Using Surveys as Proxies for Observations in

Measuring Mathematics Instruction

David Braslow and Andrea Humez

Harvard Graduate School of Education, Boston College Lynch School of Education

Abstract

Using data from elementary mathematics teachers, we examine the correspondence between self-reports and observational measures of two instructional dimensions—reform-orientation and classroom climate—and the relative ability of these measures to predict teachers' contributions to student learning.

Keywords: classroom assessment, mathematics domain, surveys, latent variables, reporting

Using Surveys as Proxies for Observations in

Measuring Mathematics Instruction

Researchers studying the relationship between aspects of instruction and student outcomes must choose between substantively different measures of instructional practices. When implemented by trained raters, observational measures provide more objective evidence than surveys, but incur high costs and contain information only about observable features of instruction from specific sampled occasions. Surveys are less costly, but raise concerns about self-report bias and misalignment between teachers' and researchers' definitions of key terms (Burstein et al., 1995).

Past research has used observations to validate survey instruments, finding that correlations between observation and survey measures of the same construct can be substantial, but that not all constructs/practices are equally well-captured by both data collection methods (Mayer, 1999; Porter & Smithson, 2001). However, existing studies have neither examined contemporary observation instruments nor addressed the potential use of teacher self-reports as proxies for classroom observations for the purpose of predicting teachers' contributions to student learning. Current interest in linking aspects of instruction to teacher value-added model (VAM) scores lends urgency to the question about how much the method of data collection might influence research findings.

This study extends the existing literature by using data from the overarching National Center for Teacher Effectiveness (NCTE) study to address the following research questions:

- To what extent do self-report and observational measures of reform-oriented instruction and classroom climate capture similar underlying constructs?

- To what extent do the different measures of these dimensions predict teachers' VAM scores?

- To what extent do the different measures of these dimensions predict the *same* portion of the variance in teachers' VAM scores?

To answer these questions, we use data collected from a survey designed by NCTE, an observational measure of reform-oriented instruction drawn from the MQI instrument, and an observational measure of classroom climate measure drawn from the Classroom Assessment Scoring System (CLASS) instrument. To answer our first question, we use a multitrait-multimethod matrix to compare the magnitudes of the correlations between survey and observational measures of the two instructional dimensions. This analysis provides evidence that the two measures of each instructional dimension are effectively capturing some of the same variation in instruction.

The second and third research questions concern the relationship between different measures of instructional dimensions and teachers' contributions to student learning. Using OLS regression, we predict VAM scores using each of the instructional measures. By examining the coefficients associated with the different measures, we can evaluate the extent to which each measure predicts value-added scores.

Furthermore, if the two instruments do indeed measure the same construct, we would expect them to be collinear and to account for the same portion of the variation in VAM scores. By including both survey and observational measures in an OLS regression as predictors of VAM scores, we can assess the extent to which different measures of instructional dimensions have different predictive power in assessing teachers' contributions to student learning. Initial analyses suggest that different measures of the same instructional dimensions do not capture different features of instruction that might explain VAM scores.

These findings have implications for researchers who wish to use surveys and/or observations to capture or investigate how features of instruction relate to teachers' contributions to student learning. Insofar as different measures capture similar portions of variance in teachers' contributions to student learning, it may be possible to use surveys as proxies for observations in teacher effects research; however, if these measures capture different portions of variance, they should not be considered interchangeable.

References

Burstein, L., McDonnell, L. M., Van Winkle, J., Ormseth, T., Mirocha, J., & Guitton, G. (1995).

Validating national curriculum indicators. Santa Monica, CA: RAND.

Mayer, D. P. (1999). Measuring instructional practice: Can policymakers trust survey data?

Educational Evaluation and Policy Analysis. 21(1)*, 29-45.

Porter, A. C. & Smithson, J. L. (2001). Defining, Developing, and Using Curriculum Indicators.

Philadelphia, PA: Consortium for Policy Research in Education. Retrieved from

http://http://www.cpre.org/sites/default/files/researchreport/788_rr48.pdf