



Teacher Effectiveness: Can We Measure this in an Informative Way?

Franklin W. Bender
University of Oregon



| Background | | Defining Teaching Effectiveness & Noting the Problem | |
|---|---|--|--|
| <ul style="list-style-type: none">The Obama administration has suggested that an updated and educational policy and accountability system could be our “moonshot” of the 21st century to ensure that the United State is economically competitive within the international community.⁴Fowler (2009) suggests that the nature of reform will be imposed by the presence of a problem where constituents have divergent values. With the reauthorization of No Child Left Behind (NCLB) currently being discussed at the federal level, the educational accountability system, vertically spanning the federal, state, and local levels, is at a precipice where operational changes need to be considered.⁵One of the educational discussion points being seriously investigated in the educational policy arena is how educators should be objectively evaluated and this data can be used to potentially predict student educational performance. | | <ul style="list-style-type: none">The general consensus within the literature is that teachers play a critical role in the performance of students. However, there are differences in defining and measuring the the constructs associated with what is referred to as “teacher effectiveness”. Teacher effectiveness will be defined here as a “teacher’s ability to improve student learning as measured by summative assessments.⁶There is a paucity of of empirical research reflecting predictive and causal relations between teacher effectiveness and student academic outcomes due to the inherent validity and reliability constraints associated current models associated with measuring teacher effectiveness.⁶The paradox of measuring teaching effectiveness is that student assessment results appear to be the best proxies for measuring teacher effectiveness. Teachers are the most proximal and malleable variable that can be controlled to impact student outcomes. However covariates that are more distal and rigid (e.g., socioeconomic status and family involvement) impact the discussion of this topic between policy actors.¹⁰ | |
| 3 Most Common Teacher Evaluation Models | Overview | Strengths | Weaknesses |
| Value Added Models | Using a predictive model for growth, an individual student’s true summative scores are compared to their predictive scores. If the true scores exceed the predictive scores, then the teacher and their instruction are considered highly effective. The converse would also be true. ⁷ This provides a metric to evaluate teacher effectiveness and medium for noting effective instructional strategies. ³ | <ul style="list-style-type: none">It can be an objective measure removing observer bias (e.g., observational models)Cost-efficient since summative data is already collectedIf combined with other models, VAM can be an objective corroborator for data analysis. | <ul style="list-style-type: none">There are validity concerns in terms of construct validity (e.g., concurrent, predictive, convergent, discriminant; in addition: generalizability, structural, & consequential).⁹<ul style="list-style-type: none">Studies that correlate VAM with teacher qualifications and characteristics produce mixed results (i.e., missing data and nonrandom assignment).VAM assume that standardized test data obtain is valid and reliable (e.g., demographics and vertical scaling). |
| Classroom Observations | This model has been used for decades to capture attributes of teaching and learning. Currently, teacher observations attempt to summarize 4 features: teacher student interactions, classroom management, school community contributions, and subject matter knowledge. ^{6, 8} | <ul style="list-style-type: none">2 instruments that are widely published and used:<ul style="list-style-type: none">Charlotte Danielson’s Enhancing Professional Practice: A Framework for TeachingUniversity of Virginia’s Classroom Assessment Scoring System (CLASS)Flexible for both formative and summative assessments⁸ | <ul style="list-style-type: none">Reliability of the protocol is questioned due to lack of internal consistency of the phenomena attempting to be measured.Validity concerns (e.g., content, predictive, and construct)Paucity of information for use in grades 6-12.Modest effect sizes from research conducted between pre-K through 5th grade (e.g., .2 - .5).Training and administration is cumbersome⁸ |
| Principal Evaluations | One of the most common forms of teacher evaluation. This is where the principal is the primary evaluator of the teacher versus an outside observer. ² | <ul style="list-style-type: none">A principal evaluation can be used for multiple evaluative purposes aligned with teacher effectiveness (e.g., formative and summative assessments of student performance, decisions related to teacher tenure or dismissal, needs for teacher remediation, and formative feedback related to instructional practice)² | <ul style="list-style-type: none">Validity issues associated with defining the constructs being measured.Reliability associated with observer biasInconsistent training opportunities to observe teacher performance in an objective manner which lead to threats to internal validity and reliability.² |
| Other models utilized to capture teacher effectiveness include: <i>Assessing classroom artifacts, evaluating teacher portfolios</i> , having teacher submit a <i>self report</i> , and <i>student submitted teacher evaluations</i> . It is noted that each of these models can provide insight to aspects of a teacher’s performance. However, as measures that can predict student achievement, these models should be considered as mechanisms to provide additive insight to a teacher’s performance and abilities rather than a means to validity and reliability assess and measure teaching effectiveness. | | | |

