



Multi-method teacher evaluation for high poverty schools: Observations and self-ratings of instructional and behavioral management

Ryan J. Kettler^{a,*}, Ilona Arnold-Berkovits^a, Linda A. Reddy Rutgers^a, Alexander Kurz^b, Christopher M. Dudek^a, Anh N. Hua^a, Adam Lekwa^a

^a Rutgers University–New Brunswick, United States

^b Arizona State University, United States

ARTICLE INFO

Keywords:

Observations
Rating scales
Student achievement
Teacher evaluation
Validity

ABSTRACT

The investigation examined the degree to which data from a multi-method, multi-informant teacher evaluation project related to growth in student achievement as measured using the Measures of Academic Progress in high poverty charter schools. Investigators used two classroom observational measures (the Framework for Teaching and the Classroom Strategies Assessment System-Observer form) and two teacher self-report measures (the Instructional Learning Opportunities Guidance System and the Classroom Strategies Assessment System-Teacher form) in 15 high poverty charter schools to assess teacher performance as part of a school reform grant funded by the US Department of Education, Teacher Incentive Fund. Correlational analyses were conducted corresponding to each combination of method, content area, and criterion score type, as well as controlling for grade level. Bivariate correlations were modest and different than those found in previous research, and findings from multiple regression analyses favored different scores based on method and content area.

1. Introduction

The need for fair and balanced teacher evaluation practices is a critical issue affecting teachers, evaluators, school districts, and other educational stakeholders. Since the landmark legislation of the No Child Left Behind Act of 2001, the country has moved from reliance on a single student proficiency score to a broader framework incorporating multiple metrics of teaching practice and student growth in achievement to assess teacher effectiveness. Many scholars advocate teacher evaluation approaches should integrate multi-methods (MMs) and sources of data to generate a comprehensive assessment of teacher performance related to student learning, which in turn informs human capital management systems decisions and professional development (PD; Reddy, Kettler, & Kurz, 2015). In 2012, recognition of the value of MM evaluation was seen through the federal appropriations of large school reform grants through the Teacher Incentive Fund (TIF) program of the US Department of Education (ED). The primary purpose of the TIF projects is to develop and implement comprehensive integrated human capital management systems which include strategies to recruit, hire, develop, retain, and reward effective teachers in high poverty settings. Specifically, TIF requires the implementation of a rigorous MM

teacher evaluation employing multiple process and outcomes metrics to determine teacher effectiveness and inform targeted PD efforts. Examining the relations among the scores used collectively to evaluate teachers is critical for implementation of such a system. This demonstration study examined the implementation of measures used in the 2012 TIF-funded, School System Improvement (SSI) Project.

For purposes of the SSI Project, 15 high poverty charter schools in New Jersey began a partnership with researchers from Rutgers University–New Brunswick, and Arizona State University to build a rigorous and innovative teacher evaluation system (TES) that recognizes and rewards effective teaching and leadership practices over the course of a five-year period. Project leaders aligned the TES with the NJ state teacher evaluation policy. The SSI Project team designed the TES using an approach to evaluate teacher effectiveness based on (a) observational data from measures such as the Framework for Teaching (FFT; Danielson, 2013) and the Classroom Strategies Assessment System-Observer Form (CSAS-O; Reddy & Dudek, 2014); (b) teacher self-report measures such as the Instructional Learning Opportunities Guidance System (MyiLOGS; Kurz & Elliott, 2012) and the CSAS-Teacher Form (CSAS-T; Reddy & Dudek, 2014); and (c) student growth data from assessments such as the Measures of Academic Progress (MAP;

* Corresponding author at: Graduate School of Applied and Professional Psychology, Rutgers University, 152 Frelinghuysen Road, Piscataway, NJ 08854, United States.

E-mail address: r.j.kettler@rutgers.edu (R.J. Kettler).

<https://doi.org/10.1016/j.stueduc.2018.08.004>

Received 8 December 2017; Received in revised form 14 August 2018; Accepted 23 August 2018

0191-491X/© 2018 Published by Elsevier Ltd.

Northwest Evaluation Association, 2011). The TES generates scores informing four performance levels, as well as empirically supported and personalized PD. Consistent with TIF program requirements, the SSI Project rewarded teachers for scores in the effective or highly effective performance levels, providing performance-based compensation in the form of stipends.

1.1. Federal policy and teacher evaluation

In recent years, teacher evaluation reform has been a focal point of federal legislation and funding opportunities. The Race to the Top Competition, for example, influenced many states to revamp their TESs (McGuinn, 2012). In 2015, the National Council on Teacher Quality noted 47 states had implemented teacher evaluation policies with classroom observations as part of the evaluation process; 45 states also had policies including student achievement. More recently, the Every Student Succeeds Act of 2015 (ESSA, 2015) appropriated funds for states to develop assessment and evaluation systems with the specific recommendation to use MMs for purposes of determining teacher effectiveness. Under Section 2101, ESSA noted the need for:

“Developing, improving, or providing assistance to local educational agencies to support the design and implementation of teacher, principal, or other school leader evaluation and support systems that are based in part on evidence of student academic achievement, which may include student growth, and shall include MMs of teacher performance and provide clear, timely, and useful feedback to teachers, principals, or other school leaders...” (c,4, B, ii).

