Abstract Title Page

Title: Can Principals Promote Teacher Development as Evaluators?

Authors and Affiliations:

Matthew A. Kraft
Brown University

Allison Gilmour
Vanderbilt University

Abstract Body

Background / Context:

District- and state-level efforts to remake teacher evaluation systems are among the most substantial and widely adopted reforms that U.S. public schools have experienced in decades (McGuinn, 2012). Research on these next generation of evaluation systems has focused overwhelmingly on policy goals, program designs, and performance measures (e.g. Kane, McCaffrey, Miller, & Staiger, 2013). However, we still know very little about how these policies are interpreted and enacted by school leaders. History clearly shows that the success of federal, state, and local policy initiatives depends on the will and capacity of local actors to implement reforms (Honig, 2006). This is particularly true in the decentralized U.S. education system where local practice is often decoupled from central policy (Spillane & Kenney, 2012). Many states and districts require principals to conduct observation and feedback cycles as part of new evaluation systems (Center on Great Teachers and Leaders, 2014; Herlihy et al., 2014). In a number of states, including the one in which our study takes place, principals are given full responsibility for determining teachers' overall summative evaluation ratings (Donaldson & Papay, 2014; Steinberg & Donaldson, in press).

Relying on principals as the primary evaluators raises important questions about their willingness, capacity, and ability to implement observation and feedback cycles and support teacher development through the evaluation process. Principals' views on the primary purpose of evaluation may differ. Some scholars (Hanushek, 2009) and journalists (Thomas, Wingert, Conant, & Register, 2010) see evaluation as a mechanism for increasing teacher effort through accountability and monitoring, and for dismissing ineffective teachers. Others view evaluation as a process that can support the professional growth of teachers by promoting self-reflection, by establishing a common language and framework for analyzing instruction, and by providing individualized feedback (Almy, 2011; Curtis & Weiner, 2012). Evaluation system reforms have also greatly expanded the demands on principal time and the role of principals as instructional leaders. The degree to which principals are prepared to assume this expanded role and the ways in which they navigate these increasing responsibilities have important implications for teacher development and evaluation.

Purpose / Objective / Research Question / Focus of Study:

In this study, we examine the perspectives and experiences of principals as evaluators in a large urban school district in the northeastern United States that recently implemented sweeping reforms to its teacher evaluation system. Our study focuses on principals' perspectives and experiences with classroom observation and feedback because this process is a primary mechanism through which evaluation is intended to promote teacher development. Principals' abilities to rate teachers accurately, to facilitate teachers' own self-reflection, to make specific actionable recommendations, and to communicate this feedback effectively are central to any evaluation process intended to improve instruction. In our view, this paper makes several contributions to the literature. First, the paper is among the first to look inside the black box of how this next generation of evaluations systems are perceived and implemented by principals. Second, we describe how, in the district we studied, four key implementation challenges resulted in unintended consequences that undercut principals' ability to support teachers' professional growth through the evaluation process. Finally, the paper discusses five different proposals to improve the quality of feedback teachers receive through observation and feedback cycles.

Setting / Population / Participants / Subjects:

The district we studied is an urban district in the northeast that serves a racially and linguistically diverse student population. Hispanic and African American students make up approximately 75% of the district student body, while the remaining 25% of students are predominantly Caucasian and Asian American. Over 70% of students in the district are eligible for free or reduced price lunch and nearly half speak a language other than English as their first language. We defined our target population of inference as all principals in the district that oversaw schools serving students in main-stream classes across grades K-12.

Early in the summer of 2013, we recruited a subset of 46 randomly selected principals to participate in the study in order to capture views that were broadly representative of principals across the district as a whole. In order to reduce chance sampling idiosyncrasies that might skew our results, we identified potential participants using a stratified random sampling framework. We chose two school characteristics, school size and level, on which to stratify our sample. Specifically, we categorized all principals into six different strata: three school types (elementary, middle, and high) and two school sizes (390 students or more, less than 390 students). We then contacted up to nine randomly selected principals within each strata by phone and email to invite them to participate confidentially in our study.