Recommendation: Comprehensive Measurement Tool

- The inherent challenges associated with measuring teacher effectiveness is that it encompasses multiple constructs nested within latent constructs (e.g., a variable that can’t be measured directly).⁶
- Because of this, a triangulated approach to collecting and interpreting multiple teaching constructs should be used to measure teaching effectiveness. This type of comprehensive evaluation would embody a weighted matrix that would yield a relevant composite score that captures the attributes of effective teaching.⁶

The MET Project

In the fall of 2009, the Bill and Melinda Gates foundation launched the Measures of Effective Teaching (MET) project in order to develop and test multiple measures of teacher effectiveness. The goal of the project is to build fair and reliable systems for teacher observation that can be used for a variety of purposes, including instructor feedback, professional development, and continuous improvement that will address educational accountability requirements. Even through the MET Project recognizes that student achievement gains need to be valid predictors of teacher effectiveness, gain scores alone cannot be the sole determinant of capturing and evaluating teacher effectiveness. **The five areas studied were: (a) student achievement, (b) classroom observations and teacher reflections, (c) teachers’ pedagogical content knowledge, (d) student perceptions of the classroom instructional environment, and (e) teachers’ perceptions of working conditions and instructional support at their schools.**¹

MET Project

Multiple Measures of Teaching

The initial findings of the project suggest a framework that uses multiple measures to capture teaching effectiveness

Testing Validity

To What Extent Do Indicators Predict Outcomes?

TEACHING INDICATORS

from each teacher working with ONE GROUP of students:

- Classroom Observations
- Student Surveys
- Gains on State Tests
- Combination of Indicators

STUDENT OUTCOMES

from same teacher working with ANOTHER GROUP of students:

- Gains on State Tests
- Gains on Supplemental Tests
- Positive Student Feedback

Student Perceptions Matter

| The 7 Cs | Sample Questions | Agreeing with Each Item | At the 25th percentile | At the 75th percentile |
|-------------|---|-------------------------|------------------------|------------------------|
| CARE | My teacher in this class makes me feel that s/he really cares about me. | 40% | 33% | 73% |
| CONTROL | My teacher really tries to understand how students feel about things. | 35% | 33% | 68% |
| CLARIFY | Students in this class treat the teacher with respect. | 33% | 36% | 79% |
| CHALLENGE | Our class stays busy and doesn't waste time. | 50% | 53% | 82% |
| CAPTIVATE | My teacher explains difficult things clearly. | 52% | 56% | 79% |
| CONFIR | In this class, we learn a lot almost every day. | 56% | 33% | 81% |
| CONSOLIDATE | My teacher makes lessons interesting. | 47% | 40% | 76% |
| | Students speak up and share their ideas about class work. | 46% | 58% | 86% |
| | My teacher respects my ideas and suggestions. | 44% | 44% | 76% |
| | The comments that I get on my work in this class help me understand how to improve. | 58% | 44% | 76% |

Survey items are differentiated based on grade level and can be administered online or on paper.

The table above, based on the Tripod survey, shows that students are able to differentiate between teachers and their classroom environments. The Tripod survey identifies seven constructs—the 7 Cs—that are core to a student’s experience in his or her classroom. For example, “Care” refers to the extent to which students report that their teacher cares about them as measured by multiple survey questions. “Control” refers to the extent to which teachers effectively manage student behavior in the classroom.

