

Trends & Early Lessons on Evaluation and Effectiveness

Prepared for 11/17/11 Legislative Education Study Committee Working Session,
New Mexico

Kathy Christie, Education Commission of the States



What we know.

What we don't.

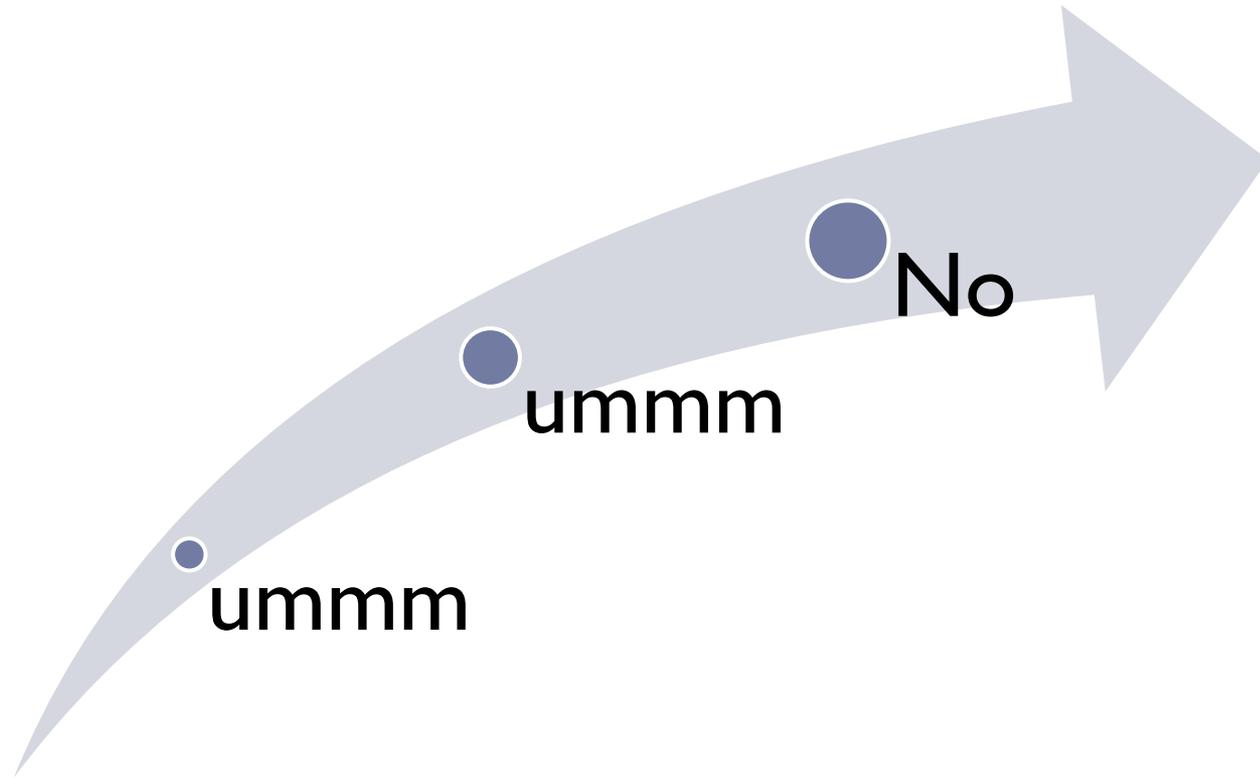
What we know

The Widget Effect
2009 study
New Teacher Project

- 99% positive ratings when “thumbs up,” “thumbs down”
- 94% positive ratings with more rating options
- Less than 1% rated unsatisfactory

2010 NBER report

“Do Principals Fire the Worst Teachers?”



Lessons from the Experts

What makes a good system?

