Building Better Measures of Effective Teaching: Year 1 Results
Organizer: Steve Cantrell, Bill & Melinda Gates Foundation

Abstract

With funding from the Bill & Melinda Gates Foundation, a number of leading researchers and education professionals are working to develop multidimensional measures of effective teaching. While a perfect measure is unattainable, educators should strive to develop indicators of teacher effectiveness that are recognized as fair, accurate, and useful. In partnership with six school districts, including over 3,000 teachers and their students, we are developing and implementing new approaches to combining student achievement data, digital videos of classroom instruction, assessments of teacher pedagogical knowledge, and student and teacher perceptions to identify effective teaching and provide objective information to teachers and administrators. We propose this symposium as a forum to discuss Year 1 results of this two-year project.

Session Summary

All children, provided appropriate and sufficient supports, can learn. However, without an effective teacher even the most advantaged student may struggle in school. For decades researchers have been building the case that teachers’ behaviors are powerfully influential in shaping student learning (for an early review, see Brophy & Good, 1986). However, there is little agreement among education stakeholders about how to identify and measure components of effective teaching. Typical teacher evaluation methods are based largely on characteristics unrelated to student achievement (Weisberg, 2009). In rare instances when student learning is considered, the focus is narrowly on test scores. Such evaluation systems do not provide meaningful feedback to help teachers improve their practice or inform administrators’ decisions regarding teacher recruitment, development, placement, and retention.

In fall 2009, the Bill & Melinda Gates Foundation launched the Measures of Effective Teaching (MET) project to develop and test multiple measures of teacher effectiveness. Led by more than a dozen organizations including, academic institutions, nonprofit organizations, and for-profit education consultants, the MET project aims to improve the quality of information about teaching effectiveness available to education professionals.

To help identify the best mix of teacher effectiveness measures, more than 3,000 teacher volunteers are participating in the MET project across six predominantly urban school districts. Participants teach math and English Language Arts (ELA) in grades 4-8, Algebra I, grade 9 English, and high school biology. In addition to the existing schedule of student assessments administered by the district, new data collection includes supplemental assessments of students’ higher order conceptual understandings, digital video recordings of classrooms, assessments of teachers’ pedagogical content knowledge, and surveys of both students’ and teachers’ school experiences. This combination of diverse data collection methods allows us to move beyond considerations of individual behaviors or interactions, toward a richer understanding of teaching and learning processes. Presently, Year 1 data collection is complete and analyses are underway.

To isolate teacher effects, the study must account for differences amongst students. To do so, participating teachers have agreed to be randomly assigned to the students they will teach during Year 2 (academic year 2010-11). By signing up for the project in groups of three or more...
colleagues teaching the same subjects and grade level within the same school, the randomization becomes feasible. The random assignment will allow us to make reasonable causal inferences regarding the impact of particular teacher behaviors on student outcomes.

In line with this year’s meeting theme, this research seeks to develop new approaches to advance our understanding of educational quality. With measures of good teaching that adequately reflect the complexities of the classroom, we will be better positioned to support effective teachers and make changes where practice is ineffective.

The symposium will begin with a ten-minute introduction provided by project leaders from the Bill & Melinda Gates Foundation. Then, each of the panelists will present for 15 minutes on the Year 1 results of their respective work. The discussant will offer ten minutes of comments followed by 25 minutes of question and answer.

References


Scoring 13,000 Lessons Using Multiple Observation Protocols
Catherine McClellan, Educational Testing Service

As part of the MET study, video captures of approximately 13,000 classroom lessons in Year 1 were scored using five observation protocols: CLASS, FfT, MQI, QST, and PLATO. The specific focus of this element of the study was to provide maximally useful data to the composite measure of effective teacher being built. Another purpose was to determine the operational and fiscal feasibility of scoring videos of this type of complex performance at a large scale with high psychometric quality.

The scoring structure of the MET study meant that the 13,000 lessons translated into about 30,000 sessions to be scored, since many lessons were scored on two or three of the study protocols. This component of the panel discussion describes the research and analyses leading up to the decisions about the final scoring design as well as implementation data. Included are descriptions of the evaluation of single- versus multiple-scale scoring for each rater; appraisal of approaches to streamlining the instruments while maintaining construct integrity; investigation of methods for sampling the video sessions; creation of online, distributed rater training of sufficient depth and quality to prepare raters for live scoring; and methodology for monitoring quality during live scoring.

Data sources for the research include scores provided by the instrument developers as the criterion measure against which scoring quality was measured; data collected in think-aloud interviews with raters attempting the proposed scoring designs; outcomes from the rater training, certification, and calibration procedures; and results from the scoring quality monitoring assessments.