Our sampling procedure resulted in a diverse collection of interview participants with demographic characteristics and school assignments that were broadly representative of the district as a whole. Twenty-four out of the 46 principals we contacted agreed to be interviewed, a participation rate of 52%. We conducted a series of t-tests to confirm that our stratified random sample of participating principals is representative of principals across the district. In Table 1, we provided the demographic characteristics and school characteristics for all principals in the district we interviewed and those we did not. We find no statistically significant differences across any measures, strong evidence that our sample is broadly representative of the district.

Research Design, Data Collection and Analysis:

We conducted interviews with principals lasting 45 to 60 minutes in July and August of 2013, the summer after the first year the new evaluation system was implemented district-wide. These interviews gave principals the opportunity to share their perspectives about teacher evaluation as well as their experiences implementing the districts' former and current evaluation systems. The authors and a research assistant conducted each interview individually in person, or by phone, based on principals' availability and preferences. We used a semi-structured protocol to ensure that each interview touched upon a common set of topics and reduced interviewer effects and bias (Patton, 2001). Our research team composed structured, thematic summaries (Maxwell, 2005) of each interview and used these summaries to develop a set of codes that captured the common themes and topics raised by principals.

We coded interview transcripts for central concepts (Strauss & Corbin, 1998) using a hybrid approach to developing codes (Miles & Huberman, 1994). We generated codes informed by our research questions, the theory of action behind classroom observation and feedback cycles, and our review of the instructional leadership literature discussed above, as well as common topics that were reflected in our thematic summaries. Each author then conducted a trial coding process with two transcripts, reviewed the other's initial coding, and debriefed about coding discrepancies and common themes that were not included in our initial set of codes. We analyzed our interview data by organizing codes around broad themes and reviewing interview passages associated with the codes. We wrote analytic memos that outlined the range of

perspectives and experiences that principals shared, and reviewed the characteristics of principals and their schools to situate quotes within context. Once the evidence on each theme was organized into an extended analytic memo, we returned to the interview transcripts to search for disconfirming evidence and counterexamples.

Findings / Results:

In the large urban district we studied, recent reforms to the teacher evaluation system provided a common framework and language that aided principals in assessing and discussing teachers' professional practice. Principals' perceived that teachers were becoming more involved in the evaluation process and that the culture around evaluation was beginning to shift towards a focus on professional growth. These changes provided necessary structures and more fertile contexts for principals to promote growth among their staff as evaluators. However, the expanded role of principals as evaluators resulted in a variety of unintended consequences. These unintended consequences illustrate that *how* an evaluation system is implemented ultimately determines whether it will be successful at promoting teacher development.

Challenge #1: Principals' views on the purpose of evaluation differ. We also found that principals' views on what the evaluation system should be used for did not always align with how the district articulated the purpose of the system or how principals felt teachers perceived the system. These differing views led principals to interpret their role in the evaluation process quite differently. Among the principals we spoke with, the vast majority, over 75%, viewed teacher evaluation as a system that should focus on helping teachers improve their practice. However, four of the administrators we spoke with highlighted the importance of dismissing teachers who were ineffective educators.

Consequence: Principals used the evaluation process in very different ways. Principals leveraged the evaluation process to achieve a range of goals that were not always aligned or consistent with the district's stated intent. Implementation approaches differed substantially even among the majority of principals who viewed improving teachers' instructional practices as their primary goal of the evaluation process. Some principals emphasized the importance of direct feedback that is "specific and actionable, and that comes from a place of knowledge and experience on the part of the administrator." Other principals saw teacher self-reflection as the primary mechanism for improvement. One principal who was a veteran middle school teacher focused on a third mechanism - monitoring and accountability - as a means of motivating teachers to improve their practice.

Challenge #2: The expanded role of principals. Nearly all principals, 88%, expressed real concerns about the increased demands of the new evaluation system. As one principal put it, "the biggest challenge is time." Principals commonly described the process of evaluating all teachers in their schools as "a nightmare" or "nuts." As one principal shared, "It's too much. It almost killed me to try to do all of it."

