Teachers' Use of Evaluation for Instructional Improvement

and School Supports for This Use

Min Sun, *University of Washington*

R. Brock Mutcheson, *Virginia Tech*

Jihyun Kim, *Michigan State University*

As states and districts reinvent teacher evaluation systems to provide better measures of teacher performance, their utilization of information to improve instruction has become key to improving student learning outcomes. One popular use is to inform human resources decisions regarding hiring, firing, promoting, and compensating teachers (Chingos & West, 2011; Rockoff, Staiger, Kane, & Taylor, 2012). Another important mechanism, discussed much less in the research literature, involves how teachers themselves use information about their past performance to improve instruction.

Conventional teacher evaluation systems have been criticized for not promoting individual teacher development (Toch, 2008; The New York Times, 2012). The lack of information on the status of an individual's teaching and what constitutes effective teaching are often identified as main barriers to improvement (Stiggins & Bridgeford, 1985; Taylor & Tyler, 2012). To overcome this issue, newly-designed teacher evaluation systems commonly use multiple measures of teacher performance that focus on classroom instruction and are closely linked to student learning outcomes. These new approaches, if effective, have the potential to promote teacher improvement in several ways. First, teachers have opportunities to reflect on their practices when they collect evidence to demonstrate their own effectiveness in promoting students' learning. Second, links between evaluation and professional development can potentially make evaluation practices constructive. Third, most new evaluation procedures include feedback from principals, peers, students, and parents.

In this chapter, we discuss teachers' reports on the use of teacher evaluation to improve their instruction and what types of school supports aid in this process. We conducted descriptive analyses of survey data that were collected in two rural school districts in Southwest Virginia during the 2012-13 and 2013-14 school years, following the state of Virginia receipt of a waiver

from requirements of No Child Left Behind Act in 2012 and moves to revamp teacher evaluation. Our analyses aimed to address the following questions: Which types of teachers are more likely to use evaluation for improvement? What types of school supports are related to teachers' use of evaluation for improvement? And how does the relationship between school supports and teachers' use vary for different types of teachers?

Anticipating our key findings, early-career teachers who were on annual contracts, on average, were more likely to use evaluation to improve their instruction than teachers who were continuing contracts. The top one-fourth and bottom one-fourth teachers were significantly more likely to use performance information for improvement than teachers in the middle of the distribution. Moreover, among a variety of potentially effective school supports for teachers' use evaluation for improvement (e.g., professional development, principal feedback, and collegial supports), the usefulness of principals' feedback particularly mattered to early-career teachers.

The remainder of this chapter is organized as follows. We first present the key components of the evaluation policy in the two districts. Next, we develop several conjectures regarding how this policy may influence teachers' instruction, potential variations across subgroups of teachers, and potentially effective school supports. We then describe the sample and measures. Lastly, we present the analytic strategies for testing these conjectures, summarize main findings, and discuss their policy implications.

Teacher Evaluation Policies in Two Virginia Districts

Virginia issued the *Guidelines for Uniform Performance Standards and Evaluation Criteria for Teachers* on July 1, 2012, which marked a significant overhaul of conventional evaluation criteria. The guidelines called for 40% of teacher evaluation to be based on the standard of *student academic progress*. Each of the other six standards is weighted 10% in

teachers' total evaluation scores: *professional knowledge, instructional planning, instructional delivery, assessment for and of learning, learning environment,* and *professionalism.* For each standard, teacher performance is rated on a four-point scale from "exemplary" to "unacceptable." The annual summative evaluation separates teacher performance into four groups: those whose performance consistently exceeds expectations on all standards ("exemplary"), those who meet the standards ("proficient"), those who have an opportunity for improvement ("developing"), and those who do not meet expectations ("unacceptable").

