Classroom observation for evaluating and improving teaching: An international perspective

Felipe Martinez, Sandy Taut*, Kevin Schaff

*Corresponding author at: Pontificia Universidad Católica de Chile, Escuela de Psicología, Centro de Medición MIDE UC, Avda Vicuña Mackenna 4860, Macul, Santiago, Chile.

E-mail address: staut@uclal.edu (S. Taut).

Contents lists available at ScienceDirect

Studies in Educational Evaluation

journal homepage: www.elsevier.com/stueduc

ARTICLE INFO

Article history:
Received 21 July 2015
Received in revised form 14 March 2016
Accepted 18 March 2016
Available online 15 April 2016

Keywords:
Classroom observation
Teacher appraisal
Teacher development
Instructional quality
Professional development
Evaluation systems

ABSTRACT

Teacher evaluation and development policies around the world are undergoing significant reform. Classroom observation often carries a considerable weight in teacher appraisal and improvement systems, and provides the critical formative anchor informing professional development. This study examined a purposively selected sample of sixteen classroom observation systems in six countries, including high performing Singapore and Japan, regional exemplar Chile, the three largest school districts in the United States, and other interesting examples in Australia, Germany and the United States to add diversity to the sample. The study offers an analytic framework for understanding classroom observation systems across contexts, distinguishing conceptual, methodological and policy aspects that shape these systems. The sixteen systems were remarkably consistent in their stated overall purposes, but there was variation in terms of how they operationalized good teaching, the degree of standardization of the observation process, emphasis on validation, and information uses. The paper describes and discusses these characteristics in order to help researchers and policymakers reflect on the available options and take more informed decisions in designing classroom observation for evaluating and improving teaching. © 2016 Elsevier Ltd. All rights reserved.

1. Introduction

Teacher evaluation and development policies are in the midst of substantial reform in education systems around the world. Such calls for reform tend to reflect continuing concerns about shortcomings in educational quality and equity, reflected in state, national, or international student achievement testing results, either in absolute terms or in relation to peer districts, states, or countries (Peuer, 2012). International reviews also reflect growing interest in revamping teacher evaluation and development at local, state, and national levels, with a focus on supporting valid inferences about teachers and providing useful feedback to help them improve their practices in the classroom (OECD, 2013a, 2013b; Santiago & Benavides, 2009). In the United States, in particular, growing interest in teacher evaluation reflects policy shifts away from highly qualified teachers to a notion of highly effective teachers (Goe, Bell, & Little, 2008). In the context of federal guidelines that require that districts and states develop comprehensive approaches for evaluating teacher effectiveness, a great deal of attention and discussion among educators, statisticians, and policymakers has focused on the use of aggregates of student achievement to evaluate teachers, typically through so-called Value Added Models, or VAMs (for reviews see e.g. Baker et al., 2010; Braun, 2005).

While the attention paid to VAMs in the research literature and the press could suggest otherwise, however, student achievement is rarely the driving indicator for teacher evaluation. In most cases, measures of teacher practice based on direct observation of teachers in classrooms receive an equal or larger weight. A common assumption to most teacher evaluation systems around the world is that classroom practice is the key mediator between education policies and student outcomes (Schleicher, 2011), and classroom observation remains the method of choice (a de facto gold standard) for gaining systematic insight into these practices in their natural setting (Kennedy, 2010; Stallings, 1977). In the context of teacher evaluation policies, observation-based assessments of teachers in the classroom are seen as key both for understanding the mechanisms linking classroom processes and desired improvements in student outcomes, and for informing formative and developmental feedback to guide teacher improvement efforts.

http://dx.doi.org/10.1016/j.stueduc.2016.03.002
091-491X© 2016 Elsevier Ltd. All rights reserved.
(Kane & Staiger, 2012; OECD, 2013a, 2013b; Reform Support Network, 2012).

Yet, even as teacher development and evaluation policies receive unprecedented attention around the world, systematic study of classroom observation in this context has been scarce. Direct observation of the work of teachers in the classroom has been a staple of research on teaching for nearly a century (see Medley & Mitzel, 1963). However, educators and policymakers need to understand how evaluative contexts influence the information collected through classroom observation, and how it is used to inform inferences about, and feedback to teachers. With this paper we start to fill this gap by investigating how a sample of notable and varied regional and national education systems around the world apply classroom observation to inform teacher evaluation and professional development. Specifically, by adopting a comparative lens for data collection and analysis, we seek to (a) outline an analytic framework for classroom observation systems based on the conceptual, methodological, and contextual commonalities and differences to help bring cohesion to the research literature, and (b) to support policymakers and system designers in systematically considering key questions involved in the design of classroom observation systems.

