

Validation of Professional Practice Components of Teacher Evaluation Systems

Courtney Bell

Reidy Interactive Lecture Series
September 13-14, 2012

Overview

- Professional Practice in Policy Context
- Teaching v. Teacher Quality
- The Domain of Teaching Quality
- The Validity Argument Approach
- A Theory of Score Use
- An Interpretive Argument for Improving Teaching
- Exemplar Analyses

Professional Practice in Teacher Evaluation

- A mixture of definitions of the problem
- A mixture of implicit theories of professional improvement
- A mixture of uses for the scores
- A mixture of measures

Mixture of Problem Definitions

- Teaching practice is mediocre and teacher evaluation is a way to improve it
- There are some really poor teachers out there and some good ones. We need a reasonable way to sort these out and treat them accordingly.
- We need stronger public accountability at all levels of the system and teacher evaluation provides accountability at the teacher level.

Mixture of Improvement Strategies and Implied Theories of Motivation

- People will improve if you teach them
- People will improve if you have strong incentives and sanctions
- People will improve if their performances are more public

Mixture of Score Uses

- To make high stakes personnel decisions
- To improve teaching

A Mixture of Measures

- Observations
- Student work
- Student belief surveys
- Parent belief surveys
- Teacher reports of professional activities
- Etc.

Overview

- Professional Practice in Policy Context
- Teaching v. Teacher Quality
- The Domain of Teaching Quality
- The Validity Argument Approach
- A Theory of Score Use
- An Interpretive Argument for Improving Teaching
- Exemplar Analyses

Teacher and Teaching Quality

Imagine two surgeons. We would like to evaluate them on the quality of their surgical skills using multiple measures. We will use the size of the scar, the rate of infection, the quality of pain management, and patient satisfaction as our measures of the quality of their surgical skills. One is in Miami, Florida, the other in Moshe, Tanzania. Both must remove a benign tumor from a 53 year old man's abdomen. The surgeon in Miami has a #10 blade steel scalpel that is designed for cutting muscle and skin. The surgeon in Moshe has a well sharpened utility knife that is used for a range of surgical purposes. The excision in Miami will occur in a sterile operating room with no external windows, fans and filters to circulate and clean the air, an anesthesiologist, and a surgical nurse. The excision in Moshe will occur in a clean operating room washed with well water and bleach, windows opened a crack to allow the breeze to circulate the stifflingly hot air, no fans or filters, and a nurse borrowed from the pediatrics unit because she was the only available help.

