

Rethinking Teacher Evaluation: Findings from the First Year of the Excellence in Teaching Pilot in Chicago Public Schools

There is a growing consensus that the way most districts and schools evaluate teachers fails to improve student learning or teacher practice. In a recent speech, American Federation of Teachers President Randi Weingarten acknowledged that “with rare exceptions, teacher evaluation procedures are broken—cursory, perfunctory, superficial and inconsistent.”

Policymakers and others have begun calling on districts to replace these flawed evaluation systems, which generally fail to distinguish between strong and weak instruction or provide teachers with meaningful feedback. This push has generated new demand for a set of clearly delineated standards that can be used to judge whether effective teaching is taking place in the classroom.

The Charlotte Danielson Framework for Teaching is perhaps the most well-known example of an observational tool that attempts to outline the main components of effective teaching. A team of researchers from the Consortium on Chicago School Research (CCSR) at the University of Chicago is studying the implementation of the Danielson Framework in Chicago Public Schools and providing real-time, objective feedback to the district. This brief summarizes the key findings and implications from the study’s first year.

The full report is available at http://ccsr.uchicago.edu/publications/Joyce_TE_yr1_finaldoc.pdf.

Findings

Overall, principals and trained experts use the rating scale consistently. To understand the reliability of the Framework, principals and highly trained external observers conducted simultaneous classroom observations but assigned Framework ratings independently. Looking across all the Framework ratings, there is no significant difference between principals and observers. In addition, principals and observers use the rating scale the same way from one observation to the next. However, there are some individual differences in rater severity—both among principals and observers.

Principals had no trouble identifying unsatisfactory teaching practices. However, when using the high end of the scale, principals inflated their ratings. Principals and external observers agreed about unsatisfactory practice. However, across all Framework components, principals were much more likely than external observers to identify instruction as distinguished.

More teachers were identified as low-performing under the new evaluation system. In previous years, only 0.3 percent of teachers in CPS had been rated as unsatisfactory. However, 8 percent of teachers in this sample received at least one unsatisfactory rating on the Framework.

Just over half of principals were highly enthusiastic about the evaluation process. These principals had positive attitudes about the Framework and their conferences with teachers, perceived teacher buy-in as high, and said they saw changes in instructional practice stem from the evaluation system. A little less than half of the principals were characterized by mixed to mostly negative attitudes about both the Framework and the conferences. These principals generally said that they were “already doing” evaluation in the “right way” and were more likely to suggest that they “just knew” if teachers were good or bad.

Implications

The Danielson Framework has potential for improving teacher evaluation systems. Our study indicates that the Framework is a reliable tool for identifying low-quality teaching, suggesting it is an appropriate tool for fairly identifying teachers in need of supports or sanctions. In addition, principals were generally positive about using the Danielson Framework. Principal and teacher buy-in is critical for the success of any initiative. This is especially true for efforts aiming to identify low-quality instruction, remove ineffective teachers, make more informed decisions about staffing schools, and ultimately to improve teaching and thereby student learning.

To realize the Danielson Framework’s potential as an evaluation tool, ongoing training and support for principals is necessary. CPS provided high-quality, ongoing professional development and support for principals in the first year of the pilot; yet, principals still struggled to rate some areas of instruction consistently. Even with high levels of training and support, there will still be challenges when using a tool like the Danielson Framework for teacher evaluation. Because evaluating instruction is complex, continued training and meaningful supports are vital to ensure evaluation tools are fair and useful. If scale up to a larger number of schools does not include training and support that is intensive and ongoing, there are likely to be problematic inconsistencies in the use of the Framework by principals.

There may be challenges in using observational tools for high-stakes decisions. The consequences of inconsistent application of the Danielson Framework become clear when we discuss using ratings for evaluation purposes. For instance, a principal who is a severe rater may have detrimental effects on the careers of borderline teachers in that school. On the other hand, lenient principals may keep teachers who should otherwise be removed due to low performance.

Successful implementation of a rigorous evaluation system requires changing the way practitioners and district leaders think about teacher evaluation. While introducing a high-quality teacher evaluation tool is an important step in revamping evaluation practices, changing the evaluation process also requires a long-term shift in the way people think about teacher evaluation. Some CPS principals revealed attitudes and assumptions about evaluation (for instance “just knowing” if a teacher is good) that need to be addressed if teacher evaluation practices are to improve. Truly transforming teacher evaluation practices relies upon finding ways to shift perceptions among principals who do not see the value in deeper evaluation practices.

Conclusion

It is important to note that our analysis and findings come very early in the implementation of the pilot, which continues to grow. Nevertheless, our preliminary analysis suggests areas of particular promise for states and districts contemplating a major overhaul of their evaluation systems. In order to improve direct evaluation, schools and districts need frameworks that are both reliable and acceptable to major stakeholders. In the Chicago pilot, strong principal buy-in and the overall consistency of ratings from principals and trained observers suggest the Danielson framework, as implemented in CPS, may meet these criteria.

In spring 2011, we plan to release another policy brief focused on the relationship between Framework ratings and student outcomes—findings that should advance our understanding of the link between academic achievement (student outputs) and instructional practice (teacher inputs).