The evaluation procedures, instruments, and materials adopted by the two districts (pseudonyms: District 1 and District 2) in this study were essentially consistent with the state guidelines and recommendations. The districts collected evidence on teachers' performance from three sources. First, teachers underwent classroom observations and walkthroughs conducted by administrators or peers. Probationary teachers (similar to pre-tenured teachers) were observed at least three times per year. Teachers employed under a continuing contract (similar to tenured teachers) were observed at least once per year. Second, teachers were asked to set up learning goals for students and to document the extent to which students had achieved these goals by the end of the semester or school year. Third, districts conducted student surveys. At the end of each semester, teachers were asked to provide a summary of the survey data that described their strengths and areas in need of improvement as well as their own reflections.

In both school districts, teachers on annual contracts received a summative evaluation during each of their probationary years (a.k.a., pre-tenure) and these teachers also received a mid-year interim review to provide systematic feedback prior to the summative review. However, the evaluation procedures for teachers on continuing contracts (a.k.a., with tenure) differed. District 1 teachers received a summative evaluation every year, while counterparts in

District 2 received a summative evaluation every three years. These attributes of evaluation policies and the variation between districts and for different teachers had important implications for the conceptual framing of this study and data analyses.

Conceptual Framing

How Do Teachers Use Performance Information? The use of performance measures based on explicit and uniformly defined criteria and metrics has long been a fundamental component of both public- and private-sector personnel incentive systems (Heinrich & Marschke, 2010). One of the most popular concepts for understanding this performance measurement system is "principal-agent theory" (e.g., Eisenhardt, 1989). This theory is based on the assertion that the principal (e.g., the employer) cannot observe the true effects of agents (e.g., employees); rather, only agents themselves know their true effort and performance. To address this asymmetric information, the employer can align the incentives for employees with the organization's goals. For example, aligning measures of teachers' performance closely with students' learning is expected to motivate teachers' efforts and in turn contribute to schools' organizational values. Moreover, performance management, such as the increase in incentives via performance accountability, is more effective if principals and agents engage in interactive dialogues that support more effective use of performance information and organizational learning (Moynihan 2008).

This conception of using performance information to initiate changes is congruent with a recent study of teacher performance evaluation in the Cincinnati Public Schools. Taylor and Tyler (2012) studied a sample of midcareer elementary and middle school teachers, all of whom were evaluated in a yearlong program based largely on classroom observation between the 2003-04 and 2009-10 school years. Their analyses showed that teachers were more effective at raising

student achievement during the school year when they were being evaluated than they were previously, and even more effective in the years after evaluation, particularly in mathematics. Yet, this study failed to identify how teachers' use of performance information and what supports for this use.

In this chapter, we hypothesize three potential processes of teacher improvement in the two districts. First, learning about the seven professional teaching standards and documenting their performance may have helped teachers reflect on their instruction. During the process of collecting student survey data and documenting students' academic progress, teachers were asked to analyze these data to identify personal strengths and weaknesses and ways to improve their instruction. Second, the links between evaluation and professional development (or individual assistance programs) may have made the evaluation practices more constructive and effective. After being informed of their strengths and weaknesses, teachers could make more intelligent decisions in terms of selecting professional development programs that cater to their development needs. The link between teachers' needs identified by evaluation and professional development can strengthen teachers' learning from evaluation.

Lastly, learning to teach is situated in social contexts through interactions with others and through one's own practices in classroom settings (Putnam & Borko, 2000). The following situations provide rich learning opportunities for teachers: observing others and being observed by others in actual classrooms, referring to one's own practices in group discussions, and integrating what is learned outside of classrooms into one's own practices. The uniform professional standards and evaluation criteria may have facilitated teachers' interactions with their administrators and peers by giving them a common "language" to use in discussing

instruction. Their discussions were likely more relevant to teachers because they were based on evidence from observing their classroom practices.

