Listening. Learning. Leading.™

Advancing the Measurement of Teaching Quality

Drew H. Gitomer
ETS

Presented at IEL Conference
April 3 2009

Today's Talk

- A brief review of the current emphasis to measuring teaching quality and the challenges we face

- A description of a research project we are undertaking to help advance our understanding of ways of measuring teaching quality

Challenges to Measuring Teaching Quality

- **Policy Goals**
  - Accountability (teachers, programs, schools)
  - Human Capital Management (i.e., hiring, firing, compensation)
  - Social Capital Development (i.e., professional development, organizational development)

The goals are ambitious and based on two fundamental assumptions:

1. *Teachers matter*
2. *Quality is defined by impact on student learning*

Challenges to Measuring Teaching Quality

- **Methodological and Scientific Status**
  - Consistent inability to map teacher characteristics to teacher effectiveness
  - Dissatisfaction with traditional methods
  - Emergence of value-added methods (VAM) that are limited:
    - Limitations of available assessments – measures matter!
    - Only address a fraction of teachers (grades 3-8)
    - Issues of stability, teacher sorting
    - Black box – no information for social capital development

Conclusion: Current methods are not up to carrying the burden of policy ambitions. The complexity of teaching and instructional goals are not captured with current methods

The Need

- **A robust set of teaching measures and measurement practices that are**:
  - Theoretically grounded
  - Empirically supported
  - Feasible in standard educational contexts
  - Useful to satisfy accountability, human capital management and social capital development purposes
  - Part of a robust system of measuring teaching quality

Understanding Teaching Quality

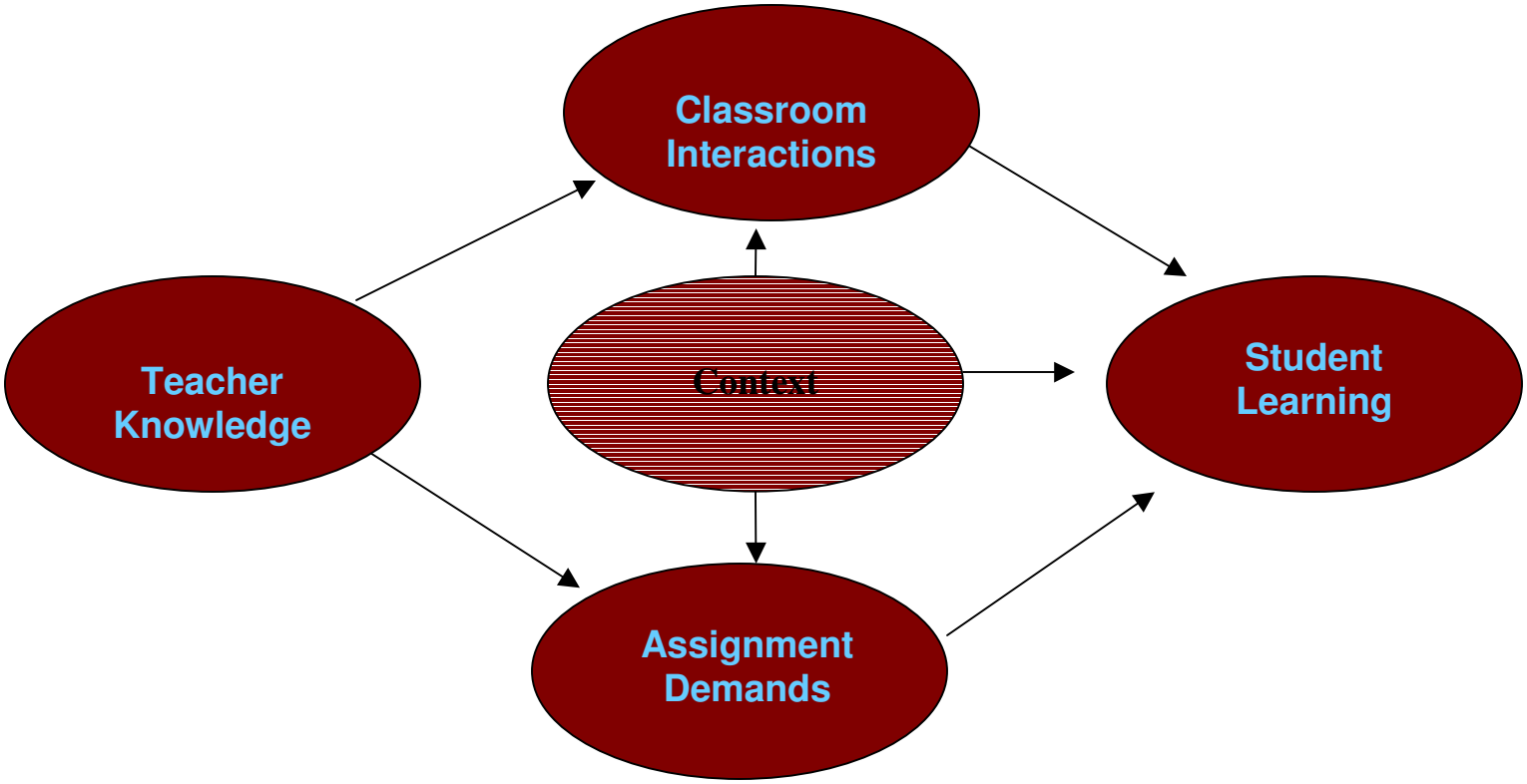
- Multi-year effort to refine, apply and understand a broad range of teaching measures
- co-PIs are Courtney Bell, ETS; Brian Rowan, University of Michigan; Dan McCaffrey, RAND
- Funded by the Bill and Melinda Gates Foundation
- Development of an engineered solution for scalability
- Middle school mathematics and language arts