Passing Muster: Evaluating Teacher Evaluation Systems

(Brown Center on Education Policy at Brookings, 2011)

- **Differentiate teachers**

(if 95% receive same, system not useful)

**Better
systems**

Passing Muster: Evaluating Teacher Evaluation Systems

(Brown Center on Education Policy at Brookings, 2011)

- Should not be driven by observable characteristics unlikely to predict effectiveness (i.e., experience not really predictive)

**Better
Systems**

- Should be predictive of future evaluations

(Identify those who will continue to perform well)

**Better
Systems**

Passing Muster : Evaluating Teacher Evaluation Systems

(Brown Center on Education Policy at Brookings, 2011)

- Use multiple measures

**Better
systems**



- Value-added on state-level tests
- Formal/informal observation and/or measures of student progress

**Multiple
Measures**

Passing Muster : Evaluating Teacher Evaluation Systems

(Brown Center on Education Policy at Brookings, 2011)

- Should be applicable to all teachers

(might require some alternative measures – other tests, other student work, assessments by leaders, peers, parents and/or students)

**Better
Systems**

RAND: *Using Student Performance to Evaluate Teachers* (2011)

- *Multiple measures over multiple years*

Assessments that support valid inferences about teachers

Recommendations or lessons

Very similar to those from *Passing Muster*

For non-testing grades/subjects, supplemental assmts.

System monitoring – for unintended consequences

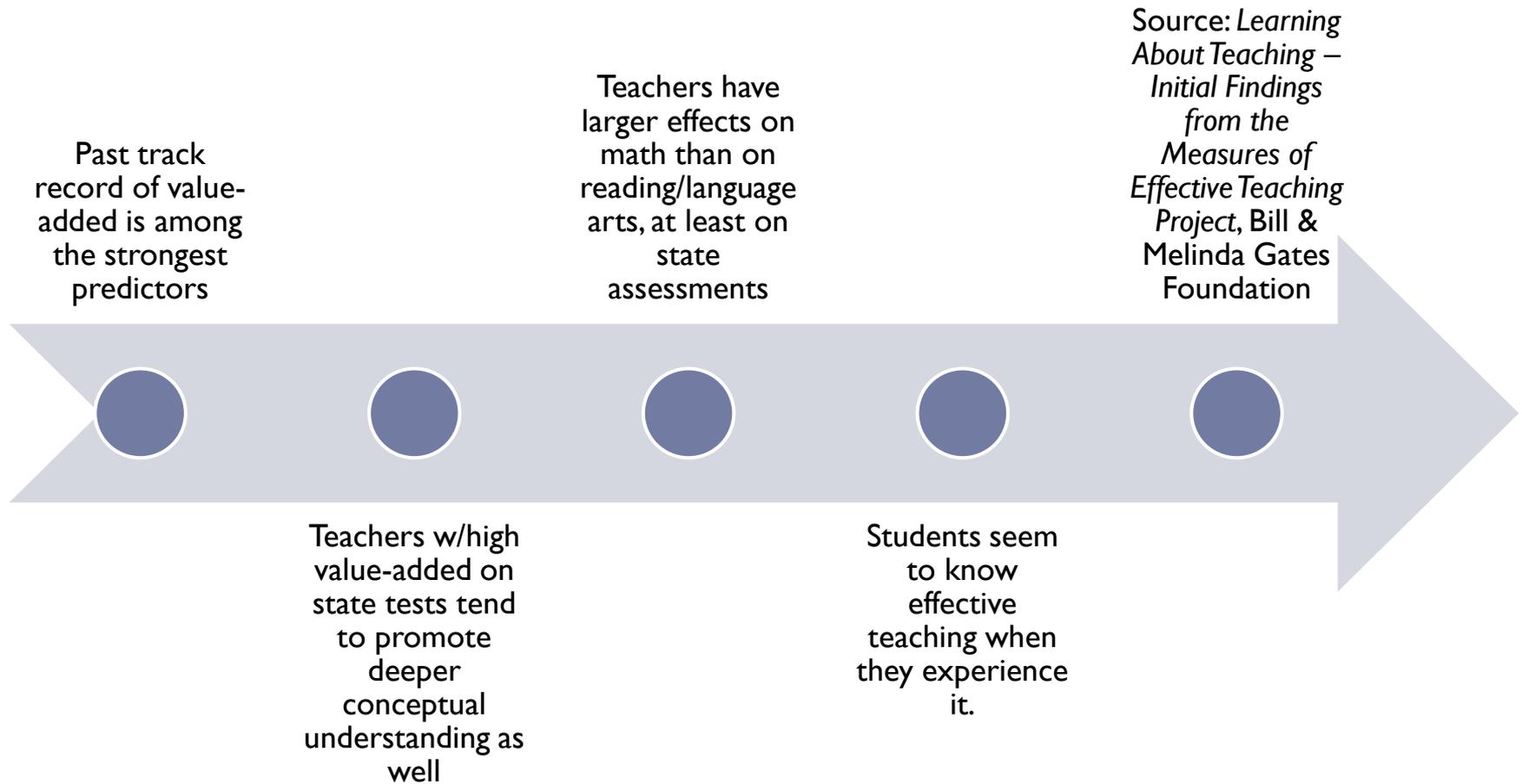
Do value-added scores reflect good teaching?