We further hypothesize that the use of evaluation for instructional improvement may vary by individuals. For instance, studies have shown that teachers make steep gains in effectiveness early in their careers (Boyd, Grossman, Lankford, Loeb, & Wyckoff, 2008; Harris & Sass, 2011; Papay & Kraft, 2013; Rockoff, 2004; Wiswall, 2013). In general, probationary teachers may show more interest in learning and improvement. Moreover, they face more risks of losing their jobs due to a bad evaluation than teachers who have continuing contracts; therefore, they may be more motivated to improve. We also expect variations by teachers' prior ratings. Teachers who received low ratings in a prior year may have an incentive to improve, while teachers who were rated at the top may be motivated to become better and stay at the top. In comparison, teachers at the middle of the distribution may have the least incentive to change, because they are not directly threatened by the negative consequences of losing their jobs or being assigned to special assistance, or by the positive consequences of receiving a bonus or recognition.

What Types of School Supports are Related to Teachers' Use of Evaluation for Improvement? Next, we describe three types of school supports that seem related to teachers' use of evaluation for improvement. First, professional development programs can help teachers understand the goals and procedures of teacher evaluation and help them see the evaluation process as a means of improving their practices. These programs should also develop teachers' capacity to collect measures of their effects on students' growth and use these measures to adjust their instructional practices (Lachat & Smith, 2005; Supovitz, 2006).

Second, principals and other observers should provide timely feedback as close to the time when the observation happens as possible (Goe, 2013). It should also be specific and tied to

aspects of effective teaching. Useful feedback often focuses on evidence about an individual teacher's pedagogy, curriculum and materials, and ways they can improve their practices and student learning outcomes. Furthermore, the tone of the conversations between principals and a teacher should be constructive by inviting the teacher to share her thoughts, with a goal of developing a common understanding about high-quality teaching.

Lastly, teachers' use of evaluation data may also be influenced by their colleagues (Sun, Penuel, Frank, Gallagher, & Youngs, 2013). We hypothesize that collegial interactions will generate a larger effect on teachers if they address specific details from the teacher evaluation process. Moreover, teachers benefit from help with general instructional issues, including knowledge about differentiating instruction, using data to adjust their teaching, and classroom management skills. In the following sections, we use empirical data to assess these conjectures about the effects of supports on teachers' use of evaluation for improvement.

Sample and Measures

Sample. The two participating school districts included 17 elementary and secondary schools: 3 high schools, 3 middle schools, and 11 elementary schools. These two districts were 30 miles apart and located in rural areas or small towns. They served students with similar demographic characteristics (predominantly white) and poverty status. In 2013-14, District 1 had all of its schools accredited by the state based its performance on the 2012-13 state standardized tests, while 9% of District 2 schools received accreditation with warning because these schools did not meet the state benchmark in mathematics. For federal accountability measures, 16.7 percent of the schools in District 1 and 36.36 percent of the schools in District 2 did not meet the Annual Measurable Objectives (AMO) for all subjects. In our later analyses, we used school

fixed effects to control for variation in school performance and, consequently, variation in the accountability pressures experienced by different schools.

In this chapter, besides using district administrative data on personnel backgrounds and evaluation ratings, we primarily drew on data from a teacher survey administered in fall 2014 to learn about teachers' practices and the implementation of teacher evaluation policy in 2013-14. Survey response rates were about 72% in District 1 and 58% in District 2. District 1 had a total of about 254 full-time equivalent teachers (FTE), while District 2 had about 387 FTE teachers. There were no differences in the percentage of white teachers or in average prior ratings between respondents and non-respondents in either district. However, probationary teachers were significantly more likely to respond to the survey in both school districts. This sample with over-representation of probationary teachers allows us to sufficiently capture probationary teachers' responses to the new evaluation system, given that only 11 percent of the teachers in District 1 and only 16 percent in District 2 were probationary.