Gitomer & Bell, in press

Teacher Quality vs. Teaching Quality

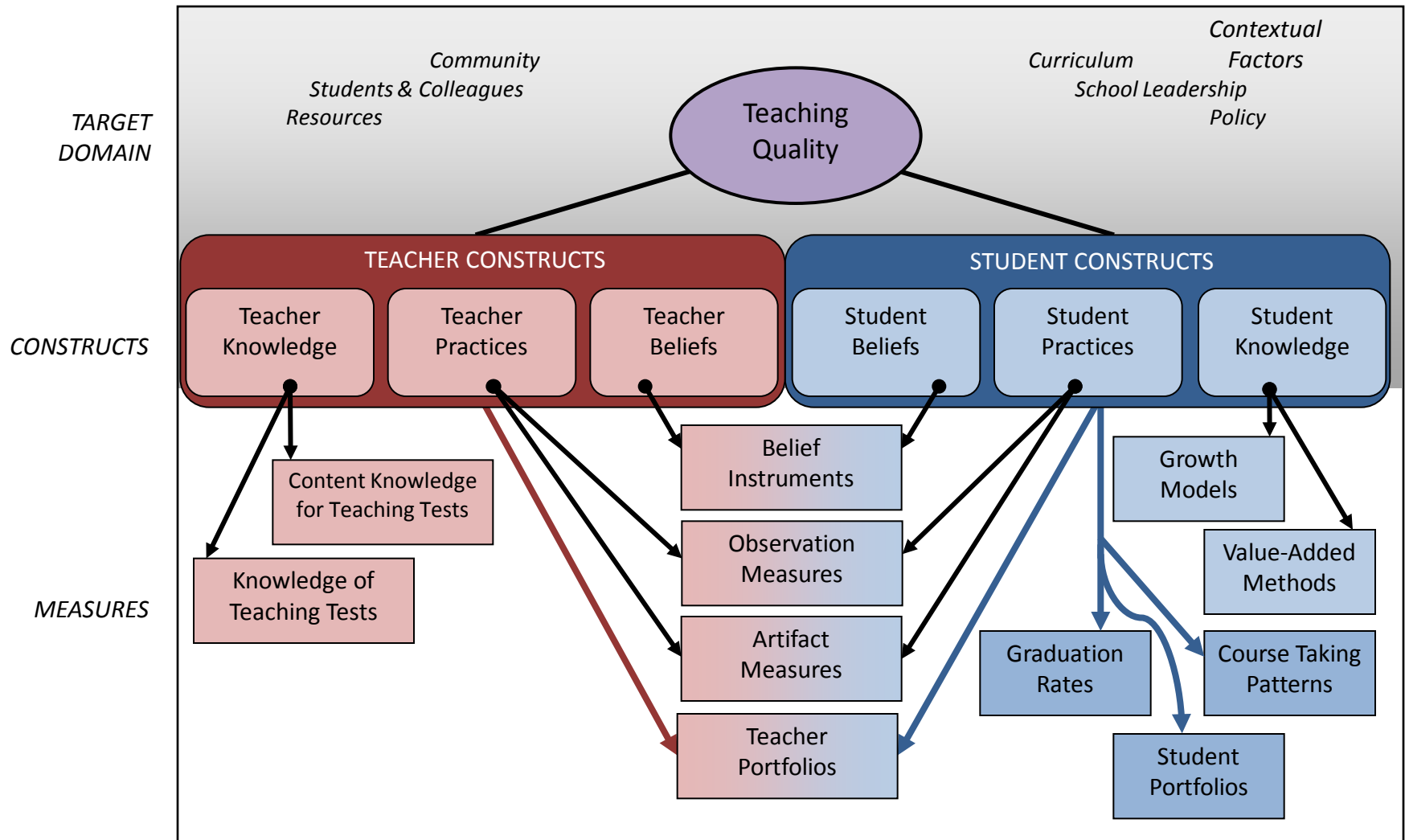
***Teacher Quality* assumes:**

- Stable attributes across contexts
- Fixed characteristics
- A focus on the characteristics of the teacher

***Teaching Quality* assumes:**

- Potential for systematic variability with context
- Potential for malleability
- A focus on the interactions of teaching and learning

The contextual factors, constructs, and measures associated with teaching quality



Bell, Gitomer, McCaffrey, Hamre, Pianta & Qi (2012)

Three Reasons to Clarify Constructs and Measures

- For making causal inferences
- For interpreting system outcomes and making sensible revisions
- For designing quality control, evaluation, and research plans

Overview

- Professional Practice in Policy Context
- Teaching v. Teacher Quality
- The Domain of Teaching Quality
- The Validity Argument Approach
- A Theory of Score Use
- An Interpretive Argument for Improving Teaching
- Exemplar Analyses

An Argument Approach to Validation

- Kane's (2006) validity framework suggests
 1. *Interpretive argument* – Describe the chain of inferences and assumptions that connect observed performance to claims
 2. *Validity argument* – Gather empirical evidence or logical support that demonstrates the “truthfulness” of the chain
- Validity argument judged in light of the proposed uses of scores
 - Improvement of teaching practice
 - Making high stakes personnel decisions

A Simple Example – An Interpretive Argument

- Luke can ride his bike on our sidewalk the same way he rides it on all the sidewalks on the way to school.
- When Luke falls down he can get up by himself and decide what to do next.
- Luke can cross the street safely.
- Luke knows the way to school.

A Simple Example – A Validity Argument

- All summer we watch Luke ride his bike on many of the sidewalks in our neighborhood and he is always able to ride.
- Luke fell twice over the summer. One time he just sat and cried. The other time he got up and kept going.
- 50% of the time Luke looks both ways.
- On 3 occasions, Luke was able to guide us to school.

Critical Aspects of Validation

- The purpose is critical
- Be clear about where you are making assumptions, using logic, using evidence
- Evidence is unlikely to be airtight
- The strength of the validity argument will be a value judgment

Overview

- Professional Practice in Policy Context
- Teaching v. Teacher Quality
- The Domain of Teaching Quality
- The Validity Argument Approach
- A Theory of Score Use
- An Interpretive Argument for Improving Teaching
- Exemplar Analyses

Theory of Use for A Professional Practice Measure

Information

- Professional practice measure creates information about the level of teaching skill demonstrated by each teacher

New Insights

- Information leads to insights about teaching

Practice and Learning

- Insights lead to the acquisition of new understandings and strategies through professional development, utilization of new tools, etc.