Teacher Quality vs. Teaching Quality

- ***Teacher* Quality assumes:**
  - Stable attributes of the teacher across contexts
  - Characteristics relatively fixed
  - Focus is on the characteristics of the teacher

- ***Teaching* Quality assumes:**
  - Potential for systematic variability with context
  - Potential for malleability
  - Focus is on the act of teaching

High Level Construct Model of Teaching Quality



The Measures: Classroom Interaction

General, Domain-Independent Protocols

- Framework for Teaching (FFT) – Charlotte Danielson
- Classroom Assessment Scoring System Secondary (CLASS-S) – Bob Pianta et al.

Domain-Specific Protocols

- Mathematical Quality of Instruction (MQI) – Heather Hill etal
- Protocol for Language Arts Teaching Observations – Pam Grossman

Framework for Teaching

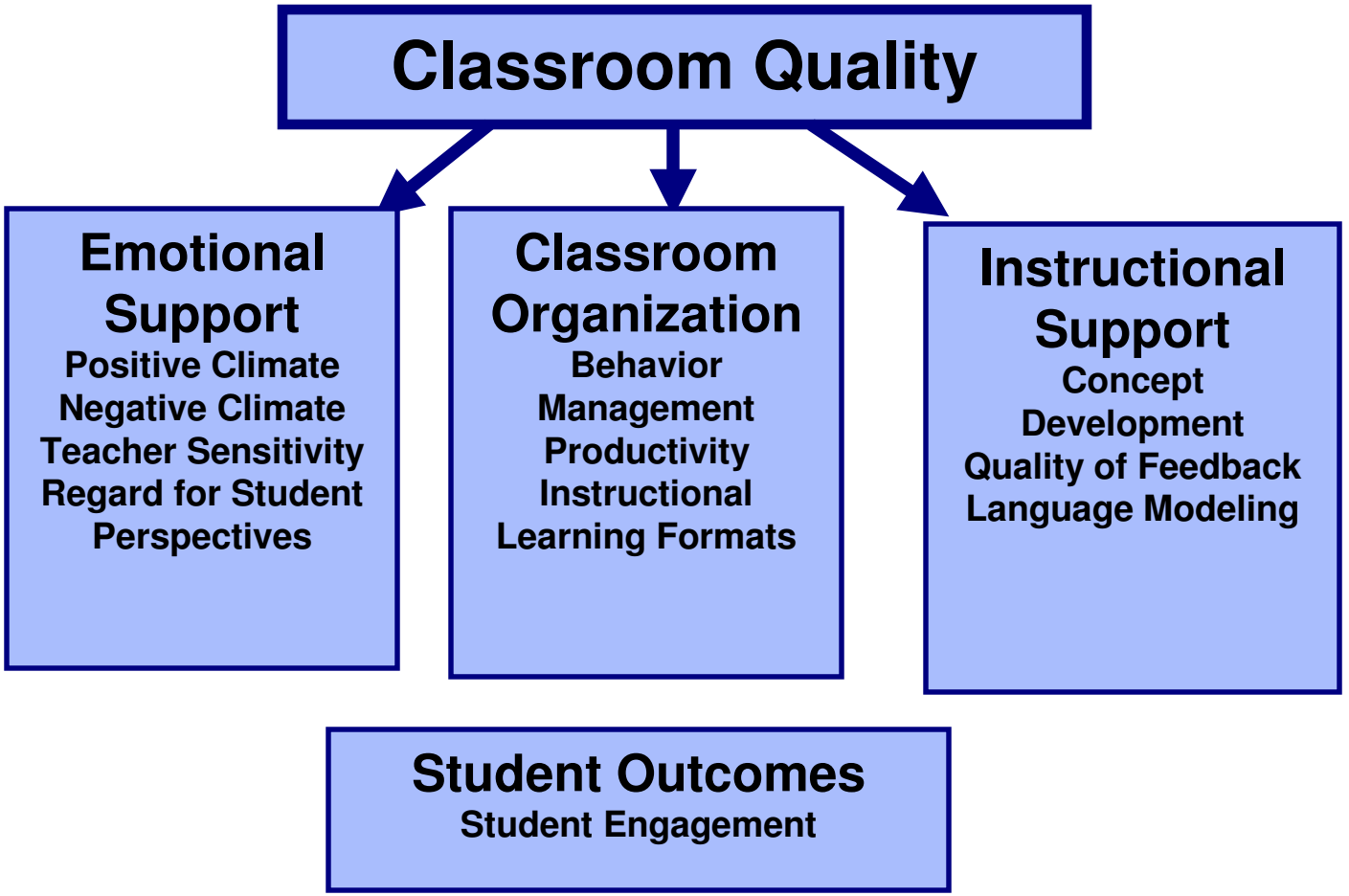
- **Domain 2 – Classroom Environment**
  - Creating an Environment of Respect and Rapport
  - Establishing a Culture for Learning
  - Managing Classroom Procedures
  - Managing Student Behavior
  - Organizing Physical Space

- **Domain 3 - Instruction**
  - Communicating with Students
  - Using Questioning and Discussion Techniques
  - Engaging Students in Learning
  - Using Assessment in Instruction
  - Demonstrating Flexibility and Responsiveness