Measures. We drew on teacher survey data to assess teachers' use of evaluation for improvement and how school supports were associated with this use in the 2013-14 school year. For composite variables (e.g., *teachers' use of evaluation for performance improvement*, *collegial support for general instruction* and *for evaluation implementation*, and *usefulness of principal feedback*), we ran exploratory factor analysis to explore the latent constructs of the survey items relevant to these composite variables. Survey items are on a four-point scale of 0=not at all, 1=some extent, 2=moderate extent, and 3=great extent. We considered item loadings of at least 0.4 necessary for inclusion in a factor and eigenvalue >1 to be considered as a factor. Each factor had high internal consistency (Cronbach's α>0.85 and eigenvalue=5.36).

Teachers' use of evaluation for performance improvement is the standardized mean of teachers' responses to four items. These included, for example, "The information collected about my teaching in the 2012-13 evaluation really helped me to know my strengths and weaknesses" and "My professional development programs were more informative because their selection was based on my evaluation results in 2012-13."

Collegial support for general instruction is the sum of teachers' responses to seven items about the extent to which each of the following aspects about instruction was discussed between the teacher and her/his colleague. These included, for example, "How the lesson relates to the curriculum and the sequence of learning for the class," "Knowledge of students and differentiate instruction," "What changes I could make if I were re-teach the lesson," and "Alternative strategies of delivering the content to my students."

Collegial support for evaluation implementation is the sum of teachers' responses to seven items about the extent to each of the following aspects about teacher evaluation was discussed between a teacher and her/his colleagues. These included, for example, "The multiple measures of teaching," "How to document student progress using formative assessments," "How to use data on student academic progress to adjust my teaching," and "Communicating my teaching and student performance with my principal."

Usefulness of principal feedback is a mean of 11 items about teachers' perceptions of principal feedback during post-observation conferences and the end-of-year summative evaluation conference. These included, for example, "How the lesson related to the curriculum and the sequence of learning for the class," "My knowledge of my students and differentiating instruction," "Most discussions were constructive and I can directly use the feedback to improve

10

my instruction," and "Overall, the conferencing helped to develop a common understanding about high-quality teaching between me and my principal."

Frequency of principal feedback is measured as the number of times a teacher reported that her/his principal conducted informal visits to her/his classrooms (e.g., walkthroughs), formal observations, post-observation conference, and/or summative evaluation conferences. We then summed the frequencies across these different types of interactions between a teacher and her/his principal.

Professional development is indicated by both the frequency and the usefulness of professional development programs that help teachers implement teacher evaluation policy and use data to improve instruction. Teachers reported the frequency by the number of hours of professional development they participated in in 2013-14 and the usefulness on a four-point scale.

Variations in the Use of Evaluation for Improvement among Teachers

Employment status. The variable *teachers' use of evaluation for improvement* was regressed on a dummy variable for whether the teacher was on either a probationary or continuing contract, controlling for teachers' gender, subject areas (1=high-stakes subjects, e.g., reading, math, and science; 0=otherwise), advanced degree (1=master's or higher; 0=otherwise), and race (1=white; 0=otherwise). We used school fixed effects to account for variation in accountability pressure due to school performance on state standardized tests, variation in school-level implementation of teacher evaluation, and variation in school contexts and practices. Probationary teachers were about 0.3 standard deviation higher than the expected use of evaluation for improvement (*p*=0.06), while teachers on continuing contracts were about 0.045 standard deviation below the expected use (*p*=0.46). Probationary teachers' estimated use

was significantly higher than the use of teachers on continuing contracts at the 95 percent confidence level ($F=4.24$, $p=0.04$).

Prior ratings. We separated teachers based on prior performance ratings in the 2012-13 school year into four groups: prior ratings in the first quartile (0-25%, low performers), prior ratings in the second quartile (26%-50%, mid-low performers), prior ratings in the third quartile (51%-75%, mid-high performers), and prior ratings in the fourth quartile (76%-100% and high performers). The variable *teachers' use of evaluation for improvement* was regressed on the dummy variables for these four groups, controlling for teachers' employment status, gender, subject area, race, advanced degree, and school fixed effects.

