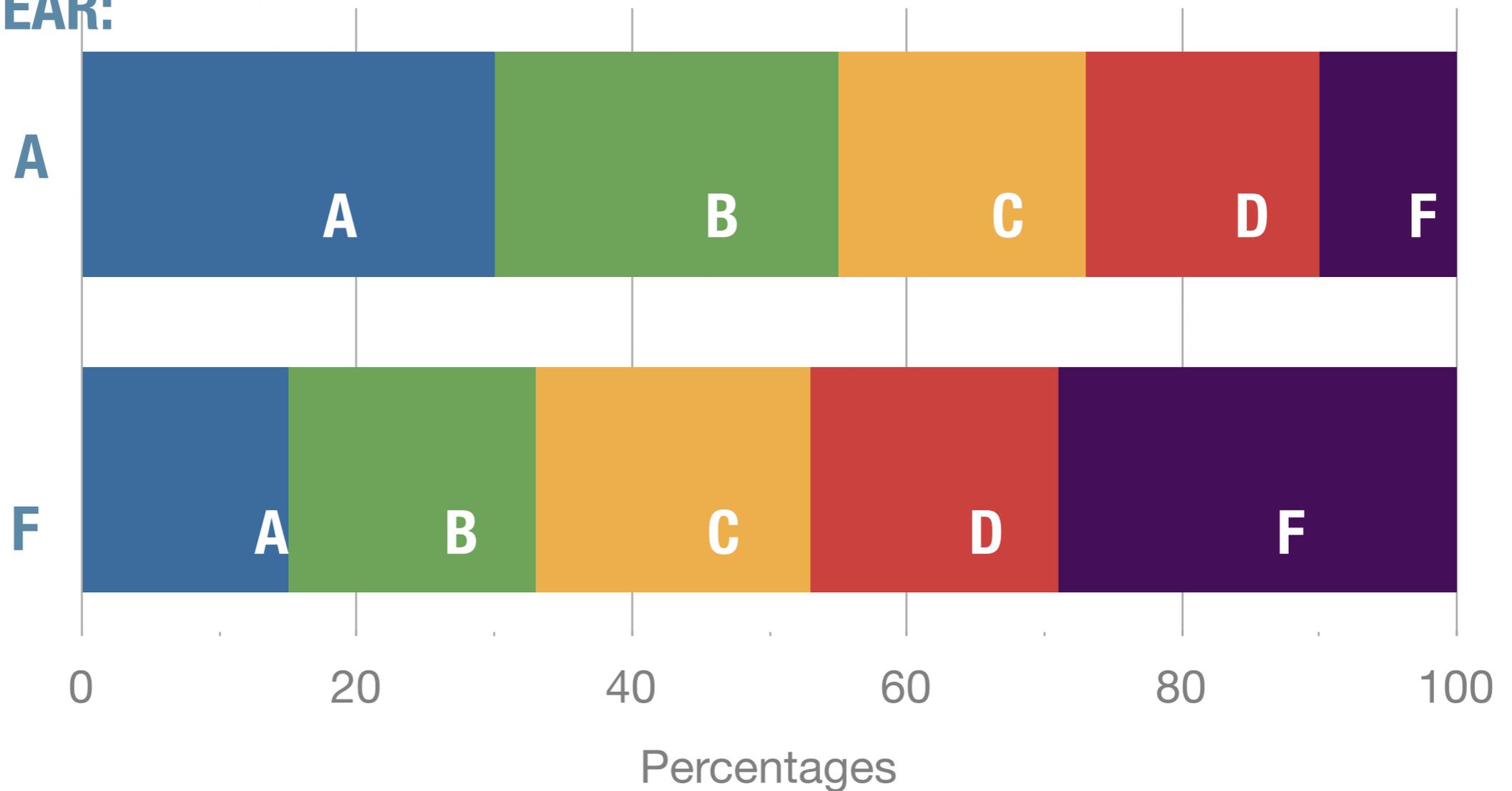
VAM estimates of teacher effectiveness ... should not be used to make operational decisions because such estimates are far too unstable to be considered fair or reliable.

- *National Research Council Board on Testing and Assessment 2009*

The Instability of “Value-Added” Measures of Teaching Effectiveness

GRADES
FIRST YEAR:

GRADES SECOND YEAR:



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

VAM ratings vary substantially, depending on:

- ◉ The statistical model used
- ◉ The measure of achievement used
- ◉ Class size, curriculum, instructional supports, and time spent with students
- ◉ Tutoring, team teaching, and parent supports
- ◉ Student characteristics and attendance

Use of VAM or “Houston, we have a problem.”