Consequence: Feedback conversations were infrequent and brief. The demands on principals and their administrative teams to conduct extensive evaluations for all teachers limited the frequency and quality of feedback teachers' received. Several principals expressed concerns that they were unable to provide the frequent feedback necessary for supporting teachers' professional growth because of the sheer number of teachers they were required to evaluate. From the perspective of one principal, if feedback cycles for improvement are "done right, it's a weekly to monthly thing that you do with teachers." Instead, it was all that most principals could do to observe and write the formative and summative evaluations for each teacher in their school.

Challenge #3: Providing feedback outside their expertise. Nineteen of the twenty-four principals we spoke with expressed concerns about their ability to provide meaningful feedback to teachers in all disciplines and levels. Elementary school principals typically characterized this challenge in terms of grade levels. A principal who taught second grade explained that his "weaker point would be the upper grades." For middle school and high school principals, evaluating teachers across different subject areas presented more of a challenge. A principal with five years of experience teaching history and English told us, "history, I do, science and math are a little bit of a challenge." When principals evaluated teachers in subjects and grades they had not taught, principals felt less comfortable and confident in their abilities to evaluate instruction accurately or provide meaningful support.

Consequence: Feedback was narrowly focused on pedagogy. Lack of content expertise led many secondary principals to narrow the focus of their evaluation to general instructional practices and strategies. Eight principals told us how they focused on pedagogy rather than content. Although narrowing the scope of feedback may have improved principal's confidence, it failed to address teachers' need to develop both their core content knowledge and their pedagogical content knowledge, which have been shown to be central elements of effective instruction particularly in math (Wayne & Youngs, 2003; Hill et al., 2008).

Challenge #4: Principals had limited training. The current evaluation system demanded a wide range of skills from principals in order to implement the new process successfully. Principals were required to accurately differentiate teachers on a four point scale, support their ratings with low-inference evidence, communicate these ratings effectively, and prescribe specific, actionable feedback for teachers on how to improve. In the district we studied, evaluator training was focused on familiarizing principals with the expansive rubric and procedural requirements, and calibrating principals to be reliable and accurate raters. At the time, principals had not received any training on how to manage their time to complete all observations or how to engage in productive feedback conversations.

Consequences: Feedback conversations focused on ratings and positive reinforcement rather than on how teachers could improve. Differentiating among teachers who had been told they were satisfactory for many years led to feedback conversations that became focused on the summative evaluation rating itself rather than areas for continued professional growth. Rating teachers lower than they felt was fair often derailed efforts to focus the conversation on professional improvement. Our interviews also suggested that some principals may have avoided difficult conversations with teachers about their weaknesses and, instead, focused on reinforcing the things that were going well in the classroom. Some principals shied away from using feedback conversations to push teachers on their growth areas for fear of jeopardizing this relational trust.

Conclusions:

The quality of feedback teachers receive through the evaluation process depends critically on the time and training evaluators have to provide individualized and actionable feedback. Districts that task principals with primary responsibility for conducting observation and feedback cycles must attend to the many implementation challenges associated with this approach in order for next-generation evaluation systems to successfully promote teacher development.

Appendices

Not included in page count.

Appendix A. References

Allen, J. P., Pianta, R. C., Gregory, A., Mikami, A. Y., & Lun, J. (2011). An interaction-based approach to enhancing secondary school instruction and student achievement. *Science*, *333*, 1034-1037

Almy, S. (2011). *Fair to everyone: Building the balanced teacher evaluations that educators and students deserve*. Washington, DC: Education Trust.

Boyd, D., Grossman, P., Ing, M., Lankford, H., Loeb, S., & Wyckoff, J. (2011). The influence of school administrators on teacher retention decisions. *American Educational Research Journal*, *48*, 303-333.

Bryk, A. & Schneider, B. (2002). *Trust in schools: A core resource for improvement*. New York, NY: Russell Sage Foundation.

Bryk, A., Sebring, P. B., Allensworth, E., Luppescu, S., & Easton, J. (2010). *Organizing schools for improvement: Lessons from Chicago*. Chicago, IL: The University of Chicago Press.