Teachers who were at the bottom and the top of the distribution of prior ratings were more likely to report using evaluation for improvement. In particular, high-performing teachers' estimated use was about 0.29 standard deviation above the expected use, which is significant at a 90% confidence level ($p=0.07$). In contrast, teachers at the middle of the distribution reported about 0.2 standard deviation lower than were expected. Moreover, the difference of use information between top performers and middle performers were significant ($F=6.18$, $p=0.01$ for mid-low performer; and $F=5.41$, $p=0.02$ for mid-high performers).

School Supports for Teachers' Use of Evaluation for Improvement. Next, we regressed teachers' use of evaluation for improvement on several types of school supports, by controlling for teacher covariates and school fixed effects. When estimating the effects of principal feedback, we replaced school fixed effects with school covariates, because school fixed effects would have absorbed all of the variation in principal leadership. We first estimated the main effects of these supports, and then added interaction effects between the school support variables and teacher subgroups by their employment status and prior ratings. These interaction terms

Table 1. Effects of School Supports for Teachers' Use of Evaluation for Instructional Improvement

	Main effects	x probationary teacher	x prior rating at the 1st quartile	x prior rating at the 2nd quartile	x prior rating at the 3rd quartile
Professional Development	0.017***	0.002	0.005	-0.011	0.031*
	(0.004)	(0.015)	(0.024)	(0.012)	(0.015)
Main effects		X	X	X	X
Teacher covariates	X	X	X	X	X
School fixed effects	X	X	X	X	X
N	167	167		115	
adj. R-sq	0.184	0.179		0.245	
Usefulness of Principal Feedback	0.439***	0.886**	-0.251	-0.019	-0.104
	(0.092)	(0.324)	(0.322)	(0.299)	(0.325)
Main effects		X	X	X	X
Teacher covariates	X	X	X	X	X
School covariates	X	X	X	X	X
N	256	256		173	
adj. R-sq	0.123	0.145		0.166	
Frequency of Principal Feedback	0.000	0.004	0.012	0.013	0.012
	(0.004)	(0.016)	(0.015)	(0.020)	(0.022)
Main effects		X	X	X	X
Teacher covariates	X	X	X	X	X
School covariates	X	X	X	X	X
N	262	262		175	
adj. R-sq	0.035	0.032		0.057	
Collegial Support for General Instruction	0.042***	0.033	0.004	0.020	-0.043
	(0.011)	(0.036)	(0.035)	(0.034)	(0.042)
Main effects		X	X	X	X
Teacher covariates	X	X	X	X	X
School fixed effects	X	X	X	X	X
N	262	262		175	
adj. R-sq	0.104	0.103		0.161	
Collegial Support for Evaluation Implementation	0.046***	0.031	-0.031	-0.021	0.005
	(0.012)	(0.029)	(0.035)	(0.036)	(0.040)
Main effects		X	X	X	X
Teacher covariates	X	X	X	X	X
School fixed effects	X	X	X	X	X
N	262	262		175	
adj. R-sq	0.112	0.112		0.204	

Notes: Teacher covariates include teachers' prior ratings, teaching high-stakes subjects, probationary contract, female, and white. School covariates include dummy variables for middle and high schools, and district dummy variable. We did not include other school covariates, such as demographics of students and location, because these two school districts served students with very similar socio-economic backgrounds.
*p<0.05 ** p<0.01 *** p<0.001

enabled us to assess the heterogeneous effects of these school supports. The first column in Table 1 includes only main effect estimates, the second column includes estimates of effects of interactions with being a probationary teacher, and the last three columns include interaction effects with teachers' prior performance ratings.

Professional development programs with both intensity and quality would positively influence the likelihood that teachers report using evaluation for improvement (β=0.017, p<0.001). There were no statistically significant differences in the effects of professional development between probationary teachers and teachers on continuing contracts. However, professional development had a significantly larger effect for teachers with mid-high prior performance ratings than for those with high prior ratings (β=0.031, p<0.05).