- Teachers teaching larger numbers of special education students in mainstreamed classrooms are also found to have lower “value-added” scores in those years.
- Teachers teaching in grades in which English Language Learners (ELLs) are transitioned in to mainstreamed classrooms are the least likely to show “added value”
- Teachers teaching gifted students have small gains because their students are near the top.

Amrein-Beardsley & Collins, [Educational Policy Analysis Archives](#) (2012)

Get involved...

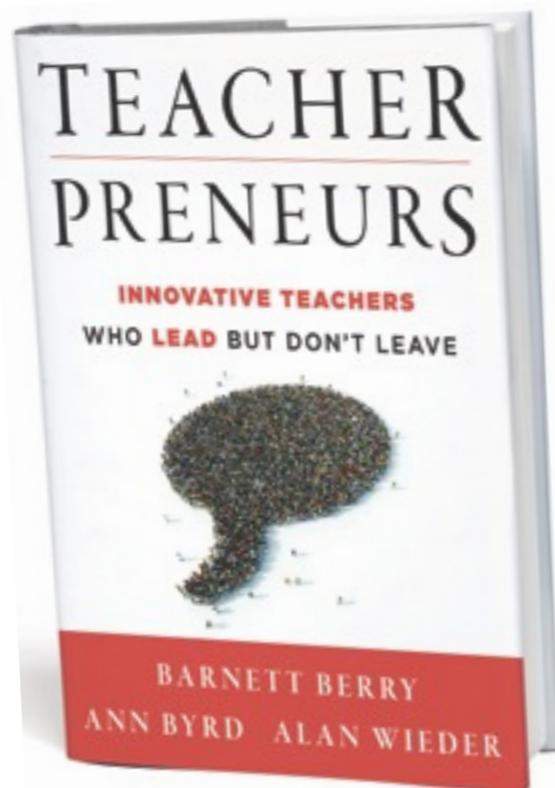
Connect with teacher
leaders in the CTQ

Collaboratory:

www.teachingquality.org

On Twitter:

- ◉ @BarnettCTQ
- ◉ @teachingquality
- ◉ #teacherpreneurs





Getting Smart About Teacher Evaluation

Barnett Berry

CEO and Partner, Center for Teaching Quality

bberry@teachingquality.org

@BarnettCTQ

Since 2009, thirty-two states and the District of Columbia have attempted to reinvent their teaching evaluation systems, fueled by evidence showing that “teachers are the most single important in-school factor for student learning” and traditional methods* of measuring quality instruction had “little to no bearing on actual student learning.”¹ School reformers were right in that old systems of teacher evaluation were “perfunctory.”² Most teacher evaluation systems rarely identified low performing teachers.

In an attempt to respond to the problem, the Obama Administration’s Race to the Top (R2T) legislation required states, in order to qualify for federal funds, to increase the number of times teachers are observed by principals and use standardized tests of student performance to assess whether or not they are effective. In New Mexico, like in many other jurisdictions, 50 percent of an individual’s rating must be calculated with a value added model (VAM), which “measure(s) the teacher contribution to student achievement numerically.” However, as reported by Education Week, the R2T “winners are struggling to implement big promises to overhaul the teaching profession”—and particularly in “implement(ing) new evaluation systems linked to student growth on test scores.”³ Even early R2T adopters like Florida and Tennessee, which placed a rigid emphasis on VAM in their evaluation calculations, well over 97 percent of teachers were “deemed effective or better.”⁴

Evidence suggests many reasons why states have struggled mightily in implementing more rigorous, valid, and reliable systems of teacher evaluation. The three most important may be:

- Reformers ignored lessons from a long history of failed efforts to improve teacher evaluation, including 1984 research (<http://www.rand.org/content/dam/rand/pubs/reports/2006/R3139.pdf>), documenting the importance of preparing principals and making sure they have time to conduct more valid classroom observations.
- Researchers have shown VAM ratings are highly unstable. Most recently the American Statistical Association (http://www.amstat.org/policy/pdfs/ASA_VAM_Statement.pdf) concluded that “most VAM studies find that teachers account for about 1% to 14% of the variability in test scores, and that the majority of opportunities for quality improvement are found in the system-level conditions.

* Others researchers have found that teaching certification and experience matters for student learning—under the right conditions.