Examples of such activities included developing high-quality evaluation tools such as classroom observation rubrics; providing training to coaches, principals, and other school leaders; and developing a system for auditing the quality of the evaluation and support systems. Under Section 2102, ESSA indicated local activities “(ii) shall include MM of teacher performance and provide clear, timely, and useful feedback to teachers, principals, or other school leaders; (b,3, A, ii).” Collectively, Race to the Top and ESSA underscore the importance of MMs for evaluating teachers, providing formative feedback, and connecting data-driven practice assessments to student academic achievement.

While passage of the ESSA (2015) reinforced the importance of states selecting and using MMs to evaluate teachers, schools, and districts, the notion of doing so originated much earlier. Baker (2003) traced this impetus to discussions surrounding the National Council on Education Standards and Testing (NCEST), held in 1991 and 1992. It is desirable to use MMs because each form of assessment has its own strengths and weaknesses, and the weaknesses of one type of assessment may be addressed by the strengths of others. While the rationale for using MMs is convincing, it is important to do so in a logical and evidence-based manner (Chester, 2003). Thus, the current study demonstrates the combined use of two observational measures of teaching performance, and two self-report measures of teaching performance in relation to student achievement.

1.2. Multiple measures in educator evaluation

Baker (2003) summarized six rationales for the use of MMs in evaluation systems. These rationales included using (a) different measures for different purposes, (b) different measures to broaden and deepen inferences about learning, (c) minimalist implementations of MMs (i.e., incorporating noncognitive measures rather than having students take additional tests), (d) different measures to diversify authority (between the state and the district), (e) accommodations and alternatives to create additional measures, and (f) different measures to broaden our definition of educational quality. Baker (2003) asserted the use of MMs is a validity issue, rather than a reliability issue, involving the coverage of a construct rather than the precision of measurement.

Baker concluded by identifying several questions that need to be answered before adopting a system based on MMs, including “How should weighting of different assessments occur?” and “How much redundancy do we want, and how much, if any, can we afford?” (p. 16, Baker, 2003). The current study is designed in part to provide information related to these two questions.

Tyler (2011) identified several considerations for implementing a TES using MMs. The author noted teacher evaluation represents multiple purposes for various stakeholders, including (a) identifying and removing low performing teachers, (b) undergirding performance-based compensations systems, and (c) informing PD for teachers. Tyler indicated evaluation data should emanate from both inputs related to the educational process and outputs of that process. While student gains on achievement tests are the typical outputs, inputs could take a variety of forms including structured observations, teacher-produced artifacts, and PD activities. Of these process variables, classroom observations may be the most direct, and would therefore receive the greatest weights within a MMs-based evaluation system. The author concluded the correlation between observation scores and growth on achievement test scores is important information for the adopted evaluation model.

Steele, Hamilton, and Stecher (2010) reviewed lessons from two states and three districts incorporating multiple measures of student quality. The researchers indicated two major challenges of incorporating MMs are generating valid estimates of teachers’ contributions to growth in student achievement and including teachers of grades and content areas not tested annually. The current study addresses these challenges by examining the predictiveness of teacher process variables to student achievement, informing both the contribution of teacher performance to student performance and the potential for using scores from process variables for situations in which outcome variables are not available. One of the key policy recommendations of Steele et al. (2010) was to create comprehensive TESs based on multiple measures; the developers of the SSI Project followed this recommendation in designing the TES.

Mihaly, McCaffrey, Staiger, and Lockwood (2013) studied several models using data from the *Measures of Effective Teaching* (MET) Project, a partnership involving seven school districts and 21 research entities, using video-based observations. The aim of the large-scale study was to determine whether information from multiple measures (value-added scores, classroom observations, student surveys) supported the existence of a composite estimator of effective teaching; the researchers explored methods of combining data from multiple measures to yield scores representing teacher quality. Findings indicated evidence of a stable component shared by indicators. The component differed greatly across models, with the optimal weights of predictor variables depending greatly on the criterion variables used. In a review of Mihaly et al.’s (2013) study, Rothstein and Mathis (2013) argued evidence did not support the existence of a general factor for teaching. The theorists concluded data from the MET Project do not indicate which combination of measures is best to use for teacher evaluation.

The SSI Project’s TES involves multiple direct observations and teacher self-report measures. Classroom observations have long been considered the gold standard in terms of capturing behavioral data, such as teachers’ use of instructional techniques, due to the objectivity with which a well-designed observational system can reflect behavior during a set period. However, despite widespread use in TESs, principals’ use of classroom observations of teacher practices has been questioned. For example, scholars have noted issues with the reliability of principals’ observational reports of teachers (e.g., Kauchak, Peterson, & Driscoll, 1985; Porter, Youngs, & Odden, 2001; Toch & Rothman, 2008; Weisberg, Sexton, Mulhern, & Keeling, 2009) and have indicated the accuracy of principal assessments remains an assumption instead of an empirically proven component to teacher evaluation (Medley & Coker, 1987). Furthermore, classroom observations are time and cost intensive in that they require a school principal to attend a classroom for a set amount of time (e.g., one class period) on multiple occasions

Download English Version:

<https://daneshyari.com/en/article/11002199>

Download Persian Version:

<https://daneshyari.com/article/11002199>

[Daneshyari.com](https://daneshyari.com)