We then examined two attributes of principal feedback: usefulness and frequency. With a one-unit change in usefulness (e.g., from "to some extent" to "to a moderate extent"), it is estimated that teachers would increase their use by 0.44 standard deviation (p<0.001). More interestingly, principal feedback had a larger effect on probationary teachers than on teachers with continuing contracts. With a one-unit increase in the usefulness of principal feedback, probationary teachers would increase their use by an additional 0.89 standard deviation (p<0.01). Furthermore, after controlling for employment status and other covariates, we did not observe variation in the effect of principal feedback by teachers' prior ratings. In contrast, the frequency of principal feedback did not have significant influence on teachers' use, nor were there differential effects among different types of teachers.

Lastly, we assessed the effects of two types of collegial supports for general instruction and for the implementation of this new teacher evaluation system. The results indicated that collegial supports for both general instruction (β=0.042) and evaluation implementation

(β=0.046) were positively related to teachers' use of evaluation for improvement at the 99.9% confidence level. The effects were relatively homogenous across different types of teachers by employment status and prior performance ratings.

Discussion

This chapter aims to provide descriptive evidence on teachers' use of evaluation for instructional improvement and school supports for this use. Data were collected from two rural school districts where multiple performance measures focused on students' academic progress had recently been implemented. The findings in this chapter are relevant for school leaders, policymakers, and researchers who are interested in understanding how to promote instructional improvement through building teacher capacity.

Probationary teachers who were on annual contracts, on average, were more likely to use evaluation to improve their instruction than teachers who were on continuing contracts. It may be that early-career teachers were eager to learn or that they perceived more pressure to improve under the performance accountability system. If the goal of using teacher evaluation is to not only hold early-career teachers accountable for continuous improvement, but also teachers later in their careers, perhaps school leaders should rethink the design of their performance evaluation and feedback systems to motivate a wider range of teachers beyond those with probationary status.

Moreover, teachers who were at the top 25 percent of the distribution of the prior year's evaluation ratings were significantly more likely to use information for improvement than teachers in the middle of the distribution. This strongly implies a theory of positive reinforcement; i.e., teachers who were rated as high performing were motivated to stay on top and were more likely to buy into the notion that this system could benefit their instruction.

Alternatively, it could also be the case that good teachers always seek to improve regardless of how they are evaluated. A future longitudinal analysis that accounts for a time-invariant measure of teacher quality will help to further separate these two explanations.

Lastly, school supports, including professional development, principal feedback, and collegial supports, were found to be related to teachers' use of evaluation for improvement. This is not surprising, given that prior studies have provided similar evidence that school supports positively contribute to teacher improvement (Kraft & Papay, 2014; Sun et al., 2013). Here we underscore three additional observations. First, professional development had a stronger effect on teachers at the upper-middle range of prior ratings than on the top-performing teachers. Second, the frequency of principal feedback had little impact on teachers' reported use of evaluation for improvement, while teachers' perceived usefulness of this feedback really mattered, particularly for early-career teachers. Third, except for the above two heterogeneous effects, the school supports had relatively homogenous influences regardless of teachers' employment status and prior ratings.

The interpretation of the relationship between school supports and teachers' use for improvement is constrained by other limitations of our data. The measure of teachers' use of evaluation for improvement and the measures of school supports both came from the same teachers' reports. This is so-called "common source bias." For example, it could be that a good teacher is more likely to seek professional development opportunities, make effective use of them and report high satisfaction with them. At the same time, this teacher is also more likely to report the use of evaluation for instructional improvement. This issue can be addressed in future studies by collecting alternative measures from different sources, such as a third party's classroom observations to measure teachers' use of evaluation for instructional improvement. In

16

addition, we were unable to assess change in teachers' use over time or the long-term effect of this new teacher evaluation system, since we observed teachers' use in only one year at a relatively early stage of the policy implementation. Lastly, we were unable to verify that using evaluation to strengthen instruction would lead to improvements in student achievement. Verifying this link will further establish the validity of this measure.