- Scale – Unsatisfactory, Basic, Proficient, Distinguished

CLASS-S

From Pianta et al, 2007



Scale – Low, Medium, High – 7 points

MQI

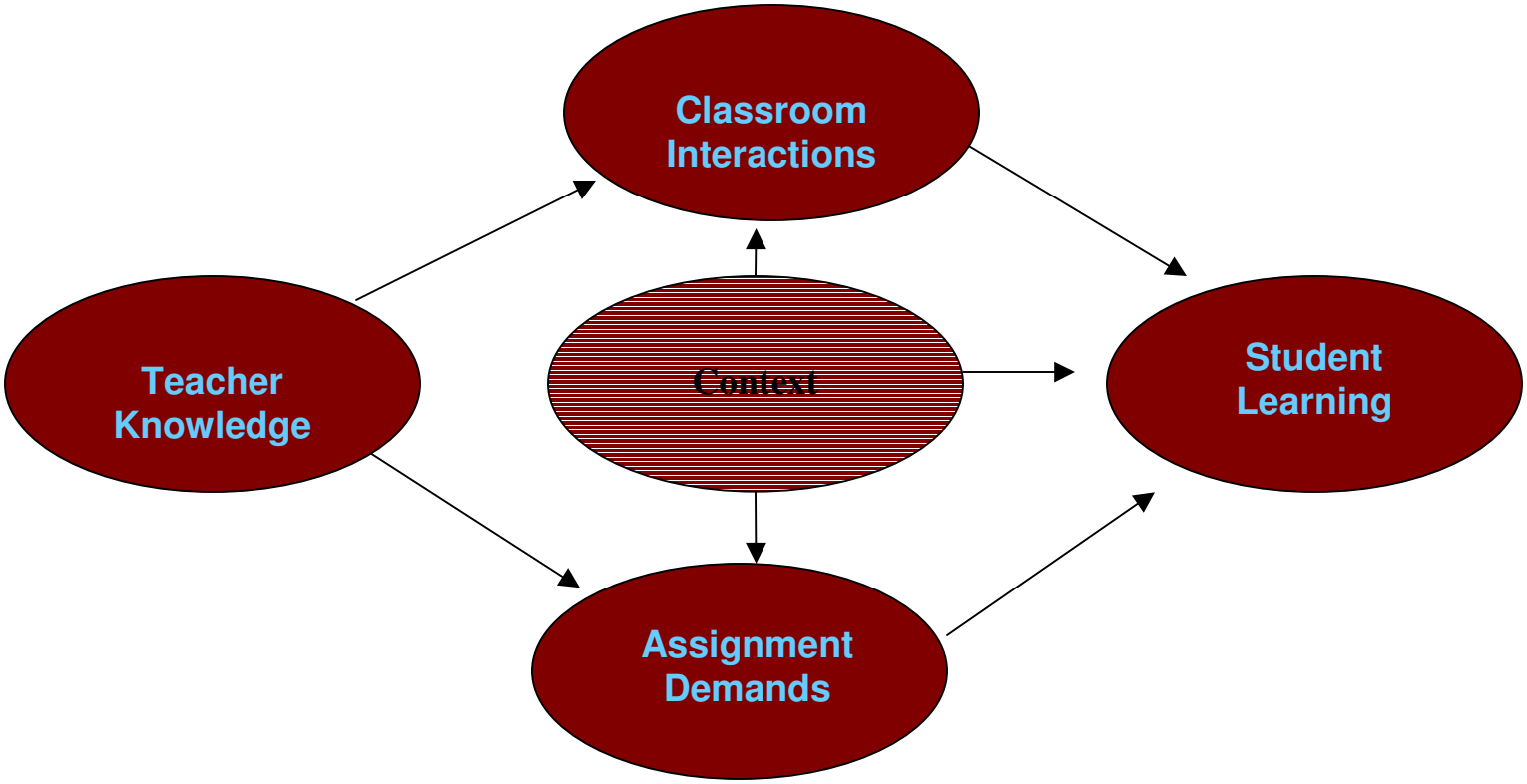
- Development of the mathematics in the classroom
  - -e.g., Multiple representations and/or models of mathematics content
- Errors in the mathematics
  - e.g. Errors in computation or serious mathematical oversight
- Mathematical interactions with students
  - e.g. Identifies mathematical insight in specific student questions, comments, work
- Student cognitive demand
  - E.g., Students analyze similarities and differences in representations, methods, or solutions
- Scale – Not present, Partially present, Characterizes the Lesson

PLATO

- Purpose
- Intellectual Challenge
- Representation of Content
- Connections to Prior Knowledge
- Connections to Personal and/or Cultural Experience
- Models/Modeling
- Explicit Strategy Instruction
- Guided Practice
- Classroom Discourse
- Text-based Instruction
- Accommodations for Language-learning
- Classroom Environment
- Content of Instruction

- Scale - Provides almost no evidence, Provides limited evidence, Provides evidence with inconsistencies, Provides consistent evidence