Value-added scores correlate with content knowledge and the quality of instruction.

Caveat: Use in combination with high quality, observational systems or as a TRIGGER for use of such a system

AERA-published research: *A Validity Argument Approach to Evaluating Teacher Value-Added Scores*
(Hill, Kapitula, Umland, 2010).

Do value-added scores reflect good teaching?



Does evaluation affect performance?>

Quality classroom-observation-based evaluation and performance measures CAN improve mid-career performance both during the period of evaluation and in subsequent years

Effect sizes represent a substantial gain in welfare given the program's costs

Dimensions of “well-structured” or “quality” “remain elusive” – best to vary the components and measure resulting differences in effectiveness

Source: NBER: *Effective of Eval. On Performance: Evidence from Longitudinal Student Achievement Data of Mid-Career Teachers*, 2011

Is evaluation useful to employers?

Objective job performance data can be useful to managers

The privacy of this information affects its usefulness as a policy tool

Private information on teacher performance contributes to the continued employment of poor performing teachers.

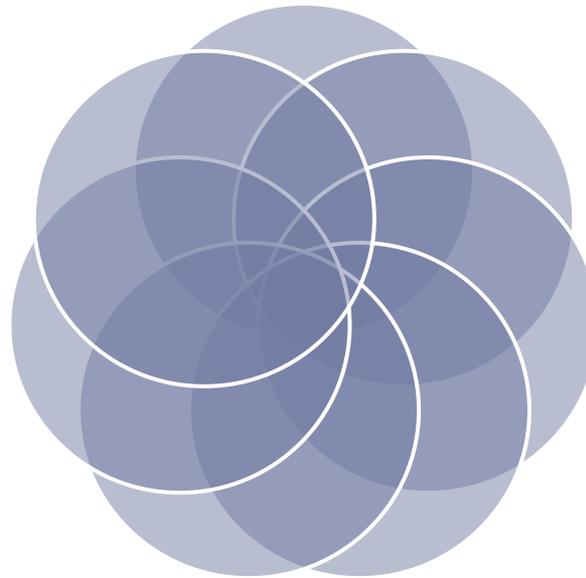
Source: *Information and Employee Evaluation: Evidence From a Randomized Intervention in Public Schools* (2010) NBER Working Paper

Lessons from DC's IMPACT System

What's
weighted?

School's overall
impact on
achievement = 5%

Commitment to the
school community =
5%



Value-added
achievement = 50%
(10% for non-tested
grades)

Classroom
Performance
(observational) = 40%
(80% for non-tested
grades)

KEY Lessons from DC's IMPACT System

Sources: Aspen Institute, Education Sector, RTTT Technical Assistance Network

Avoid Making standards too strict (100% students can articulate understanding....; 75% engaged)