Center on Great Teachers and Leaders. (2014). *National picture: A different view*. Retrieved March 31, 2014 from http://www.gtlcenter.org/sites/default/files/42states.pdf.

Center on Great Teachers and Leaders. (2013). *Databases on state teacher and principal evaluation policies*. Retrieved September 1, 2015 from http://resource.tqsource.org/stateevaldb/.

Coelli, M., & Green, D. A. (2012). Leadership effects: School principals and student outcomes. *Economics of Education Review, 31*, 92-109.

Cubberly, E. P. (1916). *Public School Administration*. Cambridge, MA: The Riverside Press

Curtis, R., & Wiener, R. (2012). *Means to an end: A guide to developing teacher evaluation systems that support growth and development.* Washington, DC: Aspen Institute.

Danielson, C., & McGreal, T. L. (2000). *Teacher evaluation to enhance professional practice*. Alexandria, VA: Association for Supervision and Curriculum Development.

Danielson, C. (2007) *Enhancing Professional Practice: A framework for teaching.* Alexandria, VA: Association for Supervision and Curriculum Development.

Donaldson, M. L. (2012). *Teachers' perspectives on evaluation reform*. Washington, DC: Center for American Progress.

Donaldson, M. L. (2013). Principals' approaches to cultivating teacher effectiveness: Constraints and opportunities in hiring, assigning, evaluating, and developing teachers. *Education Administration Quarterly, 49*, 838-882.

Donaldson, M. L., & Papay, J. P. (2014). Teacher evaluation for accountability and development. In H. F. Ladd & M. E. Goertz (Eds.), *Handbook of research in education finance and policy*. New York: Routledge.

Donaldson, M. L., & Peske, H. G. (2010). *Supporting effective teaching through teacher evaluation: A study of teacher evaluation in five charter schools*. Washington, DC: Center for AmericanProgress.

Frumkin, P. (2003) *Creating new schools: The strategic management of charter schools*. Baltimore, MD: Annie E. Casey Foundation.

Garet, M. S., Porter, A. C., Desimone, L., Birman, B. F., & Yoon, K. S. (2001). What makes

professional development effective? Results from a national sample of teachers. *American Educational Research Journal, 38*, 915-945.

Grissom, J. A., Loeb, S., & Master, B. (2013). Effective instructional time use for school leaders: Longitudinal evidence from observations of principals. *Educational Researcher, 42*, 433-444.

Grubb, W. N., & Flessa, J. J. (2006). "A job too big for one": Multiple principals and other nontraditional approaches to school leadership. *Educational Administration Quarterly*, *42*, 518–550

Halverson, R., Kelley, C., & Kimball, S. (2004). Implementing teacher evaluation systems: How principals make sense of complex artifacts to shape local instructional practice. *Educational administration, policy, and reform: Research and measurement*, 153-188.

Halverson, R. R., & Clifford, M. A. (2006). Evaluation in the wild: A distributed cognition perspective on teacher assessment. *Educational Administration Quarterly*, *42*, 578-619.

Hanushek, E. (2009). Teacher deselection. In D. Goldhaber & J. Hannaway (Eds.), *Creating a new teaching profession* (165-180). Washington, DC: Urban Institute Press.

Herlihy, C., Karger, E., Pollard, C., Hill, H. C., Kraft, M. A., Williams, M., & Howard, S. (2014). State and local efforts to investigate the validity and reliability of scores from teacher evaluation systems. *Teachers College Record*, *116*(1), 1-28.

Hill, H. C., Blunk, M. L., Charalambous, C. Y., Lewis, J. M., Phelps, G. C., Sleep, L., & Ball, D. L. (2008). Mathematical knowledge for teaching and the mathematical quality of instruction: An exploratory study. *Cognition and Instruction, 26*, 430-511

Honig, M. I. (Ed.) (2006). *New directions in education policy implementation: Confronting complexity.* Albany, New York: State University of New York Press.

Horng, E. L., Klasik, D., & Loeb, S. (2010). Principal's time use and school effectiveness. *American Journal of Education*, *116*, 491-523.

Hoy, A.W., & Hoy, W. K. (2012). *Instructional leadership: A research based guide to learning in schools*. 4th Edition. Pearson.