- Top-performing nations create teaching evaluation systems that are embedded in a larger teacher development system characterized by:
 - Preparing all teachers *deeply* to teach diverse students (including language learners) and analyze evidence of practice;
 - Providing substantial amounts of time for teachers to learn from each other—and including peers (alongside or instead of principals) to serve as the driving forces behind high-quality teaching feedback systems;
 - Offering performance pay not simply for teachers with higher student test scores, but to those who spread their expertise to one another.

A recent study found that current teaching evaluation systems in the United States using summative student performance data (e.g., value-added, growth, and achievement measures) “does not meet the analytic needs of teachers who must improve and adapt their practice on a daily, weekly, and monthly basis.”⁵ In top-performing nations like Singapore—where teachers only work directly with students about 17 hours a week—there is considerable organizational space dedicated to assembling sound evidence, analyzing it, and applying professional judgment to determine who is effective and why—and how to help them get better.

Practitioner Perspectives on What Works for Evaluation Implementation

CTQ Collaboratory teacher leaders Wendi Pillars (North Carolina) and Jessica Keigan (Colorado) have had extensive experience in the implementation of their states’ teaching evaluation systems. Jessica, as a CTQ teacherpreneur (a hybrid role combining half-time teaching with release time for leadership work beyond her building), has had opportunities to lead the implementation of Colorado’s SB 191, which is now beginning to look much different than the design of the original legislation. Their advice is far-reaching, representing a deep understanding of students and communities. They recommend that policymakers, and those who are charged with implementing evaluation systems for teachers and principals, hold the following in mind as they design state and local processes to provide feedback that strengthens teaching practice:

- The vast majority of teachers embrace accountability, yet they also seek to participate in collaborative discussion/goal setting that allows them to define what the system should look like.
- Teaching is complex, and it cannot not be quantified easily with a checklist on an observation rubric. Much of what counts for effective instruction is not always visible in classroom observations (requiring a paradigm shift from the very formal process of R2T to a more fluid and organic relationship developed between teachers and their evaluators).
- One-size fits all protocols and feedback systems cannot accommodate the different needs of teachers in helping them succeed and grow.
- Many schools are using co-teaching systems (including pull-out or push-in instruction) to serve diverse students, making it challenging to ascribe student learning results to any one teacher.
- Some schools (and teachers) have more special needs students than others. English language learner (ELL) students require very different services, and whether or not they have access to them can affect a teacher’s rating.

- ③ Many standardized tests and VAMs do not accurately measure the growth of ELL students, primarily because of their very wide ability and proficiency levels, as well as assessment tools that are not always designed to measure their vast differences in their knowledge and skill—in content as well as in English language proficiency.
- ③ Teachers are not only responsible for teaching content but also for providing emotional support to students. While student engagement is linked with persistence in school and higher graduation rates, this is especially important for students from other countries as they adjust to U.S. culture. (The Singapore evaluation system places a premium on teachers “winning the hearts and minds” of their students, not just raising test scores.)

ENDNOTES

- 1 Bellwether Partners (2014). *Genuine Progress, Greater Challenges: A Decade of Teacher Effectiveness Reforms.* Retrieved on July 9, 2014 at http://bellwethereducation.org/sites/default/files/JOYCE_Teacher%20Effectiveness_web.pdf
- 2 Rucinski, M. and Diersing C. (2014). *America's Teacher Evaluation System Revolution.* Retrieved on July 9, 2014 at <http://harvardpolitics.com/united-states/americas-teacher-evaluation-system-revolution/>
- 3 McNeil, M. (2013). *Race to Top Reports Detail Winners' Progress, Challenges: Teacher-evaluation puzzle proving difficult to crack.* Education Week. Retrieved on July 9, 2014 at <http://www.edweek.org/ew/articles/2014/03/19/26rtt.h33.html?r=1502512170>
- 4 Sawchuk, S. (2013). *Teachers' Ratings Still High Despite New Measures: Changes to evaluation systems yield only subtle differences.* Education Week. Retrieved on July 9, 2014 at http://www.edweek.org/ew/articles/2013/02/06/20evaluate_ep.h32.html
- 5 Lai, M. K., & McNaughton, S. (2013). *Analysis and discussion of classroom and achievement data to raise student achievement.* In K. Schildkamp, M. K. Lai & L. Earl (Eds.), *Data-Based Decision Making in Education: Challenges and Opportunities* (pp. 23-47). Dordrecht: Springer.

