

Abstract Title Page

Title:

Teacher Evaluation in Practice: Exploring Relationships Between School Characteristics & Evaluation Scores

Authors and Affiliations:

Susan E. Spote, University of Chicago Consortium on Chicago School Research
Jennie Y. Jiang, University of Chicago Consortium on Chicago School Research
Stuart Luppescu, University of Chicago Consortium on Chicago School Research

Abstract Body

Problem / Background / Context:

Decades of research evidence have consistently suggested teachers are the most important in-school factor related to student learning and achievement. Being taught by an effective teacher has important consequences for students' academic outcomes as well as longer-term impacts on postsecondary success and lifetime earnings (Aaronson, Barrow, & Sander, 2007; Chetty, Friedman, & Rockoff, 2011; Goldhaber, 2002). Yet how to measure effective teaching, how to develop effective teachers, and how to ensure that all students have access to highly effective teaching continue to be some of the most persistent challenges facing local, state, and federal education policy makers.

Federal policy has attempted to address these issues. The No Child Left Behind (NCLB) act of 2001 required states to ensure teachers met minimum certification standards. More recently, federal Race to the Top funding incentivized states to overhaul their teacher evaluation systems, with the hope better teacher evaluation policies and systems would be a primary vehicle for improving teaching and learning across schools. That federal focus continues as states have been given opportunities to pursue waivers exempting them from some components of NCLB; in exchange for flexibility in other areas, they are required to describe their evaluation systems and commit to timelines for evaluation system design and implementation.

Response to these efforts has been widespread and in the last decade, teacher evaluation has taken center stage in policy reform efforts to improve teaching at all schools. Although states and districts vary considerably in the measures they use and the weights those measures are given, most combine observations of teacher practice with measures of student growth. Furthermore, they share a common goal of improving teacher effectiveness through two key levers: (1) developing teachers' instructional skills through focused feedback and professional development and (2) holding teachers accountable by incorporating evaluation measures into personnel decisions such as tenure and dismissal.

New teacher evaluation systems are providing a proliferation of new data on teachers that is intended to be used for both accountability and to support teachers in adjusting and improving their instructional practice. In Chicago, over the course of only a few years, district leaders and teachers have moved from an annual checklist conveying essentially no data on teacher performance to multiple classroom observations and measures of student growth that generate detailed reports with multiple pages of ratings. The new systems provide not only additional information on an individual teacher's practice, but also on the patterns and characteristics of the district's overall teaching workforce: *Where are the highly rated teachers? Where are the low-rated teachers? And are there particular teacher or school characteristics related to high and low ratings?* Examining the distribution of teachers with high or low evaluation scores may give insight into how teachers are deployed, the nature of instruction received by students, and whether the measures may be reflecting contextual factors such as school or student characteristics. Examining whether evaluation scores are related to teacher experience and credentials may better inform school staffing decisions. Examining whether evaluation scores are related to race and gender may raise questions about how teachers are systematically assigned to schools and classrooms and about the validity and reliability of the ratings or the rating

instrument. If implemented well, new evaluations can provide more meaningful information to district and school leaders to better direct resources for support and professional development and to inform personnel decisions. With this wealth of data, we can also for the first time ask and begin to answer important questions about the district's teaching workforce and how it is distributed across schools.

Purpose / Objective / Research Question / Focus of Research

We seek to share findings from the following research questions about Chicago's new teacher evaluation system:

- *What is the distribution of observation and value-added ratings across schools?*
- *To what extent are evaluation scores related to school characteristics such as school poverty level, racial composition, measures of culture and climate? Are these relationships different for value-added and observation scores?*
- *Are evaluation scores related to teacher characteristics such as experience or certification?*

In addition to these questions we will share insights into our research-practice partnerships with both the district and the Chicago Teachers Union.