Kane, T. J., McCaffrey, D. F., Miller, T., & Staiger, D. O. (2013). *Have we identified effective teachers? Validating measures of effective teaching using random assignment.* MET Project. Seattle: Bill & Melinda Gates Foundation.

Kimball, S. M., & Milanowski, A. (2009). Examining teacher evaluation validity and leadership decision making within a standards-based evaluation system. *Educational Administration Quarterly*, *45*, 34-70.

Ladd, H. (2011). Teachers' perceptions of their working conditions: How predictive of planned and actual teacher movement? *Educational Evaluation and Policy Analysis, 33*, 235-261.

Lacireno-Paquet, N., Morgan, C., & Mello, D. (2014). *How states use student learning objectives in teacher evaluation systems: A review of statewide websites.* Washington, DC: U.S. Department of Education.

Leithwood, K., & Louis, K. S. (2011). *Linking leadership to student learning*. San Francisco: John Wiley & Sons.

Leithwood, K., Louis. K. S., Anderson, S., & Wahlstrom, K. (2004). *Review of research: How leadership influences student learning*. New York: The Wallace Foundation.

Louis, K., Dretzke, B., & Wahlstrom, K. (2010). How does leadership affect student achievement? Results from a national US survey. *School effectiveness and school improvement*, *21*, 315-336.

Maxwell, J. A. (2005). *Qualitative research design: An interactive approach.* Thousand Oaks, CA: SAGE Publications.

May, H. & Supovitz, J. A. (2011). The scope of principal efforts to improve instruction. *Educational Administration Quarterly, 47*, 332-352.

McGuinn, P. (2012). Stimulating reform: Race to the Top, competitive grants and the Obama education agenda. *Educational Policy*, *26*, 136-159.

Miles, M. & Huberman, M. (1994). *Qualitative data analysis: A expanded sourcebook* (2nd ed.). Thousand Oaks: Sage Publications.

National Council on Teacher Quality (2013*). 2013 state teacher policy yearbook*. Washington, DC: Author.

Newmann, F. M., Smith, B., & Allensworth, E. (2001) Instructional program coherence: What it is and why it should guide school improvement policy. *Education Evaluation and Policy Analysis, 23*, 297-321.

Neumerski, C. M. (2012). Rethinking instructional leadership, a review: What do we know about principal, teacher, and coach instructional leadership and where should we go from here? *Educational Administration Quarterly, 49*, 310-347.

Patton, M. Q. (2001*). Qualitative research and evaluation methods*. (2nd Ed.). Thousand Oaks, CA: Sage Publishing.

Papay, J. (2012). Refocusing the debate: Assessing the purposes and tools of teacher evaluation. *Harvard Educational Review, 82*, 123-141.

Papay, J. P., & Johnson, S. M. (2012). Is PAR a good investment? Understanding the costs and benefits of teacher peer assistance and review programs. *Educational Policy*, *26*, 696-729.

Peterson, K. (2002). The professional development of principals: Innovations and opportunities. *Educational Administration Quarterly*, *38*, 213-232.

Popham, W. J. (1988). The dysfunctional marriage of formative and summative teacher evaluation. *Journal of Personnel Evaluation in Education*, *1*, 269-273.

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, *73*, 417-458.

Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *American Economic Review*, 247-252.

Sartain, L., Stoelinga, S. R., & Brown, E. R. (2011). *Rethinking teacher evaluation: Lessons learned from observations, principal-teacher conferences, and district implementation.* Chicago: Consortium on Chicago School Research.

Scriven, M. (1995). A unified theory approach to teacher evaluation. *Studies in Educational Evaluation, 21*, 111-129.

Spillane, J. P., & Kenney, A. W. (2012). School administration in a changing education sector: The US experience. *Journal of Educational Administration*, *50*, 541-561.

Steinberg, M. P., & Donaldson, M. L. (in press) The new educational accountability: Understanding the landscape of teacher evaluation in the post NCLB era. *Education Finance and Policy*.

