

Chiefs' Research Primer:

A summary of Measures of Effective Teaching's (MET) recent findings, *Gathering Feedback for Teaching: Combining High-Quality Observations with Student Surveys and Achievement Gains*.

Prepared by the Council of Chief State School Officers, January 2012

A Product of CCSSO's Research and Development Service

CCSSO's Research and Development (R&D) service operates by convening external expertise around state education leaders' most essential questions in order to help find the most reliable and actionable evidence to direct state policy development and implementation. Among the R&D service's activities is the creation of research primers that summarize findings from recent key research and translate the findings into policy considerations for chief state school officers and their staff.

For more information, please contact:

Nina de las Alas, ninaa@ccsso.org

Jennifer Davis, jenniferd@ccsso.org

Lauren Stillman, laurens@ccsso.org

Source Reviewed:

Bill and Melinda Gates Foundation, Measures of Effective Teaching (MET). (2012). *Gathering Feedback for Teaching: Combining High-Quality Observations with Student Surveys and Achievement Gains*. Available:

http://www.metproject.org/downloads/MET_Gathering_Feedback_Research_Paper.pdf

Relevant Findings

Utility of Teacher Observations

The teacher observation instruments studied were effective in predicting teachers with greater student achievement gains and grew in reliability and predictive power as more ratings – and more non-observational data – were taken together.

- Each of the five teacher observation instruments under review was found to be effective in predicting teachers with greater student achievement gains.
 - Framework for Teaching (FFT), Charlotte Danielson, Danielson Group
 - Classroom Assessment Scoring System (CLASS), Robert Pianta, Karen La Paro, and Bridget Hamre, University of Virginia
 - Protocol for Language Arts Teaching Observations (PLATO), Pam Grossman, Stanford University
 - Mathematical Quality of Instruction (MQI), Heather Hill, Harvard University
 - UTeach Teacher Observation Protocol (UTOP), Michael Marder and Candace Walkington, University of Texas-Austin
- As teachers' observation scores increased, so did their value-added scores. This was true for all observation instruments.
 - However, while statistically significant, the differences were not large. For example, the difference in student learning gains between the top and bottom 25 percent of teachers (based on their FFT scores) was equivalent to approximately 2.7 months of schooling (assuming a 9-month school year) for state math tests, and only 0.6 month for state English language arts tests.
- Reliability of ratings increased when averaging multiple lessons by multiple raters.
 - A single observation score was not particularly predictive of a teacher's practice, as teachers' scores varied considerably from lesson to lesson.
 - Raters often disagreed on their ratings for a particular lesson, but few raters were consistently "too easy" or "too hard."
 - Teachers' scores varied little based on the students in the class or time of day.
- Predictive power and reliability increased when teachers' observation scores were combined with additional data: 1) student feedback, and 2) student achievement gains from another year or group of students.
 - The addition of student feedback to teachers' rankings further separated the top and bottom teachers as reflected in an overall difference of student learning of 4.8 months.

- The addition of both student feedback and student achievement gains from another year or group of students further separated the top and bottom teachers as reflected in an overall difference of student learning of 7.6 months – almost an entire school year.
- Graduate degrees or years of teaching experience were not as effective as the combination of observation scores, student feedback, and student achievement gains at predicting a teacher’s student achievement gains (as well as other measures of student success).
 - Years of experience separated the top and bottom teachers by a margin of only 0.5 month of schooling.
 - The difference between teachers with and without master’s degrees was equivalent to only 1.0 month of schooling.

Trends in teacher performance

Teachers’ scores clustered in the mid-range of performance with higher scores in management competencies and lower scores in instructional competencies.

- Teachers’ scores were mostly in the mid-range of performance, as defined by the instruments.
- Teacher scores were highest for behavioral-, time-, and materials-management competencies, and lowest for the aspects of instruction, including teaching students higher-order thinking skills.
- The study was not able to separate out which specific teacher competencies were most predictive of student achievement.

Policy Considerations

- 1. Evaluation Instrument Selection/Development – Instruments may vary in terms of specific competencies, but should set clear and realistic expectations for observers and be effective in predicting student achievement.**
 - Consider alignment of instrument to school system’s theory of action.
 - Keep number of competencies to a manageable level.
 - Describe competencies in sufficient detail to allow observers to score reliably.
 - Avoid simple checklists; instead, focus instrument on alignment of practice with competencies.
- 2. Observer Selection and Development – Train, assess, and monitor observers for reliability.**
 - If possible, use impartial observers. Consider using video for remote observers.
 - Train observers using combination of discussion, videos, practice, and techniques for limiting bias.
 - Require observers to demonstrate accuracy before they rate teacher practice.
 - Define minimum accuracy level for observers.
 - Periodically reassess observers using calibration assessments or double scoring comparisons (ideally with impartial observers).
 - Given limited resources, consider shorter, more targeted observations for teachers who have demonstrated mastery of basics.

3. **Professional Development** – Consider using observation instruments as professional development tools by focusing professional development on teachers' specific needs as measured by the instruments.
4. **Score Reliability** – Average multiple lessons by multiple raters. This is particularly important in high-stakes situations
5. **Instrument Validity** - To increase instrument validity, combine observation scores with additional data, such as:
 - Student feedback data
 - Student achievement gains from another year or another student group
6. **Quality Control** – Regularly verify that teachers with stronger observation scores also have stronger student achievement scores on average.
7. **Implications for Further Research**
 - Future research should explore how best to weight different measures as well as the effects of high-stakes situations on the effectiveness of the observation tools.