Improvement Initiative / Intervention / Program / Practice:

CPS' new teacher evaluation system —Recognizing Educators Advancing Chicago's Students (REACH)— began district-wide implementation in the 2012-13 school year. REACH seeks to provide a measure of individual teacher effectiveness that can meet the district's dual needs of supporting instructional improvement and differentiating teacher performance. It incorporates teacher performance ratings based on multiple classroom observations together with student growth measured on two different types of assessments. The main components of REACH include multiple classroom observations using a modified version of the Charlotte Danielson Framework for Teaching, required feedback after each observation, and the inclusion of two different measures of student growth (see Appendix B Figure 1 and 2 for more information about REACH).

Setting:

The research in this proposal is conducted in Chicago Public Schools.

Population / Participants / Subjects:

In school years 2012-13 and 2013-14, CPS employed approximately 23,000 teachers each year. Five hundred twenty-six schools are covered by this initiative, which have an enrollment of about 400,000 students each year (see Appendix B Table 2 for more details). Chicago students are likely to be from low-income families (87 percent), and 42 percent are African American and 44 percent are Latino.

Research Design: We use descriptive analysis to explore differences between teacher evaluation scores across schools with varying concentrations of economically disadvantaged students and minority students.

To explore the relationship between school characteristics and teacher evaluation scores we estimate variants of the following two-level hierarchical linear model (HLM).

$$\text{Level 1 (teachers): } Eval_{js} = \beta_{0s} + \beta_1 X_{1js} + \dots + \beta_k X_{kjs} + r_{js}$$

$$\text{(Level 2 (schools): } \beta_{0s} = \gamma_{00} + \gamma_{01} W_{1s} + \dots + \gamma_{0l} W_{ls} + \mu_{0s}$$

The subscripts denote teacher j in school s . The outcome, $Eval$, is either the teacher's observation or value-added score. The X_{kjs} variables denote teacher-level characteristics, including tenure status, level of education, certification status, and years of experience teaching in CPS. The W_{ls} variables are school-level characteristics, and the hierarchical nature of the model accounts for the fact that teachers are nested within schools.

Data Collection and Analysis:

Data for this presentation includes CPS personnel and administrative data from the 2013-14 school year, and survey data from the 2013-14 school year. Teacher and administrator personnel data includes individual-level data about tenure status, years of experience in the district, demographic information as well as evaluation data such as ratings and value-added scores (Please see Appendix B Table 2 for an overview of evaluation data). Observation score analyses include non-tenured teachers and tenured teachers with observation ratings from at least two observations during the 2013-14 school year. The analyses only include teachers who were rated using the CPS Framework for Teaching; librarians, counselors and other education support specialists rated on a different framework were excluded from analyses. Observation scores are the overall professional practice scores for teachers calculated using a weighted average of component ratings. Analyses of value-added scores include only elementary teachers who received individual value-added scores. For ease of comparison, results from both value-added and observation scores in this report were translated to the 100 to 400 scale utilized by CPS for overall professional practice and student growth scores.

For school-level characteristics such as the concentration of poverty, previous achievement and percentage of minority students, we utilized CPS administrative student data aggregated to the school level, including students' race, neighborhood poverty level and neighborhood socioeconomic status, and free/reduced price lunch status.

Measures of School Climate and Culture

Our measures of school climate and culture are derived from teacher and student perceptions of c UChicago CCSR has been partnering with CPS to survey all students in grades 6-12 and all teachers across the district since the early 1990s. This survey, entitled My Voice, My School, was administered annually from 2011 through 2015 and every other year prior to that. Sets of questions were combined into measures of general concepts using Rasch analysis. These include measures of leadership - instructional leadership, teacher-principal trust, teacher influence, and program coherence—and measures describing professional community—collective responsibility, teacher-teacher trust, school commitment, and quality professional development. Survey scales measuring these constructs have been well established over time in CPS, as has the relationship between these constructs and multiple components of school improvement (Bryk et al., 2010). Individual-level reliabilities on these measures ranged from 0.79 to 0.91.