Despite these limitations, we hope this chapter can elevate the conversation about using teacher evaluation systems to support the professional improvement for all teachers. Although performance accountability through personnel management strategies are critical to understand, exploring ways to support teachers in using performance information to adjust instruction is an important activity for educational leaders and researchers to engage in. This effort can be crucial part of broader initiatives to build school capacity to better serve students in this performance accountability era.

References

__ (Sep. 16, 2012). In Search of Excellent Teaching. *The New York Times*. Retrieved on

Dec.12.2014 from http://www.nytimes.com/2012/09/17/opinion/in-search-of-excellent-

teaching.html

Boyd, D., Grossman, P., Lankford, H., Loeb, S., & Wyckoff, J. (2008). *Who leaves? Teacher*

attrition and student achievement. NBER Working Paper No. 14022. Retrieved at

http://www.nber.org/papers/w14022

Chingos, M. M. & West, M. R. (2011). Promotion and reassignment in public school districts:

How do schools respond to differences in teacher effectiveness? *Economics of Education*

Review, 30, 419-433.

Eisenhardt, K.M. (1989). Agency theory: An assessment and review. *The Academy of*

Management Review, *14*(1). 57-74.

Goe, L. (2013). Can teacher evaluation improve teaching? *Principal leadership,* 13, 25-29.

Retrieved from http://www.nassp.org/Content/158/PLmar13_goe.pdf

Harris, D. N., & Sass, T. R. (2011). Teacher training, teacher quality and student achievement.

Journal of Public Economics, 95, 798-812.

Heinrich, C. J., & Marschke, G. (2010). Incentives and their dynamics in public sector

performance management systems. *Journal of Policy Analysis and Management, 29*(1), 183-

208.

Kraft, M. A., & Papay, J. P. (2014). Can professional environments in schools promote teacher

development? Explaining heterogeneity in returns to teaching experience. *Educational*

Evaluation and Policy Analysis, 36(4), 476-500.

Lachat, M. A., & Smith, S. (2005). Practices that support data use in urban high schools. *Journal of Education for Students Placed at Risk, 10*(3), 333–349.

Moynihan, D. P. (2008). *The dynamics of performance management: Constructing information and reform*. Washington DC: Georgetown University Press.

Papay, J. P., & Kraft, M. A. (2013). *Productivity returns to experience in the teacher labor market: Methodological challenges and new evidence on long-term career growth*. Working Paper. Available at: *http://scholar.harvard.edu/files/mkraft/files/papay_kraft_-_rte_may_2013_-_v2.pdf*.

Putnam, R. T, & Borko, H. (2000). What do new views of knowledge and thinking have to say about research on teacher learning? *Educational Researcher, 29*(1), 4-15.

Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *American Economic Review,* 94 (2), 247-252.

Rockoff, J. E., Staiger, D.O., Kane, T. J., & Taylor, E.S. (2012). Information and employee evaluation: Evidence form a randomized intervention in public schools. *American Economic Review, 102*(7), 3184-3213.

Stiggins, R. J., & Bridgeford, N .J. (1985). Performance assessment for teacher development. *Educational Evaluation and Policy Analysis, 7*(1), 85-97.

Sun, M., Penuel, W. R., Frank, K. A., Gallagher, H.A., & Youngs, P. (2013). Shaping professional development to promote the diffusion of instructional expertise among teachers. *Educational Evaluation and Policy Analysis, 35(*3), 344-369.

Supovitz, J. A.(2006). Th*e case for district-based reform: Leading, building, and sustaining school improvement*. Cambridge, MA: Harvard Education Press.

Taylor, E S., & Tyler, J. H (2012). The effect of evaluation on teacher performance. *American Economic Review, 102* (7): 3628-3651.

Toch, T. (2008). Fixing teacher evaluation. *Educational Leadership, 66*(2), 32-37.

Wiswall, M. (2013). The dynamics of teacher quality. *Journal of Public Economics,* 100, 61-78.