Using Different Measures of Teaching Quality to Predict Student Learning in Mathematics: An Exploratory Study

Charalambos Y. Charalambous & Ermis Kyriakides
Department of Education, University of Cyprus
Structure of Presentation

- Why should we measure teaching quality accurately?
- Different approaches to measuring teaching quality
- Research purpose and research questions
- Methods
- Selected findings
- Discussion and tentative conclusions
- Lessons learned and open issues
Why Measuring Teaching Quality Accurately?

- Teachers matter for student learning:
  - Empirical studies have repeatedly documented teachers’ role for student learning (Hattie, 2009; Nye, Konstantopoulos, & Hedges, 2004; Strong, 2011)
  - Teacher effects have been found to explain a higher percentage of variance in student achievement compared to school-effects or system-level effects (Muijs & Reynolds, 2001; Scheerens & Bosker, 1999)

- Increased accountability pressures
  - Need to ensure that public expenditure on education is well spent (cf. Papay, 2012) –especially during an era of economic crisis
Several approaches pursued to measure teaching quality:

- Classroom observations (e.g., Douglas, 2009)
- Teacher logs (e.g., Rowan & Correnti, 2009)
- Principal ratings (e.g., Harris, Ingle, & Rutledge, 2014)
- Teacher ratings (e.g., Kyrgiridis et al., 2014)
- Student ratings (e.g., De Jong & Westerhof, 2001; Fauth et al., 2014)
Different Approaches to Measuring Teaching Quality

Classroom observations:

The “gold standard” of measuring teaching quality (Rowan & Correnti, 2009)

Can avoid many of the biases associated with self-reported data (Strong, 2011) → can yield more reliable data

Can produce stronger effects than those obtained through teacher self-reports or student surveys (e.g., Seidel & Shavelson, 2007)

Expensive to obtain

Estimates are influenced by a variety of factors, including the observational instrument, the recruitment and training of raters, the number and the length of observations to be conducted etc. (cf. Casabianca et al., 2013; Hill, Charalambous & Kraft, 2012; Praetorius, Lenske & Helmke, 2012)
Different Approaches to Measuring Teaching Quality

Teacher ratings:

Provide **inexpensive measures** of teaching quality with increased face validity (Kunter & Baumert, 2006)

**Correlations** between teacher self-reported data and student learning have been **moderate** (e.g., Mayer, 1999; Porter, 2002)

Teachers might deliberately (Blank, 2002) or unwittingly (Cohen, 1990) delineate their work in ways that depart notably from their actual practice → **significant bias**

Teachers’ reports on annual surveys **hardly capture the complexity and variability of their instruction** (Rowan & Correnti, 2009)

Friday, January 8, 2016
Different Approaches to Measuring Teaching Quality

Student ratings:

- Can have **even higher predictive validity** than classroom observations when aggregated at the classroom level (De Jong & Westerhof, 2001)
- Can **accurately delineate teachers’ day-to-day work** (Fauth et al., 2014; Hastie & Siedentop, 1999)
- **Cheaper** to obtain than classroom observations
- Can produce trustworthy measures of teaching quality, largely when students are asked questions about **easily observed behaviors** (Fauth et al., 2014; Panayiotou et al., 2014)
- Can be affected by factors such as **teacher popularity** (Kunter & Baumert, 2006)
Research Purpose and Research Questions

**Purpose:**
- Contribute to the ongoing dialogue about measuring teaching quality effectively and accurately
- Explore the predictive validity of classroom observations, student ratings, and teacher ratings
- Consider both cognitive and affective learning outcomes

**Research questions:**
- Which approach has more predictive power in determining student learning outcomes?
- Are these approaches differentially effective in predicting student learning when it comes to different types of learning outcomes?
Methods

- **Participants:**
  - 948 3rd to 6th elementary school students
  - 50 elementary school teachers

- **Data collection:**
  - **Cognitive learning outcomes:**
    - students completed a test measuring their performance in mathematics at the beginning and end of the academic year 2014-2015; test validated in prior studies (Kyriakides & Creemers, 2008)
  - **Affective learning outcomes:**
    - students completed a questionnaire measuring their attitudes and beliefs towards doing and learning mathematics (administered at the beginning and end of the academic year 2014-2015; questionnaire based on TIMSS survey)
Methods

Data collection:

Classroom observations:
- Each teacher was observed three times during the academic year by three independent raters, using two observational rubrics:
  - the *Dynamic Model of Educational Effectiveness* (Creemers & Kyriakides, 2008): generic teaching practices
  - the *Mathematical Quality of Instruction* (Learning Mathematics for Teaching, 2011): content-specific teaching practices

Student and teacher ratings:
- Student and teacher surveys completed at the end of the academic year 2014-2015
- Surveys explored certain generic or content-specific aspects of teaching quality
Methods

Data analyses:

- Rasch model applied to the student test data → a scale with satisfactory psychometric properties was developed
- Exploratory factor analyses applied to the student survey: three factors consistently yielded for both administrations; two met acceptable reliability thresholds (positive attitude toward mathematics; positive self-efficacy beliefs)
- Confirmatory factor analyses applied to observations/student ratings
  - Richness of the mathematics and cognitive activation (low inference classroom observation rubric)
  - Richness, cognitive activation, and focusing on mathematical procedures (high-inference classroom observation rubric)
  - Richness, cognitive activation, and working w/students & math (st. ratings)