Findings / Outcomes:

Findings are described below:

- **Schools serving disadvantaged students have a disproportionate share of the lowest-rated teachers.** Observation scores of teachers teaching in high poverty schools are substantially lower than the observation scores of teachers in lower poverty schools. There are smaller between-school differences in their teachers' value-added scores.
- **Teachers in schools with stronger organizational climates have higher evaluation scores.** Controlling for school-level characteristics such as poverty and achievement, teachers in schools with better professional climate tend to have higher value-added and observation scores.
- **There are some differences in teachers' evaluation scores depending on experience and credentials.** Teachers with more experience have higher scores on both value-added and observation measures than new teachers. Differences between teachers with National Board Certification and/or advanced degrees compared to those without those credentials were found only on observation scores.
- **Minority teachers have lower observation scores than white teachers but their value-added scores are not significantly different; male teachers have lower observation and value-added scores than female teachers.** Male teachers scored 12 points lower on observations and about four points lower on value-added than their female counterparts with similar levels of experience teaching in similar schools. On average, African American teachers scored about 10 points lower, and Hispanic and other minority teachers scored about seven points lower than white teachers with similar levels of experience teaching in similar schools. However, there were no significant differences by race/ethnicity in either reading or math value-added scores.

Conclusions:

The new evaluation system provides the opportunity to assess teaching with information that is believed to better measure instructional quality than the data available in the past. It is now possible to gauge the degree to which different groups of students in Chicago have access to high- and low-quality teaching using actual metrics of teaching practice, rather than proxies for teaching quality derived from teacher qualifications. This information could be used to set priorities for teacher recruitment, assignment and support to schools that are most in need of better teachers. While the new measures are likely to improve the identification of schools with more effective *teaching* practices, there are still questions about whether they can fairly assess *teacher* quality. That is, it is not clear whether all teachers have an equal chance of receiving strong ratings, given the skills they bring to bear to the job. The teaching that occurs in a classroom comes from the interaction of the teacher with her students within a school context (citation). Thus, there is a risk that ratings of teacher effectiveness depend on factors other than the teacher herself, and that teachers who work in the contexts that are most difficult for high ratings will systematically be rated lower. This could provide perverse incentives for teachers to avoid teaching in schools where high quality teachers are most needed, and to penalize those teachers who do decide to work in those contexts. Furthermore, if particular types of teachers (e.g., minority teachers, male teachers) are more likely to work in contexts where it is difficult to get good ratings, the composition of the workforce could be affected by the evaluation system itself.

Appendices

Appendix A. References

- Aaronson, D., Barrow, L., & Sander, W. (2007). Teachers and student achievement in the Chicago public high schools. *Journal of Labor Economics*, 25(1), 95-135.
- Bryk, A. S., Sebring, P. B., Allensworth, E., Luppescu, S., & Easton, J. Q. (2010). Organizing schools for improvement: Lessons from Chicago. *The University of Chicago*.
- Chetty, R., Friedman, J. N., & Rockoff, J. E. (2011). The long-term impacts of teachers: Teacher value-added and student outcomes in adulthood. *National Bureau of Economic Research*.
- Goldhaber, D. (2002). The mystery of good teaching. *Education Next*, 2(1), 50-55.

Appendix B. Tables and Figures

Table 1: Number of teachers included in analyses of REACH evaluation data

	Observation Scores	Individual VA Scores
Elementary Non-Tenured Teachers	3,271	1,147
Elementary Tenured Teachers	8,974	3,577
High School Non-Tenured Teachers	1,134	
High School Tenured Teachers	3,532	