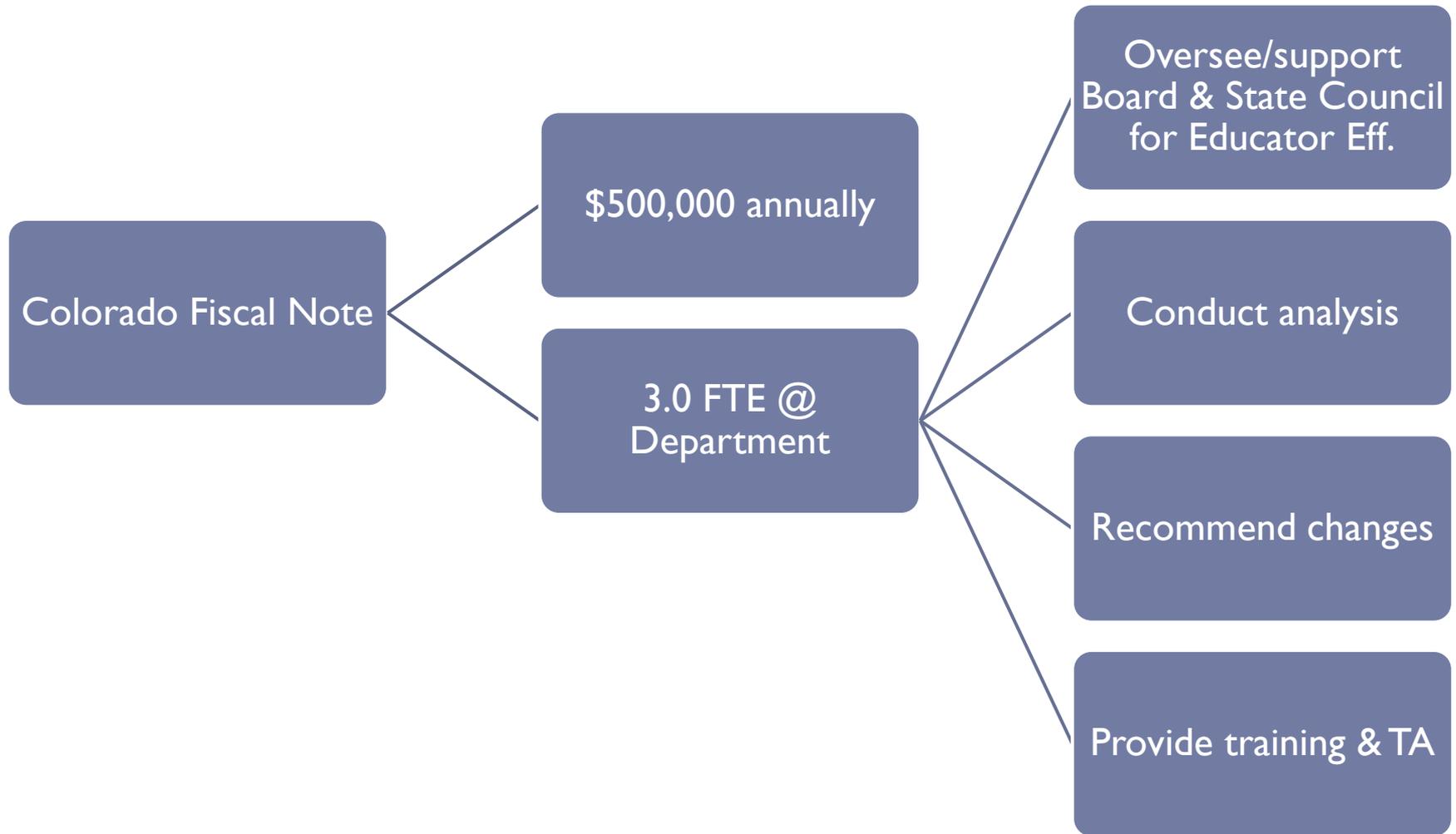
Define standards clearly and provide training BEFORE

Ensure everyone gets adequate training and fully understands

Analyze data from all to ensure findings are consistent

Look for high correlation between test scores and observations

Cost Estimates?



Cost Estimates?