Contact Information and Acknowledgements

For further information, please contact Franklin W. Bender, University of Oregon: fbender@uoregon.edu.

Methods & Results

- Large scale study randomly assigning 1,300 teachers to classroom from school districts from 6 different states.¹
- 5 Observation Instruments were used (e.g., Framework for Teaching – FTT; Classroom Assessment Scoring System –CLASS; Protocol for Language Arts Teaching Observations – PLATO; Mathematical quality of Instruction – MQI; Uteach Teacher Observation Protocol – UTOP)¹
- 7,491 videos of lessons were scored at least 3 times by raters.¹
- Test scores and student surveys were incorporated from 44,500 students.¹

Teachers with Higher Observation Scores Had Students Who Learned More

Balanced Assessment in Mathematics

Combining Measures Added Predictive Power

Multiple Measures and Value-Added on State Math Test

Multiple Observations Led to Higher Reliability

Each rater is observing a different lesson

NOTES: The number in each circle is the percentage of variance in average FTT scores attributable to teacher effects. The area of the inner circle represents variance in aspects of teachers' practice that is consistent across lessons, while the area of the outer circle adds in what we call "other factors," such as day development and lesson-to-lesson variation. As the number of observations increases, the percentage due to consistent teaching practice increases, while the percentage due to other factors declines. It is anticipated that more observations will lead to higher reliability estimates. These reliability estimates are based on having trained raters, with no personal relationship to teachers, observe digital video. The reliability achieved by school systems could be higher or lower. See Table 11 in the research paper at www.metproject.org for results with the other instruments.

3 Key Take Ways from MET Project

- High-quality classroom observations will require clear standards, certified raters, and multiple observations per teacher. Resources allocated for observation training, purchase of evaluation tools, and continued professional development will be essential.¹
- Combining the 3 approaches (classroom observations, student feedback, and value-added student achievement gains) capitalizes on identifying teacher strengths and offsetting weaknesses.¹
- Combining new approaches to measuring effective teaching (while not yet perfect) significantly outperforms traditional unidimensional measures. By providing more valid and reliable evidence, better educational decisions can be made.¹

References

- Bill and Melinda Gates Foundation, Measures of Effective Teaching (MET) Project. (2012). *Gathering Feedback for Teaching: Combining High Quality Observations with Student Surveys and Achievement Gains*. Retrieved from <http://www.metproject.org>.
- Brandt, C., Mathers, C., Oliva, M., Brown-Sims, M., & Hess, J. (2007). *Examining district guidance to schools on teacher evaluation policies in the Midwest region* (Issues & Answers Report, REL 2007-No. 030). Washington, DC: US Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance.
- Braun, H., Chudowsky, N., & Koenig, J. (Ed.) (2010). *Getting value out of value added: Report of Workshop*. Washington, DC: The National Academy Press.
- Duncan, A. (2009, July 24). Education reform's moonshot. *The Washington Post*. Retrieved from <http://www.washingtonpost.com/wp-dyn/content/article/2009/07/23/AR2009072302634.html>
- Fowler, F. C. (2009). *Policy studies for educational leaders: An introduction* (3rd Ed.). Boston: Pearson.
- Goe, L., Bell, C., & Little, O. (2008). *Approaches to evaluating teacher effectiveness: A research synthesis*. Washington, DC: National Comprehensive Center for Teacher Quality.
- Goldhaber, D., & Anthony, E. (2004). *Can teacher quality be effectively assessed?* Seattle, WA: Center on Reinventing Public Education.
- Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2007). *Classroom assessment scoring system: Observation protocol scoring manual*. Baltimore: Paul H. Brookes.
- Rothstein, J. (2008). Student sorting and bias in value added estimation: Selection on observables and unobservables. Paper presented at the National Conference on Value-Added Modeling, Madison, WI.
- Shulman, L. S. (1988). The paradox of teacher assessment. In Educational Testing Service (Eds.), *New directions for teacher assessment* (pp. 13-27). Princeton, NJ: ETS.

POWER PRESENTATIONS
www.PowerPresentations.com