High Level Construct Model of Teaching Quality

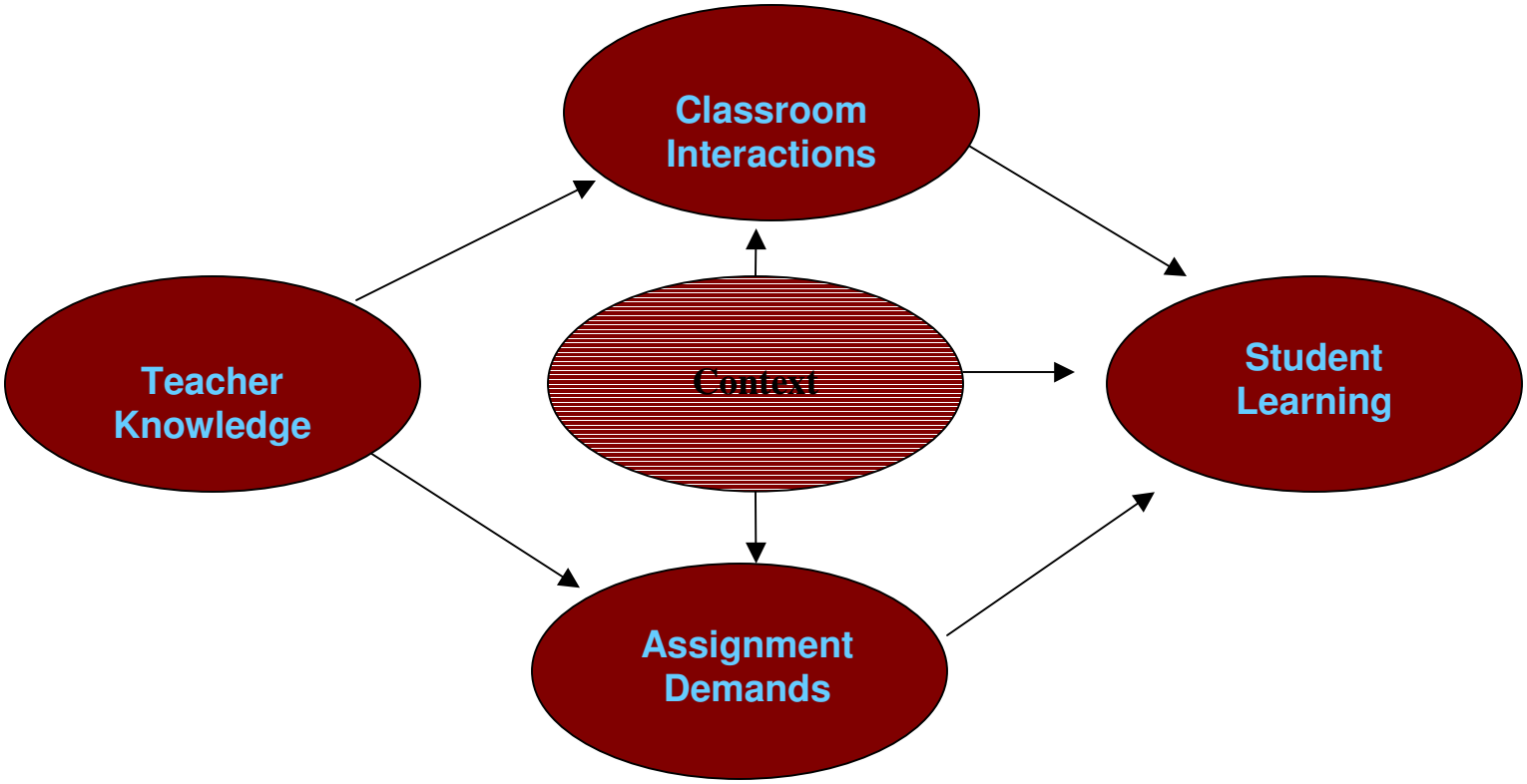


Assignment Demands

- Intellectual Demand Assignment Protocol (IDAP) – Fred Newmann etal. (writing and math protocols)
  - Construction of knowledge
  - Elaborated communication
  - Connection to student lives

