

Evaluating Teacher Evaluation Systems: Introduction and Conference Overview



SCOTT MARION

smarion@nciea.org

CENTER FOR ASSESSMENT

www.nciea.org

REIDY INTERACTIVE LECTURE SERIES

SEPTEMBER 13-14, 2012

Purpose

2

- To outline the problem and challenges we are trying to address during the next two days
 - How do we know what we're doing—in terms of teacher evaluation—will lead to more good than harm?
 - ✦ What would be “good?”
 - Improved student learning; increases in average teacher quality with a reduction in poor teachers
 - ✦ What would we consider “harm?”
 - Increased attrition of good teachers; narrowing/corrupting of curriculum to limit learning opportunities
- To describe the range and common features of the new generation of educator evaluation systems



Rationale new forms of teacher evaluation?

3

- **Teacher Quality:**
 - The influence of teacher quality on student achievement has been well documented
 - Traditional evaluation systems rarely provide feedback for improving teaching performance
- **Political-Economic Forces:**
 - Economic/incentives-based framework
 - Traditional evaluation systems rarely differentiate highly effective or ineffective teachers
 - Democrats for Education Reform and related and unrelated groups



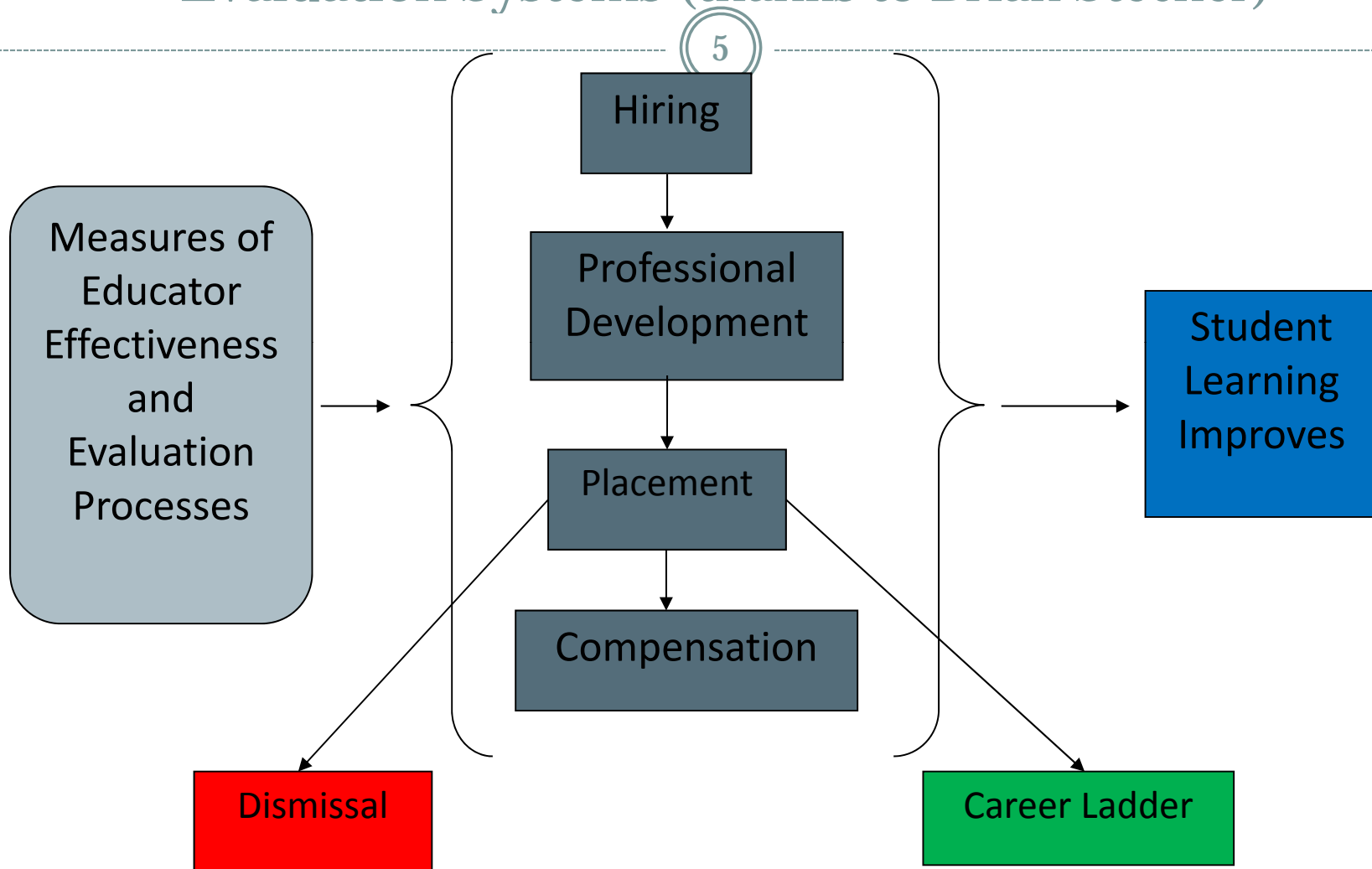
A Policy Maker's Implicit Theory of Action...