2. Theoretical background

Large-scale policy initiatives on teacher evaluation and professional development have traditionally been informed by two different models. The standards-based model emphasizes explicit frameworks to model quality instruction and classroom practice (Peterson, 2000), while the outcome-based model privileges productivity in terms of student achievement and other relevant outcomes (Kennedy, 2010). These two models have started to converge in recent years, as a new crop of systems of teacher evaluation and development combines an emphasis on student achievement with explicit and detailed models of instructional practice (Kane & Staiger, 2012). Importantly, despite the growing international consensus around the importance of evaluating and improving teaching, the specific characteristics of a system depend on a series of factors that include assumptions about the goals and purposes, stakes and incentives attached to the evaluation results, frameworks or models that define the domains and criteria to be assessed, and methods used to collect and analyze information about these domains (OECD, 2013a, 2013b).

A common feature in recent efforts to revamp teacher evaluation across countries and contexts is a focus on formative uses of information. For example, the federal Race to the Top legislation in the U.S. calls for data systems “that inform teachers and principals about how they can improve instruction” (U.S. Department of Education, 2009), while in Australia the national agreement on teacher performance seeks to promote professional conversations that improve teaching (Australian Institute for Teaching and School Leadership [AITSL], 2016). Similarly, Chilean legislation sets up a formative teacher evaluation system focused on “improving the pedagogical work of educators and promoting their continuous professional development” (Law, 2004, Law No. 19.961). This formative focus has renewed interest in classroom observation as a method for collecting information to support improvement efforts regarding teaching quality.

Classroom observation is a powerful tool that offers an unobstructed view of classroom practice (Millman & Darling-Hammond, 1990) and allows us to understand how teachers teach within a realistic context (Putnam & Barko, 2000). Although observation protocols can be found in the research literature since at least the 1920s (Stallings, 1977), classroom observation enjoyed widespread popularity as a method of large scale data collection in the 1950s and 1970s, when studies investigating the validity of observation protocols became a staple of education research (see e.g. Brophy & Good, 1986; Medley & Mitzel, 1963; Shavelson, Webb, & Burstein, 1986; Stallings, 1977). Validation research is again being conducted, as a new generation of researchers works to develop observation protocols and instruments to help us understand the relationship between classroom practices, teacher effectiveness, and student achievement (Bell et al., 2012; Hill, Charalambous, & Kraft, 2012; Mihaly, McCaffrey, Staiger, & Lockwood, 2013).

Growing interest in classroom observation also crucially reflects the widening reach of teacher evaluation policies around the world (Bruns & Luque, 2014). Despite the ubiquity of efforts to use student achievement to evaluate teachers, in practice most teachers in the United States are evaluated through procedures that rely heavily on classroom observation (Kane & Staiger, 2012; Loup, Garland, Ellett, & Rugutt, 1996). Indeed, all states granted funding under the new Race To The Top legislation in the United States included a new or redesigned classroom observation component for teacher evaluation (Reform Support Network, 2012), and a similar trend is evident in international reviews (see e.g., OECD, 2013a, 2013b; Santiago & Benavides, 2009). In the 2007 Teaching and Learning International Survey (TALIS), for example, over 70% of teachers in participant countries reported that classroom observation was an important component in the process of evaluation and feedback (OECD, 2009). This is not unexpected because classroom observation enjoys considerable face validity among educators and policymakers, and is seen as the key source of information supporting teacher formative evaluation and feedback (Danielson, 1996; Goe et al., 2008; Pianta, La Paro, & Hamre, 2007; see also Chait, 2010; Weisberg, Sexton, Mulhern, & Keelng, 2009), teacher self-reflection (Richards, 1991), and the diagnosis of areas of strength and those in need of improvement (Protheroe, 2002).

As with teacher evaluation more generally, however, what it means to do classroom observation for teacher evaluation can differ across systems and contexts. Below we outline an analytic framework for systems of classroom observation that involves three main dimensions, related to their basic conceptualization of instructional practice and teacher effectiveness (i.e., conceptual issues), the sources of evidence and methods used to gather information about these constructs (i.e., methodological issues), and policy context, processes, and decisions that shape the evaluation (i.e., policy issues).

2.1. Conceptual issues

The first step in designing an observation system is to decide upon the theoretical or conceptual underpinnings that will provide the basis for understanding, describing, and assessing teacher practice. Standards of effective teaching have become widespread in the past decade (e.g., Campbell, Kyriakides, Muijs, & Robinson, 2004; Danielson, 1996) and many education systems are standardizing their approaches to classroom observation to align them to such frameworks. The expectation is that clear and explicit teaching standards, and matching observation rubrics, will provide useful guidance to teachers and administrators for understanding and promoting high-quality instruction (Darling-Hammond, 2013; Ingvason & Rowe, 2008). One example is the Framework for Teaching (Danielson, 1996, 2011) adapted in a number of school districts in the United States (e.g., New York, Los Angeles, Cincinnati) and abroad,—e.g., as “Marco para la Buena Enseñanza” in Chile (Ministry of Education, 2004). Within each of these sets of standards, protocols and rubrics are typically developed targeting a narrower set of dimensions, deemed most amenable to direct observation in classrooms. These protocols may characterize teacher performance in general pedagogical aspects of instruction.
دانلود مقاله

http://daneshyari.com/article/372569