Teacher ratings

- Richness, cognitive activation, mathematical procedures, and working with students and mathematics (no factor analysis applied because of small sample size)
Methods

Data analyses: Multi-level analyses

\[ Y_{ij} = \pi_{0j} + \pi_1 X_{1ij} + \sum_{s=2}^{S} \pi_s X_{sij} + e_{ijk} \] (Eq. 1)

Where:

- \( Y_{ij} \) is the end-of-year outcome (cognitive or affective) of student \( i \) taught by teacher \( j \);
- \( X_{1ij} \) is the variable corresponding to students’ initial cognitive or affective performance [grand-mean centered)] (entered in Model 1);
- \( X_{sij} \) are the student background characteristics (gender [dummy variable], and SES indicators) (entered in Model 2);
- \( \pi_{0j} \) is the adjusted mean performance for students of teacher \( j \) after controlling for student initial performance and background characteristics;
- \( \pi_1 \) is the fixed effect of student beginning-of-year performance;
- \( \pi_s \) are the fixed effects of student background characteristics;
- \( e_{ijk} \) is the random “student effect,” that is the deviation of student \( i \) of teacher from the teacher-group mean.
Methods

- Data analyses:
  - Multi-level analyses

\[ \pi_{0j} = \beta_{00} + \sum_{m=1}^{M} \beta_{0m} W_{mj} + u_{0j} \]  
(Eq. 2a)

\[ \pi_{0j} = \beta_{00} + \sum_{n=1}^{N} \beta_{0n} W_{nj} + u_{0j} \]  
(Eq. 2b)

\[ \pi_{0j} = \beta_{00} + \sum_{p=1}^{P} \beta_{0p} W_{pj} + u_{0j} \]  
(Eq. 2c)

Where:

- \( \beta_{00k} \) is the grand mean;
- \( W_{mj} \) are the content-specific teaching practice scores from lesson observations of teacher \( j \) (grand-mean centered);
- \( W_{nj} \) are the content-specific teaching practice scores from student ratings for teacher \( j \) (grand-mean centered);
- \( W_{pj} \) are the content-specific teaching practice scores from teacher ratings for teacher \( j \) (grand-mean centered);
- \( \beta_{0m} \) are the effects of content-specific practices for the observational scores;
- \( \beta_{0n} \) are the effects of content-specific practices for student ratings;
- \( \beta_{0p} \) are the effects of content-specific practices for teacher ratings;
- \( u_{0j} \) is the random “teacher effect,” that is the deviation of teacher \( j \)’s mean from the grand mean.
Selected Findings

- Cognitive learning outcomes:
  - 28% of the variance at the teacher level in the null model, but only 3% remained unexplained after introducing pre-test results.
  - Used student learning as the dependent variable:
    - 9.69% of the variance at the teacher level.
    - Percentage of unexplained teacher-level variance explained when introducing:
      - Classroom observations (factors): 17.65%
      - Student ratings (factors): 0%
      - Teacher ratings (composites): 0%
      - Classroom observations (individual codes): 58.82%
      - Student ratings (individual statements): 8.40%
      - Teacher ratings (individual statements): 57.14%
Selected Findings

- Affective learning outcomes (positive attitudes):
  - 14.88% of the variance at the teacher level in the null model
  - 8.76% of the variance at the teacher level remained unexplained once introducing the initial measure
  
  Percentage of unexplained teacher-level variance explained when introducing:
  - Classroom observations (factors): 0%
  - Student ratings (factors): 37.63%
  - Teacher ratings (composites): 0%
  - Classroom observations (individual codes): 30.11%
  - Student ratings (individual statements): 59.14%
  - Teacher ratings (individual statements): 44.09%
Selected Findings

- Affective learning outcomes (positive self-efficacy beliefs):
  - 4.43% of the variance at the teacher level in the null model; 2.99% of the variance at the teacher level remained unexplained once introducing the initial measure.
  - Used the difference as the dependent variable (4.70% unexplained variance at the teacher level).

- Percentage of unexplained teacher-level variance explained when introducing:
  - Classroom observations (factors): 0%
  - Student ratings (factors): 25.71%
  - Teacher ratings (composites): 22.86%
  - Classroom observations (individual codes): 28.57%
  - Student ratings (individual statements): 31.43%
  - Teacher ratings (individual statements): 37.14%
Discussion and Tentative Conclusions

- Some interesting patterns:
  - Using factors or composites:
    - **Cognitive results**: classroom observations > student/teacher ratings
    - **Affective results**: student ratings first and classroom observations last
  - Using individual statements:
    - **Cognitive results**: classroom observations ≈ teacher ratings > student ratings
    - **Affective results**: student/teacher ratings > classroom observations
  - **Which measurement approach is best?**
    - It depends on the type of the learning outcome considered
    - It depends on whether composites or individual statements are being used
Lessons Learned and Open Issues

- Importance of considering different learning outcomes; **cognitive or affective learning outcomes in isolation yield only part of the story**
  - Why these differences occur calls for future (more qualitative?) studies

- Results concern content-specific teaching practices; it remains an open issue whether these patterns are **replicated for generic teaching practices**

- Importance of **combining different approaches** to better understand student learning: difficult in the present study because of the small percentage of teacher-level variance and issues of multicollinearity

- Using **composites or individual statements**?
  - Do composites have more noise than individual statements?
Thank you for your attention!

- Comments
- Questions
- Suggestions

Charalambos Y. Charalambous
cycharal@ucy.ac.cy