Pennsylvania

- First Round pilot (20%), \$800,000 from Gates
- Next phase est. \$1.3 M

California estimates for proposal

- \$25-\$30 M for evaluations
- \$42-\$84 M in support

Lessons Learned from Temple Grandin

Animals in Translation (Grandin, 2005)

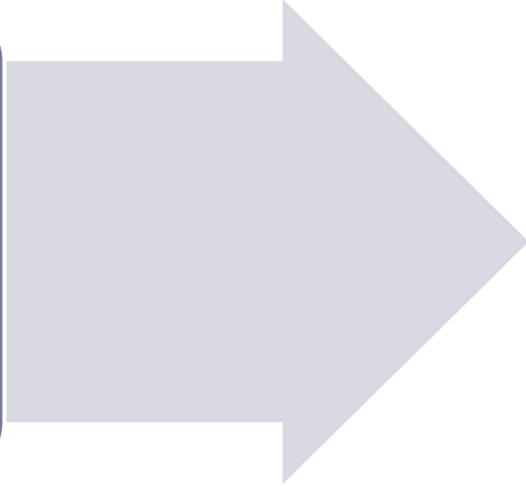
Animals Make Us Human (Grandin, 2009)

Grandin's "Critical Control Points"

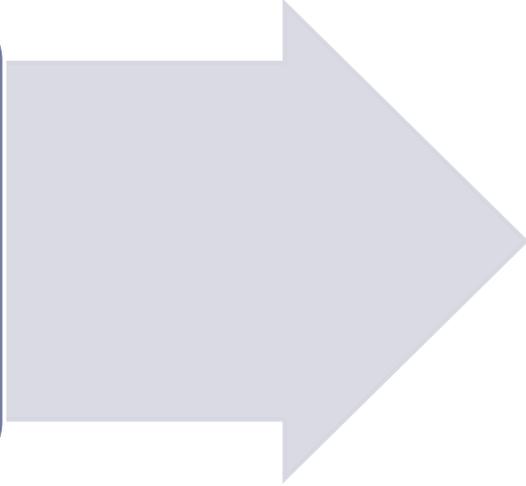
Each standard should reflect:

“a single measurable element that covers a multitude of sins”
(Grandin, 2005)