Changes in Teaching

- Understandings and strategies are incorporated into teaching practice, making it more effective

Changes in Learning

- Effective teaching practice results in improved student learning

Information on Skill Level

1. The scoring rule is appropriate.
2. The scoring rule is applied accurately & consistently.
3. The scoring is bias free.
4. The data fit the scoring model.
5. The sample adequately represents the quality of all relevant lessons.
6. The score on all lessons is related to the teaching quality teachers and students are able to enact.
7. There are not systematic errors that undermine the extrapolation to teaching quality.

Information on Skill Level Analyses (1)

- Assess the clarity of scoring criteria.
- Assess the degree to which observers use all the scorepoints and whether that is reasonable given other knowledge of variation in teaching performance.
- Assess the degree to which observers and master observers' scores on the same lessons match.

Information on Skill Level Analyses (2)

- Assess the degree to which similarities and differences across various groups of teachers are appropriate (e.g., by grade level, subject area, race, school).
- Investigate the sources of variation in scores (e.g., observer, lessons, time of year).
- Investigate the degree to which scores vary over school years.
- Investigate the influence of systematic errors (e.g., the influence of curriculum, co-teaching).

New Insights

8. Stakeholders understand the information the system produces.
9. Insights are appropriately and accurately related to the system information.
10. Insights are actionable.

New Insights – Analyses

- Investigate evaluators' and teachers' perceptions of what they learned from the information the system produced.
- Investigate the degree to which evaluators' written and verbal feedback was accurate and appropriate.

Practice & Learning

11. The implications associated with teaching performance are appropriate.
12. The properties of the observed scores on the lessons support the implications associated with the judgments of teaching performance.
13. Stakeholders link insights with appropriate tools, learning opportunities, etc.
14. Learning opportunities and resources are appropriate and of sufficient quality to support teacher learning.

Practice & Learning – Analyses

- Share videos of performance at different points in the score distribution and investigate the degree to which stakeholders view the implications associated with those scores points as appropriate.
- Conduct analyses that specify misclassification errors and levels of uncertainty over time.
- Investigate the alignment between the tools and PD activities teachers engage and their areas of demonstrated strength and weakness.
- Analyze the prevalence, coherence, timing, and quality of PD opportunities available to teachers.

Change in Teaching

15. Organizational context supports incorporation of new understandings/strategies.
16. Changes in teaching practice substantively align and are attributable to use of resources/professional learning.

Change in Teaching- Analyses

- Investigate stakeholders' perceptions of what they have learned, to what degree they have implemented their learning, and what impact implementation has had on their practices.
- Investigate teachers' perceptions about the support for their implementation of new understandings and strategies.
- Randomly assign teachers with similar needs to different professional development/resources and study changes in teaching practices.

Changes in Learning

- 17. Measures of student learning are sensitive to changes in teaching practice.
- 18. Changes in student learning are attributable to changes in teaching practice.

Changes in Learning - Analyses

- Investigate the relationship between changes in student learning and changes in scores on teaching practice measures at different points of proficiency for the teaching practice measures.
- Investigate the degree to which other measures that should reflect changes consistent with improved effectiveness actually demonstrate the theorized changes.
- Compare student learning on standardized tests with teacher reports of content coverage (Polikoff, 2010).

Don't Have 10 Million Dollars or Infinite Time?

1. Distinguish between QC and validation work
2. Disproportionate scrutiny on score validity
3. Focus on measures of concern

Two Cautions

1. Treat teaching quality seriously
2. Be mindful of the increasing validity burden moving toward teacher quality

References

- Bell, C.A., Gitomer, D.H., McCaffrey, D., Hamre, B., & Pianta, R., Qi, Y. (2012). An argument approach to observation protocol validity. *Educational Assessment, 17*, 1-26.
- Gitomer, D. & Bell, C. (in press). Evaluating Teachers and Teaching. In K. F. Geisinger (Ed.), *APA Handbook of Testing and Assessment in Psychology*.
- Polikoff, M.S. (2010). Instructional Sensitivity as a Psychometric Property of Assessments. *Educational Measurement: Issues and Practice, 29*, 3-14.