Rethinking Teacher Evaluation in Chicago

Lessons Learned from Classroom Observations, Principal-Teacher Conferences, and District Implementation

Authors: Lauren Sartain, Sara Ray Stoelinga, and Eric R. Brown; with: Stuart Luppescu, Kavita Kapadia Matsko, Frances K. Miller, Claire E. Durwood, Jennie Y. Jiang, and Danielle Glazer

CONSORTIUM ON
CHICAGO SCHOOL RESEARCH
AT THE UNIVERSITY OF CHICAGO
URBAN EDUCATION INSTITUTE

Research Report
November 2011
Executive Summary

In 2008, Chicago Public Schools (CPS) launched the Excellence in Teaching Pilot, an effort to revamp how teachers are evaluated and how they receive feedback on their performance. The pilot was at the forefront of a national movement to redesign teacher evaluation. The work in Chicago and across the country to improve evaluation was motivated by two main factors. First, evaluation systems were failing to give teachers either meaningful feedback on their instructional practices or guidance about what is expected of them in the classroom. Second, traditional teacher evaluation systems were not differentiating among the best teachers, good teachers, and poor teachers. Chicago, for example, relied on a system that both teachers and principals viewed as arbitrary and unfair.1 Moreover, the system identified 93 percent of teachers as either Superior or Excellent—at the same time that 66 percent of CPS schools were failing to meet state standards, suggesting a major disconnect between classroom results and classroom evaluations.

This report summarizes findings from a two-year study of Chicago’s Excellence in Teaching Pilot, which was designed to drive instructional improvement by providing teachers with evidence-based feedback on their strengths and weaknesses. The pilot consisted of training and support for principals and teachers, principal observations of teaching practice conducted twice a year using the Charlotte Danielson Framework for Teaching, and conferences between the principal and the teacher to discuss evaluation results and teaching practice.

Although the findings from this report focus on a specific pilot in a specific city, they have broad implications for districts and states nationwide that are working to design and develop evaluation systems that rely on classroom observations to differentiate among teachers and drive instructional
improvement. Overall, we found that the Excellence in Teaching Pilot was an improvement on the old evaluation system and worked as it was designed and intended, introducing an evidence-based observation approach to evaluating teachers and creating a shared definition of effective teaching. At the same time, the new system faced a number of challenges, including weak instructional coaching skills and lack of buy-in among some principals. Specific findings include:

- **The classroom observation ratings were valid measures of teaching practice;** that is, students showed the greatest growth in test scores in the classrooms where teachers received the highest ratings on the Danielson Framework, and students showed the least growth in test scores in classrooms where teachers received the lowest ratings.

- **The classroom observation ratings were reliable measures of teaching practice;** that is, principals and trained observers who watched the same lesson consistently gave the teacher the same ratings; however, 11 percent of principals consistently gave lower ratings than the observers and 17 percent of principals consistently gave higher ratings than the observers.

- **Principals and teachers said that conferences were more reflective and objective than in the past and were focused on instructional practice and improvement.** However, many principals lack the instructional coaching skills required to have deep discussions about teaching practice.

- **Over half of principals were highly engaged in the new evaluation system.** Principals who were not engaged in the new evaluation system tended to say that it was too labor intensive given the numerous district initiatives being simultaneously implemented in their schools.

This report is divided into five chapters:

**CHAPTER ONE** presents the national impetus for revitalizing teacher evaluation practices and details the teacher evaluation pilot in Chicago and the Charlotte Danielson Framework for Teaching. This chapter provides information that is particularly useful for those interested in how the local and national discussion around teacher evaluation has played out in Chicago, as well as the details of the research behind this report.

**CHAPTER TWO** investigates the relationship between teacher classroom observation ratings and value-added measures, providing information that might be particularly useful for those working to build an observation system that is linked to learning gains.

**CHAPTER THREE** describes how principals rated teaching practice across schools when compared to another observer. This chapter provides information for those working to build an observation system where teachers reliably receive the same rating for exhibiting the same quality of instruction.

**CHAPTER FOUR** highlights the conference component of the evaluation system—principal and teacher reports on the quality of the conversations, as well as an investigation of the nature of these conversations. This chapter provides information for those interested in how a comprehensive evaluation system can be designed to drive instructional improvement.

**CHAPTER FIVE** follows a different format than the previous chapters and is meant to serve as a design guide for districts and unions that are revitalizing teacher evaluation systems. Instead of providing explicit answers to design questions, the chapter draws attention to key design and logistical considerations and brings evidence from the Chicago pilot to bear on these issues. The goal is to give policymakers and practitioners empirical evidence that they can use to arrive at informed solutions for their own states and districts. Specifically, we discuss criteria for assigning formal evaluation ratings, decisions about classroom observation logistics, training for principals and teachers, principal engagement, and evaluator feedback and accountability. For each of these topics, we provide evidence from the Chicago pilot. We also list some questions for stakeholders to consider when thinking about each of these issues.