A 100-
item
checklist

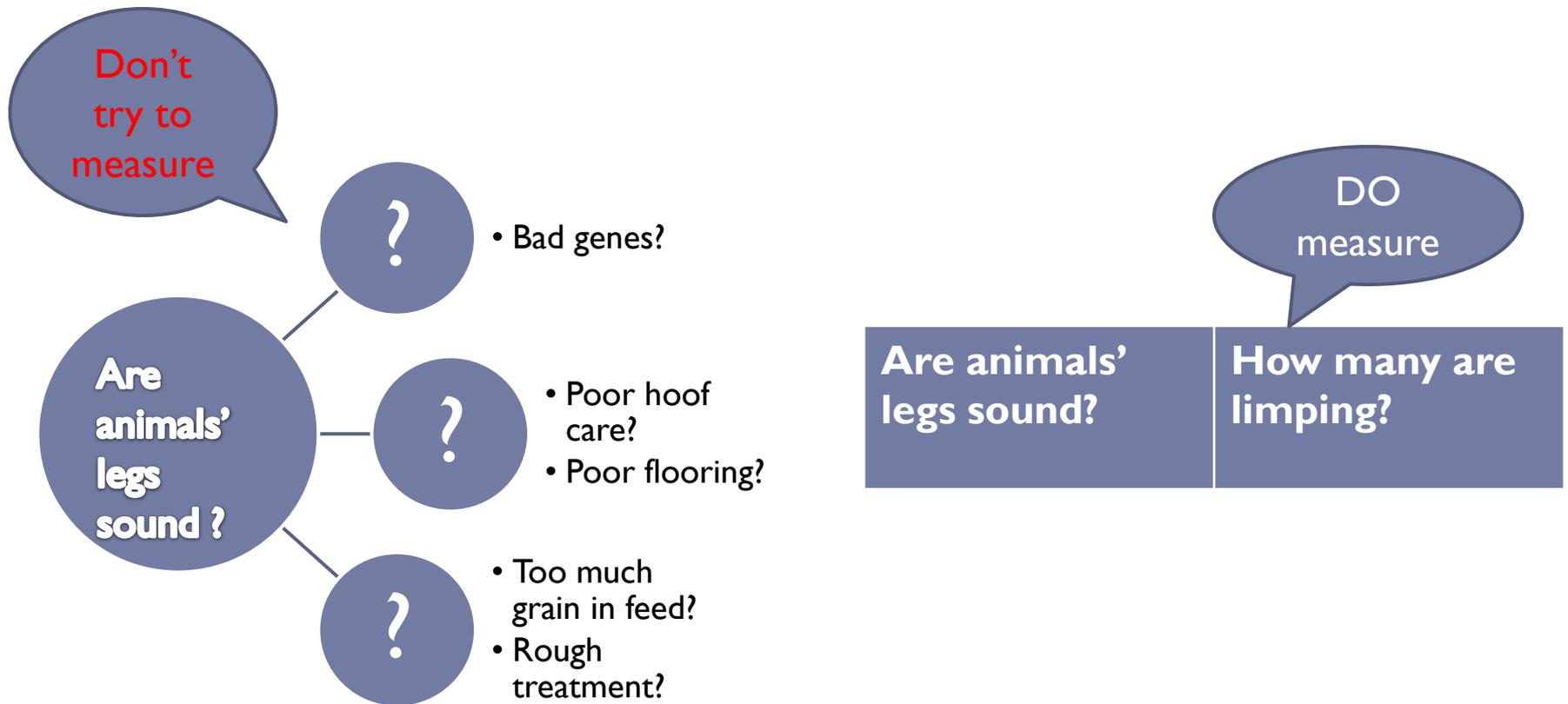


Doesn't
work as
well as a
10-item
checklist



Instructional observational audit tools should reflect “critical control points”

Is each standard “a single measurable element that covers a multitude of sins?” (Grandin)



Are states acting on these “lessons?”

Recommendations vs trends

Do results differentiate?

Maryland: 3 levels recommended

- Highly effective/effective/ineffective

Ohio: 4 levels

- Accomplished/Professional/Developing/Ineffective

Virginia: 4 levels

- Exemplary/Professional/Needs Improvement/Unacceptable

Illinois: 4 levels

- Excellent/Professional/Needs Improvement/Unsatisfactory

Oklahoma: 5 levels

- Superior/highly effective/effective/needs improvement/ineffective

Do results differentiate? (NCTQ, 2011)

4+ performance levels

- 11 states

Multiple Measures (Quality Control)

(NCTQ, 2011)

State bd/dept. develop state assessment tool; districts use

- DE, DC, LA

Districts expected to use state tool, but can design own

- MI, RI, TN

District responsible to design (but state model available)

- CO, IL, IN, OK

District designs but must submit for approval

- FL, ID, MD

District responsible for design

- AZ, MN, NV, NY, OH

Multiple Measures

Arkansas 6-17-2803

Defines “artifacts” –

- alignment of lesson plans with standards
- samples of students work/portfolios, writing and projects
- formative assessments
- Career and technical assessments

Multiple Measures (NCTQ, 2011)

Student
achievement
data used for
50% or more

**12
states**

States
accounting for
non-tested
grades

9 states

Other common trends

Standardizing timelines and requirements for conducting reviews

Training evaluators

Amending appeals processes and timelines

Recognizing and retaining highly effective

Professional development support for all to improve

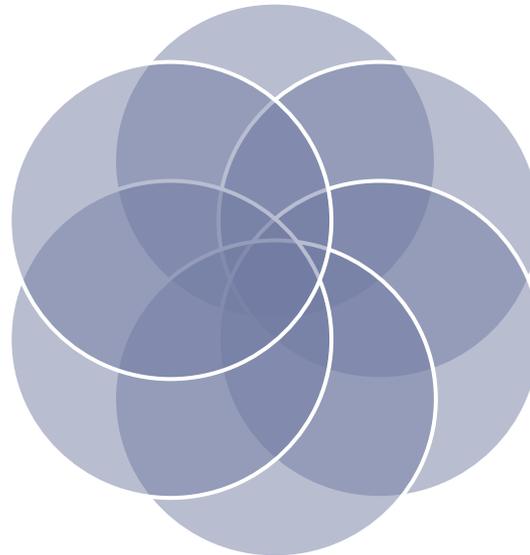
Support for new and/or low-performing

Linkages to continuing contract/tenure decisions

Other critical elements (Data Quality Campaign)

Ability to link teachers to student data

Can match records across teachers and students by course and/or subject :YES



Statewide Teacher Identifier with a Teacher-Student Match: YES

Each teacher has a unique identifier: YES

Procedures to ensure that a teacher does not have two different IDs: YES

Other critical elements

Multiple years

Timeliness

- (Back-mapped to contract notification dates?) 