Steinberg, M. P., & Sartain, L. (in press). Does teacher evaluation improve school performance? Experimental evidence from Chicago's excellence in teaching project. *Journal of Policy Analysis and Management*.

Strauss, J. & Corbin, A. (1998). *Basics of qualitative research: Grounded theory procedures and techniques*. (2nd Ed.). Thousand Oaks, CA: SAGE Publications.

Stronge, J. H. (2005). *Evaluating teaching: A guide to current thinking and best practice*. Corwin Press.

Supovitz, J., Sirinides, P., & May, H. (2010). How principals and peers influence teaching and learning. *Educational Administration Quarterly, 46*, 31-56.

Taylor, E. S., & Tyler, J. H. (2013). The effect of evaluation on teacher performance. *American Economic Review, 102*, 3628-3651.

Thomas, E., Wingert, P., Conant, E., & Register, S. (2010). Why we can't get rid of failing teachers. *Newsweek, 155*(11), 24-27.

Toch, T., & Rothman, R. (2008). *Rush to judgment: Teacher evaluation in public education*. Washington, DC: Education Sector.

Tucker, P. D. (1997). Lake Wobegon: Where all teachers are competent (or, have we come to terms with the problem of incompetent teachers?). *Journal of Personnel Evaluation in Education*, *11*, 103-126.

Turnball, B. J., Haslam, M. B., Arcaira, E. R., Riley, D. L., Sinclair, B., & Coleman, S. (2009). *Evaluation of the School Administration Manager Project*. Washington, DC: Policy Studies Associates, Inc.

Turnball, B. J., White, R. N., & Arcaira, E. R. (2010). *Achievement trends in schools with School Administration Managers (SAMs)*. Washington, DC: Policy Studies Associates, Inc.

Waters, T., Marzano, R. J., & McNulty, B. (2003). *Balanced leadership: What 30 years of research tells us about the effect of leadership on student achievement*. Aurora, CO: Mid-continent Research for Education and Learning.

Wayne, A. J., & Youngs, P. (2003). Teacher characteristics and student achievement gains: A review. *Review of Educational Research, 73*, 89-122.

Weatherley, R., & Lipsky, M. (1977). Street-level bureaucrats and institutional innovation: Implementing special-education reform. *Harvard Educational Review*, *47*, 171-197.

Weisberg, D., Sexton, S., Mulhern, J., Keeling, D., Schunck, J., Palcisco, A., & Morgan, K. (2009). *The widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness*. Brooklyn: The New Teacher Project.

Wexler Eckman, E. (2006). Co-principals: Characteristics of dual leadership teams. *Leadership and Policy in Schools*, *5*, 89–107.

Yoon, K. S., Duncan, T., Lee, S. W. Y., Scarloss, B., & Shapley, K. (2007). *Reviewing the evidence on how teacher professional development affects student achievement*. Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southwest.

Appendix B. Tables and Figures

Table 1
Principal and School Demographic Information

	Interviewed	Non-Interviewed	*p*-value
Principals Characteristics			
African American	0.46	0.39	0.54
Caucasian	0.38	0.44	0.60
Hispanic	0.08	0.16	0.32
Asian American	0.08	0.01	0.06
Male	0.42	0.28	0.21
Age (years)	47.52	47.21	0.90
School Characteristics			
Elementary	0.46	0.41	0.66
Middle	0.13	0.06	0.27
High	0.17	0.21	0.65
Traditional	0.63	0.69	0.58
African American (%)	34.76	34.75	1.00
Hispanic (%)	41.47	44.46	0.48
White (%)	11.54	12.46	0.76
Asian (%)	10.05	5.52	0.06
Independent Education Plans (%)	17.03	19.12	0.18
English Language Learners (%)	29.00	29.55	0.89
Low Income (%)	70.06	71.02	0.77
Proficient in English language arts (%)	49.29	46.99	0.64
Proficient in mathematics (%)	42.57	41.80	0.86
Observations	24	86	

Note: *P*-values are derived from two-sample *t*-tests of the mean difference in a given characteristic across interviewed and non-interviewed principals. Proportions of schools that are elementary, middle, and high school do not sum to one because of schools with non-traditional grade configurations.