4

- We'll be hearing a lot more about theories of action or theories of use throughout the conference
- The theory of action on the following slide is oversimplified and somewhat naïve, but it helps to illuminate what appears to be driving much of the policy



A Policy Makers Theory of Action for Reformed Educator Evaluation Systems (thanks to Brian Stecher)



What could go wrong?

6

- The previous theory of action makes this seem so straightforward
- In fact, many policy makers seem surprised when they hear how hard this is



What's the Purpose?

7

- Many policy makers, including those who operate behind the scenes, appear to favor sorting and selecting
- Others, recognizing that it will be hard to *fire their way to the top*, are trying to design systems to support improving the skills of the majority of educators
- How should the system designs differ?
- This is an important distinction, but we argue that designing for improvement should be the focus



Validity

8

- Is about the quality of inferences that the system can support related to the purposes the system is designed to serve
- Requirements for validity evidence will be different for systems designed for sorting/selecting compared with those designed for improvement
 - Both are challenging, but I contend the supporting the inferences associated with systems designed for improvement are even more so...



Validation or Evaluation

9

- There is much disagreement in the field whether the concept of validity is broad enough to address the types of outcomes intended for educator evaluation systems
- Some argue that validity should be restricted to the measurement aspects while evaluation, drawing on research in program and personnel evaluation, should describe our efforts here



Validity and Evaluation

10

- We need both...
- Courtney and Henry will provide examples of very specific measurement challenges that clearly fall within a validity framework
- When we move to all the socio-political-economic intend and unintended outcomes, a case could be made that we shift to evaluation
- But not everyone agrees with this distinction
- More importantly, we don't need to get caught in academic crossfire when there is so much work to do



Theory of Action

11

- Several researchers have suggested that a theory of action can be a useful heuristic
- Can highlight the claims to be addressed in a validity argument
- Can point out intended (unintended) program effects that might be evaluated with a more traditional program evaluation orientation
- Also useful for helping to prioritize the work to get going!



An Overview of Current Systems

12

- As Courtney will describe, we do not often see an explicit definition of teaching or teacher quality
- Rather, there is an implicit recognition that teaching is complex enough to require attacking the challenge from multiple angles
- Teacher (or teaching) quality has therefore been operationally defined as some combination of the *things teachers do* and the *outcomes they produce...*
 - But not in all cases, and we will hear about a powerful example from Montgomery County, MD



Aspects of Teacher Quality and Measurement Tools

13

- **Measures of teacher practice (*things teachers do*)**
 - Observations
 - Artifacts
 - Professional portfolios
 - Student/peer input

- **Measures of student performance (*outcomes they “produce”*)**
 - VAM, SGP, or other “growth” measure for “tested” subjects
 - Approaches for non-tested subjects and grades (e.g., SLOs)



Aspects of Teacher Quality and Attribution

14

- **Attributing (generalizing) practice to a teachers requires good sampling and high quality measurement**
 - Courtney will talk a lot about this issue
- **Causally attributing student outcomes to teachers requires a strong causal evidence...and good sampling, quality measurement...and...**
 - Henry will be discussing these challenges



Major Classes of Challenges

15

- What is this thing called teacher (teaching) quality?
- What are some tools to measure specific aspects of this construct? Are they implemented well? How good are the data?
- What happens when these measurement results get transformed?
- How do the results from the various transformations lead to an overall determination?
- How do these results get attributed and used?
- What happens as a result of implementing these systems?



Why Must We Evaluate These Systems?

16

- Evaluation of newly developed educator evaluation systems is critical because we have an embarrassingly thin research base on which to rely
- Some very good emerging research, especially from the Measures of Effective Teaching (MET) project...
- Yet there is essentially no research or evaluation of these systems in the high stakes context in which they will be operating in the very near future (or now!)



Campbell's Law

17

- "The more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and **corrupt the social processes it is intended to monitor.**" (emphases added)

http://en.wikipedia.org/wiki/Campbell%27s_Law

- Educator accountability systems will invite significantly more implicit and explicit corruption and distortion than has been seen with school accountability



Theory of Action and Campbell

18

- Too often theories of action turn out to be pretty pictures of how systems should work
- A theory of action should also identify ways in which the system inadvertently incentivizes or invites *corruption*
- For instance, a TOA might specify that:
 - *higher scores on our district's SLOs reflects greater learning and better teaching of important content frameworks,”*
 - But we should also be alerted to considering alternative hypotheses:
 - *higher scores on our district's SLOs reflects a narrowing of the curriculum*
- A theory of action should anticipate and try to specify these possibilities



Overview of Conference

19

- Courtney and Henry will outline some critical features associated with conducting validity evaluation and research and together address the first four challenges
- They will provide specific examples of types of questions that can and should be investigated and offer a framework for doing so
- Our district leaders will share with us how these systems operate on the ground and the issues with which they are wrestling
 - We specifically choose districts because they have a much longer history of doing this work than states and have a much closer view of the challenges and effects



Conference Overview

20

- **Two sets of panel discussions tomorrow intend to:**
 - help us wrestle with the questions of both how to not lose the forest for the trees, while also providing some thoughts about how and where to start
 - will try to translate the important conceptual underpinnings into practical approaches for doing this work while the work is going on
- **Finally, we need you to wrestle with these ideas, challenge (respectfully) our speakers, propose solutions, have an open mind, and enjoy....**