Timely processes for improvement & appeals

Review of the reviewers

The devil's in the details

Few are talking about this

Colorado's attention to details

These elements:
unique and
important

- ▶ **Protects principals**
 - ▶ Don't have to accept staff they don't want
- ▶ **Evaluates principals**
 - ▶ On retention of highly effective teachers
- ▶ **Hiring pool**
 - ▶ For those rated satisfactory or effective but that principals don't choose to new
 - ▶ Have 1st crack at interviewing for other vacancies
 - ▶ Are paid until 12 months or 2 hiring cycles, then unpaid leave

More on inclusion of principals

achievement & growth

#/% of personnel rated effective or highly effective

Colorado principals rated on:

#/% rated ineffective but improving

Systems that include principals

At least 50% on indicators of students learning over 3 years

Measures related to appropriate use of evaluation criteria

Florida principals
rated on:

Recruitment/retention of highly effective teachers

Improvement in % of these

Other states that include principals

- ▶ Arizona (30-50% student progress)
- ▶ Idaho
- ▶ Illinois
- ▶ Louisiana
- ▶ Maryland
- ▶ Maine (appears to be optional)
- ▶ Minnesota
- ▶ New York
- ▶ Ohio
- ▶ Tennessee
 - ▶ Growth plus other such as graduation, ACT scores, attendance
 - ▶ Of note: contract with principals cannot exceed contract of superintendent)

Anticipate the need for:

Quality-assurance in training for evaluators

Political cover for evaluators
(use of outside evaluators?)

Evaluators to easily & regularly discuss issues
they're facing – online and in person

Continuous improvement of the system

Anticipate the need for:

Tweaking or reconsideration of rating levels

Next generation questions: What if “satisfactory” isn’t good enough for a particular principal or school? Will you have a “hiring pool” to cover principals who need or want to set a higher bar?

Immediate feedback systems

Hand-held technology that will help automate observational interactions and automate feedback as much as possible

Anticipate need for:

Minimizing time constraints

- Triage approach where see issues – NOT required to meet w/all?
- Illinois: all evaluated IX every 2 years but every year for “needs improvement” or “unsatisfactory”
- Reconsidering frequency of evaluations

Will your system get the results that old systems got?

Illinois legislation actually cites a recent Illinois study

41,174
evaluations
over 5 years

- 92.6%
superior or
excellent

7% satisfactory
.4% unsatisfactory

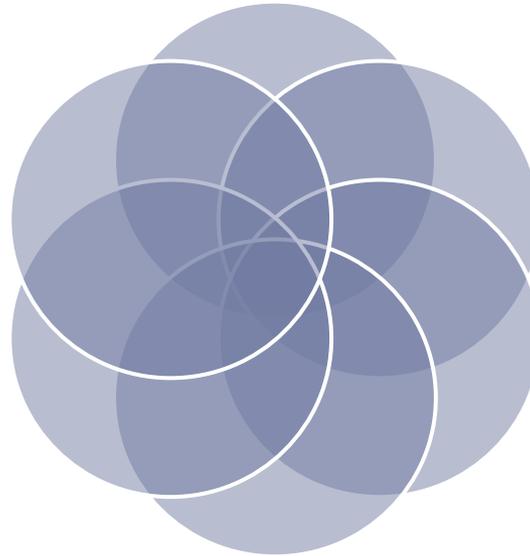
Remember the Widget Effect? (New Teacher Project)

99% positive rating
when thumbs
up/thumbs down

94% w/more
rating options

Will your
principals make
the difficult
decisions?

Less than 1%
rated satisfactory.



Performance Pay

What's the research say?

What's the new generation look like?

What's the research say?

The downside



Insufficient student test data to determine the real impact on student achievement



Some positive results – but no evidence that fostered student achievement gains



No evidence that raised student test scores or positively impacted teacher retention



What's the research say?