Figure 2: The CPS Framework for Teaching

<h1 style="margin: 0;">The CPS Framework for Teaching</h1>	
<p style="margin: 0;">Adapted from the <i>Danielson Framework for Teaching</i> and Approved by Charlotte Danielson</p>	
<p style="text-align: center;">Domain 1: Planning and Preparation</p> <p>a. Demonstrating Knowledge of Content and Pedagogy Knowledge of Content Standards Within and Across Grade Levels Knowledge of Disciplinary Literacy Knowledge of Prerequisite Relationships Knowledge of Content-Related Pedagogy</p> <p>b. Demonstrating Knowledge of Students Knowledge of Child and Adolescent Development Knowledge of the Learning Process Knowledge of Students’ Skills, Knowledge, and Language Proficiency Knowledge of Students’ Interests and Cultural Heritage Knowledge of Students’ Special Needs and Appropriate Accommodations/Modifications</p> <p>c. Selecting Instructional Outcomes Sequence and Alignment Clarity Balance</p> <p>d. Designing Coherent Instruction Unit/Lesson Design that Incorporates Knowledge of Students and Student Needs Unit/Lesson Alignment of Standards-Based Objectives, Assessments, and Learning Tasks Use of a Variety of Complex Texts, Materials and Resources, including Technology Instructional Groups Access for Diverse Learners</p> <p>e. Designing Student Assessment Congruence with Standards-Based Learning Objectives Levels of Performance and Standards Design of Formative Assessments Use for Planning</p>	<p style="text-align: center;">Domain 2: The Classroom Environment</p> <p>a. Creating an Environment of Respect and Rapport Teacher Interaction with Students, including both Words and Actions Student Interactions with One Another, including both Words and Actions</p> <p>b. Establishing a Culture for Learning Importance of Learning Expectations for Learning and Achievement Student Ownership of Learning</p> <p>c. Managing Classroom Procedures Management of Instructional Groups Management of Transitions Management of Materials and Supplies Performance of Non-Instructional Duties Direction of Volunteers and Paraprofessionals</p> <p>d. Managing Student Behavior Expectations and Norms Monitoring of Student Behavior Fostering Positive Student Behavior Response to Student Behavior</p>
<p style="text-align: center;">Domain 4: Professional Responsibilities</p> <p>a. Reflecting on Teaching and Learning Effectiveness Use in Future Teaching</p> <p>b. Maintaining Accurate Records Student Completion of Assignments Student Progress in Learning Non-Instructional Records</p> <p>c. Communicating with Families Information and Updates about Grade Level Expectations and Student Progress Engagement of Families and Guardians as Partners in the Instructional Program Response to Families Cultural Appropriateness</p> <p>d. Growing and Developing Professionally Enhancement of Content Knowledge and Pedagogical Skill Collaboration and Professional Inquiry to Advance Student Learning Participation in School Leadership Team and/or Teacher Teams Incorporation of Feedback</p> <p>e. Demonstrating Professionalism Integrity and Ethical Conduct Commitment to College and Career Readiness Advocacy Decision-Making Compliance with School and District Regulations</p>	<p style="text-align: center;">Domain 3: Instruction</p> <p>a. Communicating with Students Standards-Based Learning Objectives Directions for Activities Content Delivery and Clarity Use of Oral and Written Language</p> <p>b. Using Questioning and Discussion Techniques Use of Low- and High-Level Questioning Discussion Techniques Student Participation and Explanation of Thinking</p> <p>c. Engaging Students in Learning Standards-Based Objectives and Task Complexity Access to Suitable and Engaging Texts Structure, Pacing and Grouping</p> <p>d. Using Assessment in Instruction Assessment Performance Levels Monitoring of Student Learning with Checks for Understanding Student Self-Assessment and Monitoring of Progress</p> <p>e. Demonstrating Flexibility and Responsiveness Lesson Adjustment Response to Student Needs Persistence Intervention and Enrichment</p>

Table 2: Mean observation and value-added ratings by school-level poverty

Free/Reduced Price Lunch Group	Observation Scores		VA READ MULTI		VA MATH MULTI	
	Mean Observation Score	After Controlling for Tenure, Advanced Degree, NBCT	Mean Observation Score	After Controlling for Tenure, Advanced Degree, NBCT	Mean Observation Score	After Controlling for Tenure, Advanced Degree, NBCT
(Lowest Poverty) 1	332 (42)	331	256 (34)	256	251 (41)	251
2	312 (49)	308	254 (41)	252	251 (47)	252
3	312 (48)	305	256 (45)	258	255 (51)	256
4	304 (48)	298	247 (48)	250	248 (49)	251
(Highest Poverty) 5	289 (44)	288	246 (51)	247	248 (54)	249