Techniques used in the analysis include factor and cluster analyses, frequency counts, correlation, agreement statistics, and optimization of rater selection using a newly-developed penalty function approach. The results of this work are expected to inform developing policy around teacher effectiveness, professional development, and career track determination garnering a great deal of public attention in Race to the Top grants and the ESEA re-authorization.
Challenges and Opportunities in Using New Measures of Effective Teaching to Inform Human Capital Management in the Pittsburgh Public Schools.
Paulette Poncelet, Pittsburgh Public Schools

Pittsburgh Public Schools (PPS) is a medium sized urban school district. Administration and union efforts to increase student achievement and college readiness are focused on increasing the number of highly effective teachers in the District, increasing the exposure of high-need students to highly effective teachers, and ensuring that all teachers work in learning environments that support their ability to be highly effective. These strategic priorities have benefited from a unique set of conditions in Pittsburgh, including the Pittsburgh Promise scholarship program, the development of a new teacher evaluation system known as the Research Based Inclusive System of Evaluation (RISE), selection as an Intensive Partnership Site by the Bill & Melinda Gates Foundation and participation in the Measures of Effective Teaching (MET) research project as a pilot site.

Seven initiatives are underway to achieve the identified strategic priorities. These initiatives include the development of a set of measures of effective teaching to inform teacher practice and evaluation, the creation of a career ladder with differentiated compensation that rewards and recognizes effective teachers and provides the type of growth opportunities necessary to retain talented professionals, the establishment of a new professional learning Academy; the development of the Promise Readiness Corps, a team of highly effective teachers collectively responsible for supporting students transitioning into high school and ensuring they arrive at 11th grade Promise-ready; the transformation of the human resources department, the adoption of a new information technology system to support the initiatives, and the measurement and improvement of teaching and learning environments throughout the District.

The District’s Office of Research, Assessment & Accountability is charged with facilitating the application of quantitative measures to inform the development and implementation of the initiatives and to monitor progress. Challenges have included: 1) communicating technical measurement concepts to practitioners so that they embrace and use them to advance practice and 2) expanding measurement to teachers of non-tested subject areas. Multiple data sources including Value Added estimates derived from existing assessment data, RISE data, and data collected through the pilot of MET project measures, including Student Perceptions Survey, Teacher Working Conditions Survey, and Assessments of Pedagogical Content Knowledge, will be explored for their ability to validate, expand and improve the accuracy with which effective teaching is quantified. Multiple working groups have been assembled to inform this work, including the Student Learning Sub-Committee, consisting of approximately 50 volunteer teachers and principals from across the District; the VAM Development Team, consisting of representatives from the Pittsburgh Federation of Teachers (PFT); central administration; the American Federation of Teachers (AFT); Mathematica Policy Research, Inc. and a Technical Advisory Board, consisting of a small group of national experts in the application of Value Added Measurement to teacher evaluation.

The practical significance of these efforts includes the ability to more accurately and objectively identify effective teaching. Additionally, the ability to reward and recognize teachers based on their effectiveness, provide a career ladder and individualized professional learning will improve the retention of effective teachers and encourage teachers to continue to grow professionally.
Correlations between Student Perception Survey Results and Value-Added Scores
Ronald Ferguson, Harvard University

1. Objectives or purposes
The purpose of this work is to document the degree to which classroom-level student perceptions as reflected in student survey responses are predictors of value added achievement gains. This is the first phase of work toward measuring causal relationships between students’ classroom experiences and learning.

2. Perspective(s) or theoretical framework
Student survey questions are based upon several strands of research in teaching and learning. To capture student experience in the classroom, we measure what we call the “Seven C’s” of effective teaching: Caring, Controlling, Clarifying, Challenging, Captivating, Conferring and Consolidating. These capture themes from various overlapping literatures on achievement motivation, classroom management, cognition, and youth development.

3. Methods, techniques, or modes of inquiry
The analyses in the paper will use a combination of graphical representations, correlations, and multiple regressions. The presentation will report statistical reliabilities of the measures employed. Using multi-item indices, the measurement model for reliabilities is a four-level model in which the levels are: 1) item-level; 2) student-level; 3) classroom-level; and 4) school level. The model has been developed in collaboration with Stephen Raudenbush of the University of Chicago, who is an expert on measurement modeling in the social sciences.

4. Data sources, evidence, objects, or materials
The data come from classroom-level student surveys and standardized test-scores in several large urban school districts. All are part of the Gates Foundation project on Measuring Effective Teaching.

5. Results and/or substantiated conclusions or warrants for arguments/point of view
This project builds on the author’s past work as part of the Tripod Project for School Improvement. In that work, student survey responses have been predictors of value-added achievement gains in mathematics. Most of the findings to be reported in association with this conference proposal will be generated between now and the early fall of 2010 using data from schools participating in the MET project.

6. Scientific or scholarly significance of the study or work
Finding ways to measure effective teaching is a major national priority. As part of the MET Project, the findings to be reported in this paper will contribute importantly to our understanding of how student perceptions of teaching are related to learning outcomes.
This presentation will provide details for the method used in the MET project to estimate a teacher's contribution to student performance. This method supposes that the value-added estimate is composed of two components, a stable component (representing the "true" teacher effect for a given type of test) and all other non-persistent classroom shocks and sampling variation.

The authors estimate the portion of the variation in teacher-level value-added estimates that is "stable" by studying the correlation within teacher across sections in a given year or between years. They calculate the correlation in the "stable" components of the teacher effects for two tests by taking the covariance in teacher-level value-added estimates for different tests across different sections and dividing it by the square root of the product of the standard deviations in the stable component for each of the tests.