The upside



Some promising early data from
Teacher Incentive Fund sites



Some promising results from 2nd
and 3rd years of Tennessee's
Project on Incentives in Teaching
(POINT) – but not sustained, and
weak evidence standards for study



Denver's ProComp (External Evaluation Report 2006-2010)



At school level, meeting student growth objectives (SGOs) did not appear to be related to student growth



On average, though, meeting SGOs did reward effective teachers





But...

Here come the next generation models...

Common features, RTTT proposals

Most informed by
annual evaluations

Differentiated pay
such as signing
bonuses, additional
pay for high-need
schools or
subjects,
leadership roles,
etc.

Range from
defined bonuses
to complete
overhauls

Florida

After
7/1/14,
new
hires

- Performance salary schedule
- Others can opt in

Differentiated pay for
teachers &
administrators

- Additional responsibilities
- School demographics
- Critical shortages
- Level of job difficulty

Florida

Salary adjustments for “highly effective”

- must be greater than highest of others in same classification

Adjustments for “effective”

- Equal to 50%-75% of “highly effective” in same classification

No adjustments for employees NOT rated as “effective” or “highly effective”

Oklahoma

No school board can provide incentives to any teacher of more than 50% of salary

Incentives, including differentiated pay, are excluded from retirement calculations

Indiana

“Ineffective” or “improvement”
cannot receive any raise or
increment

Teachers can appeal

Ohio

Replaced salary schedules with salary "ranges"

Requires performance-based pay for teachers and nonteaching school employees.

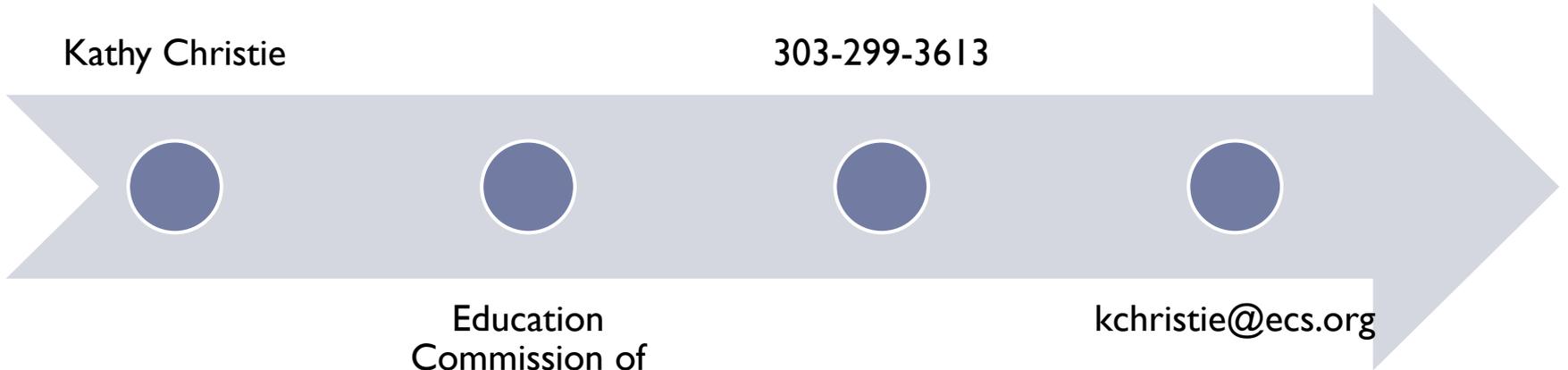
Michigan

Requires districts & schools to implement & maintain a “method of compensation for its teachers and school administrators that includes job performance and job accomplishments as a significant factor”

More questions?

Kathy Christie

303-299-3613



Education
Commission of
the States

kchristie@ecs.org

Education Commission of the States

- ▶ Legislatively established
- ▶ P-20 focused
- ▶ Cross-agency focused
- ▶ Policy focused
- ▶ Nonprofit
- ▶ Non partisan
- ▶ We work hard so you don't have to.

New Mexico's 2011 ECS Commissioners