- Assessment of both the teacher's assignment and the student work

High Level Construct Model of Teaching Quality



Teacher Knowledge and Student Learning Measures

- **Teacher Knowledge**
  - Mathematical Knowledge for Teaching (MKT) (Ball, Hill et al)
  - Praxis II items for English Language Arts
- **Student Learning**
  - Value-added measures based on statewide achievement tests

Research Questions

- How do different measures, examined separately and together, help us understand the strengths and weaknesses of teachers, their practice, and impact on their students?
- To what degree do different measures of teacher knowledge and teaching practice predict gains in student achievement?
- What is the relationship between measures designed to apply to all content areas and those designed for specific subjects?
- How stable are measures of the same teacher across different lessons and different groups of students (i.e., classrooms)?
- How can scalable systems of measures be engineered to provide detailed information about teaching effectiveness?

Study Overview

- Participants – 450 middle school teachers (225 math and 225 ELA)
- Data collected during 2009-2010 and 2010-2011
- Measures of teaching quality
  - 2 generic observation protocols (CLASS-S and FFT)
  - subject-specific observation protocol (PLATO or MQI)
    - Four observations times across two class sections
  - a subject-specific assignment protocol (IDAP)
    - 6 teacher assignments/ 2 with selected student work
  - a test of subject matter knowledge (MKT or Praxis)
  - Value-added measures based on state achievement tests

How can scalable systems of measures be engineered to provide detailed information about teaching effectiveness?

- Video as a data capturing tool
  - Supports potential for multiple judgments
  - Preserves practice for professional development
  - A database for continuing research and development
  - The ACOS System

- Video study for comparability

Automated Classroom Observation System

How do different measures, examined separately and together, help us understand the strengths and weaknesses of teachers, their practice, and impact on their students?

- Reliability of the measures
  - Ease of training observers
  - Observer reliability
  - Reliability across observations
  - Use of full scale for each measure
- Construct evidence
  - Empirical support for theoretical structure within and between measures
  - Empirical justification for domain score aggregation

To what degree do different measures of teacher knowledge and teaching practice predict gains in student achievement?

- **Do different lenses provide different results?**
  - What are the unique and joint relationship of different teaching measures to VAM?
  - What are the unique and joint relationship of different constructs (cross-instrument) to VAM?

What is the relationship between measures designed to apply to all content areas and those designed for specific subjects?

- How do the general and subject-specific observation protocols compare in their relationship to VAM? What is the unique contribution of subject-specific constructs?
- How do scores on content-related elements on the general protocols compare with subject-specific measures?
- How do the different measures compare on indicators of psychometric quality (e.g., reliability, scale use)?
- How do assignment demand and teacher knowledge measures compare with the different observation protocols?

How stable are measures of the same teacher across different lessons and different groups of students (i.e., classrooms)?

- To what extent do teaching practices vary across observations?

  Variability of lessons (4 observations)

- Can we begin to estimate context effects associated with different classrooms?
  - 2 observations in each of two sections
  - Value-added estimates also made for individual sections
  - Do teaching practices and student outcomes vary systematically across sections?

How can scalable systems of measures be engineered to provide detailed information about teaching effectiveness?

- **How do we evaluate the effectiveness of video?**
  - Feasibility of implementation
  - Issues of the scoring protocol (e.g., going back to review with video)
  - Compare scores to live observation – are live scores more "truthful"?
  - Correlations with VAM
  - Fit to theoretical model
  - Reliability within dimensions and overall instrument
  - Scale use
  - Cost/benefit analysis

Anticipated Outcomes

- A refined set of measures of teaching quality
- The basis for developing a new generation of teaching measures that have theoretical and empirical support
- An initial understanding of context effects on teaching practice
- A unique database for research exploration
- A beginning attempt at understanding the potential of video for evaluating teaching in a systematic fashion